**Title**
Multicellularity, genetics, and the discovery of a choanoflagellate sexual life cycle



**Author**
Levin, Tera Catherine

**Publication Date**
2014

Peer reviewed|Thesis/dissertation

# Multicellularity, Genetics, and the Discovery of a Choanoflagellate Sexual Life Cycle

By

Tera Catherine Levin

A dissertation submitted in partial satisfaction

of the requirements for the degree of

Doctor of Philosophy

in

Molecular and Cell Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Nicole King, Chair
Professor Iswar Hariharan
Professor Michael Eisen
Professor John Taylor

Spring 2014

Multicellularity, Genetics, and the Discovery
of a Choanoflagellate Sexual Life Cycle



by Tera Levin

# Abstract

## Multicellularity, genetics, and the discovery of a choanoflagellate sexual life cycle

by

Tera Catherine Levin

Doctor of Philosophy in Molecular and Cell Biology

University of California, Berkeley

Professor Nicole King, Chair

The biology of the first animals provided the foundation upon which all animal diversity later evolved. Yet, we know surprisingly little about the first animals or how they arose. My doctoral research uses the choanoflagellate *Salpingoeca rosetta*, one of the closest living relatives of animals, to shed light on the cellular and molecular changes that led to the origin of multicellularity in animals. To this end, I have examined the genetic basis of choanoflagellate multicellular development and discovered the sexual life cycle of choanoflagellates.

Multicellular development is linked to sexual reproduction in many eukaryotes, including animals. Indeed, the single-celled bottleneck of the fertilized egg is thought to be critical for reducing genetic conflict and limiting the emergence of cheater cells in multicellular organisms. Choanoflagellates exhibit their own form of facultative multicellularity when they develop into "rosette" or "chain" colonies, but it was not known if this multicellularity was linked to a sexual life cycle, nor if these organisms possessed a sexual life cycle in the first place.

My dissertation details the discovery of a sexual life cycle in *S. rosetta*, based on several independent lines of evidence. First, *S. rosetta* laboratory cultures can asexually reproduce as either haploid or diploid cells, and they can be induced to switch ploidy from haploid to diploid and back. Second, the presence of genome-wide haplotype blocks in laboratory isolates demonstrated a history of meiotic recombination in *S. rosetta*. Finally, during the haploid-to-diploid transition, time-lapse microscopy revealed instances of cell fusion, as would be expected during sexual reproduction. We concluded that *S. rosetta* has a sexual cycle. Future work will determine whether the sexual cycle is linked to the development of any of the morphological cell types, including the multicellular forms.

Although choanoflagellate and animal multicellular development bear morphological similarities and both evolved at similar points in evolutionary history, it is not known whether these traits evolved independently or had a single, ancient origin. Thus, I sought to identify the genes underlying choanoflagellate

multicellular development to determine if they were related to genes used in animal development. Chapter three of my dissertation describes the first choanoflagellate forward genetic screen, which I used to study the genetic basis of rosette colony development. From 15,344 clonal lines screened, we isolated nine mutants with distinct defects in rosette development, including mutants with altered cell packing within rosettes, others that developed into large, amorphous multicellular structures, and one that rarely differentiated into multicellular forms.

I focused on a single mutant, Rosetteless, that was never observed to form rosette colonies but appeared to be otherwise wild type. A combination of genome re-sequencing and the first choanoflagellate cross revealed that the Rosetteless phenotype was perfectly linked to a mutation that causes misregulated splicing in a predicted C-type lectin. Thus, the *rosetteless* lectin is the first gene to be linked to a role in rosette colony development. Additionally, the genotypes of the cross progeny allowed for the construction of a preliminary linkage map and revealed that *S. rosetta* inheritance fits the expectations of standard, Mendelian genetics. Further characterization of the other eight mutants promises to reveal additional genes in the rosette formation pathway.

With the advent of choanoflagellate forward genetics, it should now be possible to fully elucidate the genetic basis of rosette formation, or other interesting aspect of choanoflagellate biology, even in the absence of transgenic tools. It is particularly intriguing that the Rosetteless lectin is related to an important class of animal genes but not one that we would have anticipated with a candidate gene approach. This demonstrates the utility of forward genetics for understanding rosette formation and hints at possible homology between the rosette development pathway and pathways in animals.

# Table of Contents

# List of Figures and Tables

# Acknowledgements

The act of scientific research necessarily involves a series of ups and downs, times when nothing makes sense and times when the results finally seem to fit together. Choanoflagellate research in particular has a distinct wandering-in-the-wilderness feel, where you might discover something exciting but there are few tools and little previous research to guide you. I have therefore been extremely grateful for the ideas, advice, feedback, and guidance I have received from my mentors, lab mates, colleagues, and peers.

First and foremost, I would like to thank my advisor Nicole King for taking a chance on my crazy project and allowing me to go on this scientific adventure, for showing me how to write beautiful papers, for tolerating my nagging prodding, and for helping me grow into an independent scientist.

I am also very lucky to have had the chance to mentor and work with Allie Greaney, an undergraduate researcher, for the past several years. Many of the experiments described here would not have been remotely possible without her dedication, ideas, attention to detail, and hard work.

A great deal of thanks also goes out to all of the members of the King lab, present and past, for their help with choanoflagellate wrangling and troubleshooting. In particular, I'd like to thank Dan Richter for sharing his bioinformatics expertise, Pawel Burkhardt for his advice on antibodies and protein biochemistry, Rosie Alegado for her input on all-things-bacteria and her career counseling, and Steve Fairclough for generating the first *S. rosetta* monoxenic line (a usually thankless job). More importantly, I'd like to thank all of the King lab members for making the lab a collegial, friendly, and helpful environment to work and do science.

I would also like to acknowledge the useful suggestions from my thesis committee members, Iswar Hariharan, Mike Eisen, and John Taylor, in addition to other members of the MCB faculty who lent me their time and ideas, including Jasper Rine and Zac Cande.

Some of my most enjoyable times in grad school came from the scientific discussions and brainstorming sessions with my peers. I would thus like to thank Justin Bosch for insights into the world of genetic screens, Andrew Glazer for his help with statistics and making a linkage map, and Phil Cleves for his off-the-wall experiment ideas, to name just a few of the amazing friends I've made here. I will sorely miss you all.

Finally, a big thank you to my friends and family for believing in me, for encouraging me to go out and have fun sometimes, and for supporting me in so many ways over the past several years.

# Chapter 1

# Choanoflagellates, rosettes, and the evolution of animal development

## The origin of animal multicellularity

Across the eukaryotic tree of life, multicellularity has evolved independently at least 22 times [1]. Notable examples of independent multicellular lineages include the animals, land plants, red algae, brown algae, slime molds, and two separate groups of fungi [2]. Each of these lineages can ultimately be traced back to a unicellular, eukaryotic progenitor that evolved multicelluarity either through aggregation or failed separation following cell division [3].

At the base of the animal lineage, the origin of multicellularity was an essential pre-requisite for the evolution of the first animals. In turn, the biology of the first animals provided the foundation upon which all animal diversity later evolved. Thus, understanding the earliest steps in animal history can allow us to reconstruct animal evolution and highlight core, conserved features of animal multicellularity and development.

One challenge in reconstructing animal origins is that the first animals lived over 600 million years ago and these small, soft-bodied organisms left little mark on the fossil record [4]. Even so, we can gain insight into the genomes, cell biology, and development of the first animals through comparisons among extant species. For example, all modern animals reproduce through sperm and egg, undergo gastrulation during development, and differentiate into multiple cell types, implying that the last common ancestor (LCA) of animals also shared those features [5-9]. At the genomic level, many of the genes required for cell-cell adhesion, cell-extracellular matrix adhesion, and developmental patterning are shared among nearly all animals [10-14]. We can thus infer that the LCA of animals was multicellular and had already assembled a complex genetic and developmental toolkit. Therefore, if we wish to understand earlier steps in animal evolution, such as the transition to multicellularity, we must look to phylogenetically relevant organisms outside of animals to bracket this earlier period of evolutionary history.

## Choanoflagellates can be used to reconstruct animal origins

The closest living relatives of animals are the choanoflagellates, a clade of aquatic microeukaryotes that are found in marine and freshwater environments worldwide [15]. The phylogenetic and comparative genomic evidence for choanoflagellates as the closest relatives of animals is robust and comes from studies using a variety of organisms, phylogenetic methodologies, gene sets, and genomes [16-22] (Fig. 1.1).

Choanoflagellate cell biology also points to a close relationship with animals. In fact, choanoflagellates first captured biologists' attention in the mid-1800s, when early

microscopists noted that choanoflagellates are nearly indistinguishable from the collar cells of sponges [23, 24]. Both choanoflagellates and sponge collar cells use a single apical flagellum surrounded by a collar of actin-filled microvilli to capture prey bacteria (Fig. 1.2). This shared cellular architecture implies that the first animals may have also had collar cells, especially as phagocytic collar cells are seemingly conserved in other animal lineages such as cnidarians, echinoderms, ascidians, and hemichordates [9, 25-31], although, see [32].

As the closest sister group of animals, choanoflagellates can help to inform our understanding of the genome evolution of early animals. For instance, we can infer that any genes conserved among animals and choanoflagellates were also present in both the animal LCA and the animal-choano LCA, while genes that are restricted to animals likely evolved later. The genomes of choanoflagellates have revealed that many genes critical for animal multicellularity are conserved between choanoflagellates and animals, such as Myc, homeobox genes, receptor tyrosine kinases, cadherins, and C-type lectins [1, 9, 33-37]. However, although these genes are expressed in choanoflagellate cells [35, 38, 39] and are often biochemically similar to their animal homologs [37, 40-42], the functions of such genes within choanoflagellate physiology remains unknown. Thus, we are left asking: which aspects of animal development evolved prior to the origin of animals? And to what extent did animals evolve through the co-option of older genes to new developmental tasks? Clarifying the functions of these genes in choanoflagellates can help to answer these questions; any homologs with similar roles in animals and choanoflagellates likely served the same role in the ancestor, while homologs with divergent roles were likely co-opted to new functions in at least one clade.

## Rosette development and its implications for animal origins

Beyond the gene- and genome-centric approach, choanoflagellates may also allow us to reconstruct the cell biology and physiology of animal progenitors. Like many microeukaryotes, choanoflagellates can differentiate into a variety of cell types in response to environmental conditions [43, 44]. For example, the choanoflagellate *Salpingoeca rosetta* is able to differentiate into diverse cell types and morphologies, including rosette colonies, chain colonies, fast swimmer cells, slow swimmer cells, and thecate cells that attach to substrates (Fig. 1.2D) [1, 44]. Understanding choanoflagellate cell differentiation may be relevant for reconstructing animal origins, as mechanisms for the spatial and temporal regulation of differentiation during animal development were likely co-opted from pre-existing cell differentiation pathways in the progenitors of animals [1, 2, 7, 45, 46].

The colonial choanoflagellate morphologies are of particular interest for reconstructing animal origins. All choanoflagellates have a unicellular stage to their life history, but phylogenetically diverse choanoflagellate species can also develop facultatively into multicellular, colonial morphologies [3, 18]. In fact, current estimates of the number of choanoflagellates species that can form facultative colonies are likely underestimates, as most species have been observed under limited environmental conditions, potentially in the absence of colony-inducing cues.

Colony development has been perhaps best studied in the choanoflagellate *Salpingoeca rosetta*, where two different colony morphologies have been observed: flexible, fragile, chain colonies and robust, spherical rosette colonies (Fig. 1.2D) [4, 44, 47]. Both colonial forms appear during rapid, log phase growth and they have similar global transcriptional profiles [9, 39], but chain colonies can be distinguished from rosettes by: 1) the linear (rather than spherical) arrangements of cells, 2) the presence of filopodia penetrating extracellular matrix in the center of rosette colonies, and 3) the high sensitivity of chain colonies to mechanical sheer (e.g. by pipetting) as compared to the robust structures of the rosettes [10, 44].

Differentiation into rosette colonies is triggered by specific sulfonolipids from the bacterium *Algoriphagus machipongonensis* [15, 48]. In the presence of either live *A. machipongonensis* bacteria, conditioned media harvested from an *A. machipongonensis* culture, or purified sufonolipids, *S. rosetta* develops into clonal rosette colonies through serial cell division and incomplete cytokinesis [16-22, 47]. Successive rounds of cell division result in a sphere of cells in which all collars and flagella are oriented outward (Fig. 1.2B). Mature rosettes can also reproduce themselves through the process of rosette fission, in which a spherical rosette colony elongates and divides into two rosette colonies [23, 24, 44].

It remains unknown whether animal and choanoflagellate multicellularity are homologous, or if they resulted from two independent origins of multicellularity (Fig. 1.1). However, several observations support the idea that there may be some underlying homology. First, as colony development is spread fairly broadly across the choanoflagellate tree, simple multicellularity may have been present in the choanoflagellate LCA and, therefore, potentially in the animal-choanoflagellate LCA (Fig. 1.1). Second, both animal and choanoflagellate multicellularity are achieved through clonal, serial cell division rather than aggregation of different cells. Third, rosette colonies resemble the morula stage of early embryonic development, which is common to many animal lineages [9, 25-31]. Finally, hypothesized models for the morphology of the earliest animals, such as Haekel's Blastea and Nielsen's Choanoblastea, also bear marked similarity to rosette colonies [6, 32, 39, 49].

If rosette formation and animal multicellularity are entirely non-homologous, understanding the mechanisms of rosette development will still be informative to the overall study of multicellularity and how it can evolve from unicellular progenitors. However, if rosette formation is homologous to animal multicellularity, then understanding rosette formation in extant choanoflagellates may provide a more direct window into the events surrounding animal origins. Assessing the homology of these developmental processes will require comparisons of the molecular and genetic bases of rosette development and animal multicellularity to see if these processes use similar cellular mechanisms and/or homologous sets of genes.

## Open questions regarding choanoflagellate rosette development

### 1) Is there a link between rosette development and sexual reproduction?

Multicellular development is linked to sexual reproduction in many eukaryotes, including animals. Indeed, the single-celled bottleneck of the fertilized egg is thought to be critical for reducing genetic conflict and the limiting emergence of cheater cells in multicellular organisms [1, 9, 33-37, 45, 50-53]. But beyond the observation that multicellularity exists in choanoflagellates, it was not known if multicellularity was linked to a choanoflagellate sexual life cycle, nor if choanoflagellates possessed a sexual life cycle in the first place.

Sexual reproduction is ancient and widespread in eukaryotes, although some lineages have lost the ability to undergo meiosis and/or sexual reproduction [35, 38, 39, 54-59]. The genome of the choanoflagellate *Monosiga brevicollis* contains a full "toolkit" of genes required to undergo meiosis in other organisms [37, 40-42, 60]. It was therefore likely that *M. brevicollis* was capable of sexual reproduction under unknown conditions, but neither sex nor meiosis had previously been either observed or indirectly detected for any choanoflagellate species. In diverse microeukaryotes, sexual cycles are triggered by changes in environmental nutrient levels [43, 44, 57, 61-63], suggesting that a similar pattern might hold for choanoflagellates, but little is known about choanoflagellate ecology or what nutrients might serve as relevant signals. In addition, basic features of choanoflagellate biology, such as the ploidy or chromosome number were also unknown (although the genome sequence of *S. rosetta* tentatively suggested a minimum of 33 chromosomes in this species [39]). Thus, at the start of my thesis research, the existence of a sexual cycle in choanoflagellates and any relationship to choanoflagellate multicellularity was a complete black box.

### 2) What is the genetic basis of rosette development?

Prior work has established that rosette development can be triggered by exposing a subset of slow swimmer *S. rosetta* cells to sulfonolipids from *A. machipongonensis* [44, 48]. However, there is still much to be learned about how *S. rosetta* detects the bacterial signals, which signal transduction pathways are activated within the choanoflagellate cell, or how *S. rosetta* alters its cell division and cell biology to develop into rosette colonies as opposed to chain colonies or single cells. Transcriptome sequencing performed as part of the *S. rosetta* genome project revealed that chain colonies and rosette colonies also have remarkably similar genome-wide transcriptional profiles [39], although the experimental design in this study precluded the identification of genes that were differentially expressed between chain colonies and rosette colonies. Thus, no genes involved in rosette formation have previously been discovered. Unveiling the genetic basis of rosette development will be important because it will allow us to understand the fundamental mechanisms underlying this morphological transition, in addition to allowing for an assessment of whether choanoflagellate and animal multicellularity are homologous.

**FIGURES**

**Figure 1.1. Phylogenetic relationships between choanoflagellates and animals.** Triangles represent the species-rich clades of the fungi, animals, or choanoflagellates. To understand the events that led to the evolution of the last common ancestor (LCA) of animals (red circle), we compare animal biology with that of their closest relatives, the choanoflagellates, such that shared features between animals and choanoflagellates are presumed to have been present in the animal-choanoflagellate LCA (blue circle). Comparisons among choanoflagellates can allow for inferences about which features of choanoflagellate biology are derived in particular species versus which were present in the choanoflagellate LCA (green circle). Multicellularity in animals and choanoflagellates could have evolved independently in each clade (gray boxes) or could have a shared, ancient origin (black box).

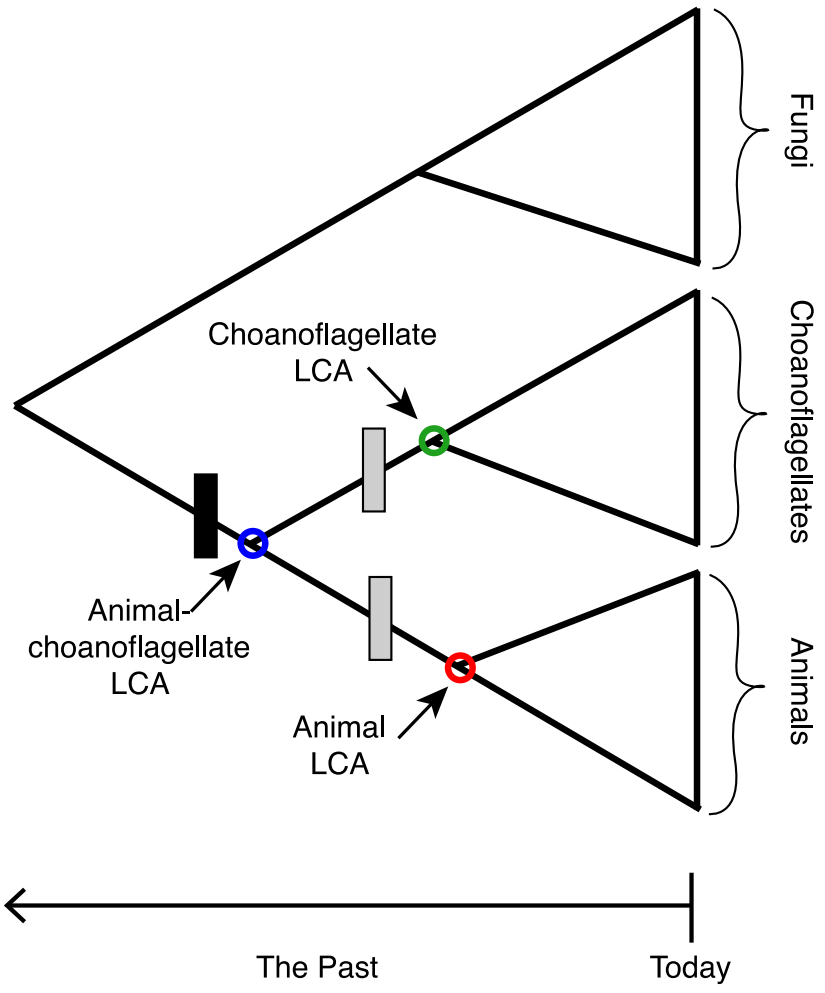
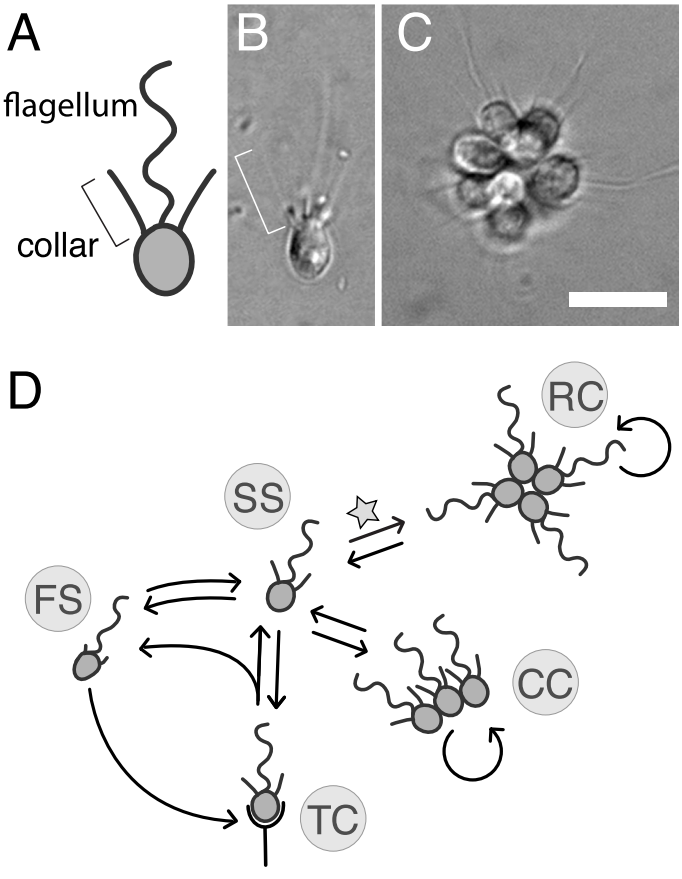


**Figure 1.2. Cellular architecture and life history of *S. rosetta*. A**) Cross-sectional diagram of an *S. rosetta* cell, with an apical flagellum encircled by a collar of microvilli (bracket). **B-C**) DIC images of *S. rosetta* as a (**B**) single cell or (**C**) rosette colony. In rosette colonies, each cell is oriented around a central point, with the flagella facing outward. Bacterial prey (~1um rods) are seen attached to the collars of some cells. Scale bar = 10μm. **D**) Life history diagram showing transitions between morphologically-differentiated cell types: rosette colonies (RC), chain colonies (CC), slow swimmers (SS), fast swimmers (FS), and thecate cells (TC). The transition from slow swimmers to rosette colonies (indicated with a star) is induced by specific lipids from the bacterium *Algoriphagus machipongonensis*. Figure modified from: Levin, T. C., Greaney, A. J., and N. King. 2014. *A genetic screen implicates a C-type lectin in the regulation of choanoflagellate multicellular development*, currently in review.



# Chapter 2

# Discovery of a choanoflagellate sexual life cycle

*The results presented here were published as part of the following paper:*

Levin, T. C. and N. King. 2013. *Evidence for sex and recombination in the choanoflagellate* Salpingoeca rosetta. Current Biology. 23 (21): 2176 – 2180.

*I performed all experiments and analyses.*

## Summary and Background

Nearly all animals reproduce sexually through the production and fusion of sperm and egg cells, yet little is known about the ancestry of animal sexual reproduction. Moreover, the sexual cycle of the closest living relatives of animals, the choanoflagellates [17, 38], remains completely unknown. The choanoflagellate *Monosiga brevicollis*, possesses a "meiotic toolkit" of genes [60], but the lack of polymorphisms detected during genome sequencing precluded inferences about its ploidy or sexual cycle [38]. Here we report that a related choanoflagellate, *Salpingoeca rosetta* [44, 47], has a sexual life cycle and transitions between haploid and diploid states. Haploid cultures of *S. rosetta* became diploid in response to nutrient limitation. This ploidy shift coincided with anisogamous mating, during which small, flagellated cells fused with larger flagellated cells. Distributions of polymorphisms in laboratory strains of *S. rosetta* provided independent evidence of historical recombination and mating. The ability of *S. rosetta* to produce morphologically differentiated gametes and to engage in sexual reproduction has implications both for reconstructing the evolution of sex in the progenitors of animals and for establishing classical genetics in choanoflagellates.

## Materials and Methods

### Growth media

Unenriched sea water was prepared by adding 32.9 g Tropic Marin sea salts (Wartenberg, Germany) to 1 L water to a salinity of 32 – 37 parts per thousand. Cereal grass (CG) medium was made by infusing unenriched sea water with cereal grass pellets (Basic Science Supplies, Rochester NY) [64]. High nutrient (HN) medium (250 mg/L peptone, 150 mg/L yeast extract, 150 µL/L glycerol in unenriched sea water) is made by diluting Sea Water Complete Medium [65] to 5% (vol/vol) in unenriched sea water.

## Px1 and the *S. rosetta* reference genome sequence

All strains of *S. rosetta* currently in culture were derived from an environmental isolate that was started from a single, rosette colony (ATCC 50818) [44]. Px1 (ATCC PRA-366) was derived from the environmental isolate through a combination of antibiotic treatments and serial dilution, and includes a single feeder bacterium, *Algoriphagus machipongonensis* [39, 44, 66]. Px1 has been sequenced and its genome is publicly available [39].

At the time that the genome was sequenced, the ploidy of *S. rosetta* and the ability of this organism to undergo sexual reproduction in the lab were unknown. However, although we do not have information about the ploidy of the culture that was sequenced, we can reconstruct its ploidy in retrospect based on the data presented here.

Given that isolates A, B, and C were all clonally derived from Px1, we infer that Px1 was a mixed population of haploid and heterozygous diploid cells. This fact initially seemed paradoxical, as Px1 was itself clonally derived from the environmental isolate. However, our finding that rapidly passaged isolate B cells became largely haploid (Fig. 2.3E) suggests the following scenario: the single cell that founded Px1 was a heterozygous diploid, and during the rapid passaging to expand the culture, a subset of the clonal population underwent meiosis. The resulting culture was no longer clonal, and was instead a mixture of heterozygous diploid cells (similar to isolate A) and the haploid, meiotic progeny of these cells. We note that our analysis of the ploidy of a Px1 culture yielded a similar mixture of haploid and diploid cells (Fig. 2.1B).

During preparation of the cells for genome sequencing, the mixed-ploidy Px1 culture was passaged extremely rapidly to propagate a large number of cells. Based on our finding that rapid passaging induces *S. rosetta* cultures to become more haploid, we infer that at the time of harvesting, the sequenced culture was dominated by haploids. As *S. rosetta* grows exponentially in high nutrient conditions, if a single haploid clone was slightly more prevalent than its sister cells early on, then a single genotype could later predominate the culture and the resulting genome sequence. Any polymorphisms that might have existed in the non-dominant haploids and diploids in the culture would have represented a small proportion of reads, and may have been dismissed as sequencing error. Supporting this scenario, upon reexamination of SNP data generated during the genome sequencing project [39], we in fact detected some of the same segregating polymorphisms shown in Fig. 2.4A.

## Establishment of *S. rosetta* isolates

Isolates A, B, and C were derived from Px1 through multiple clonal isolation steps (Fig. 2.1A). To generate isolates A and B, a culture of Px1was passaged in CG medium every 3-4 days for approximately 9 months. Using an approximation of 1 generation per 10 hours (derived from the growth curve in [47]), we estimate that this corresponds to 500-1000 generations. Subsequently, this long-term culture was mutagenized with EMS prior to clonal isolation. A sample of $5 \times 10^5$ cells/mL sea water was exposed to 0.3% (vol/vol) EMS (ethyl methanesulfonate, Sigma) at room temperature for 1hr. Following mutagenesis, the

EMS was deactivated and diluted by washing the cells 3 times in 5% (wt/vol) sodium thiosulfate (Sigma) in unenriched sea water. Each wash included a spin at 1000 x g for 10min to pellet the cells. Cells were resuspended in 2 mL 10% (vol/vol) CG medium diluted in unenriched sea water and incubated at room temperature for 24 hours to recover. All of the above steps included a control sample that experienced the same washes and incubations, but was not exposed to EMS. Isolate A was derived from the control sample and isolate B was derived from the mutagenized sample. Clonal isolation was achieved using 2 serial dilution-to-extinction steps, similar to [67]; cells were diluted to approximately 1 cell per well in 10% (vol/vol) CG medium in sea water and aliquoted into 96-well plates. After 5-7 days, plates were examined visually to find isolated clones, which were then expanded and frozen [68].

By counting the number of empty wells in the 96-well plates, we can estimate the probability of clonal isolation using the Poisson distribution. Estimates of lambda may be obtained by –ln (proportion of wells with no cells), and the probability of obtaining a clonal isolate is calculated as the (probability of getting exactly 1 cell per well)/(the probability of getting 1 or more cells per well). Using this method, we estimate that there is a >99% probability that isolates A and B each underwent a single cell bottleneck at least once during the two clonal isolation steps.

In isolate C, *A. machipongonensis* was replaced with a new feeder species, *Echinicola pacifica* (DSM 19836) [69] through repeated antibiotic treatment and limiting dilution. As *E. pacifica* and *A. machipongonensis* have very similar antibiotic sensitivities, we first generated a strain of erythromycin-resistant *E. pacifica*. Briefly, *E. pacifica* cultures were mixed with *E. coli* carrying the pNJR5 vector [70] and allowed to mate for 48 hours on polycarbonate mating filters (Millipore; 25 mm diameter, 0.45 μm pore). Bacteria were then plated on Sea water Complete Media plates [65] with 10 μg/mL erythromycin and 200 μg/ml gentamycin to select for transformed *E. pacifica*. An individual *E. pacifica* colony was isolated and retested to confirm erythromycin resistance. Meanwhile, a culture of Px1 *S. rosetta* with *A. machipongonensis* was grown for 5 days in unenriched sea water with 20 μg/mL erythromycin to reduce the amount of bacteria in the culture. A single colony of erythromycin-resistant *E. pacifica* was added to this culture, and choanoflagellates were isolated by limiting dilution as above. These antibiotic treatment and limiting dilution procedures were repeated until *E. pacifica* was the only remaining bacterial species in the culture, as monitored by plating and 16S RFLP analysis [44]. After this time, antibiotic treatment was discontinued. There were a total of 5 serial clonal isolation steps, yielding a >99.9% probability that isolate C experienced a clonal bottleneck. This strain is publicly available as ATCC PRA-390 under the name SrEpac, for *S. rosetta* grown with *E. pacifica* bacteria.

**Culturing conditions**

Isolate C was propagated in HN medium, yielding primarily haploid cultures of chains and slow swimmer cells [44]. Px1 and isolates A and B were propagated in CG

medium, which yielded primarily diploid cultures of thecate cells. All cultures were propagated by diluting 1 mL of culture into 9 mL fresh medium every 2-3 days.

To induce sex under nutrient poor conditions, a stationary phase culture of isolate C was pelleted and resuspended in an equivalent volume of unenriched sea water. These conditions did not provide nutrients to support bacterial growth, and therefore the choanoflagellates grazed down the bacteria and themselves became starved within 3-6 days.

To return diploid isolate C cells to a haploid state, 1 mL of isolate C culture in unenriched sea water (prepared as above) was added to 9 mL HN medium. Ploidy was assessed 3 days later. To induce isolate B cells to become haploid, 2 mL isolate B culture was added to 8 mL CG medium daily for 3 weeks.

**Quantification of DNA content**

We used flow cytometry to measure the DNA content of isolated, propidium iodide-stained *S. rosetta* nuclei. Previously, this type of analysis was not practical in choanoflagellates, as the large quantity of DNA from the co-cultured bacterial prey obscured the signal from choanoflagellate DNA. Therefore, we developed a Percoll density gradient technique (similar to [71]) to separate and eliminate a large fraction of the bacteria from the culture. Three to four million choanoflagellate cells were concentrated to 250 µL and loaded on top of 2 mL Percoll/sorbitol/sea water mixture [500 mM sorbitol, 50 mM Tris pH 8, 15 mM $MgCl_2$, 0.4% (vol/vol) artificial sea water, 10% (vol/vol) Percoll, PVP-coated (GE Healthcare Biosciences)] in a 15mL conical tube. This solution was centrifuged at 1000 x g for 10 minutes on a low brake setting in an accuSpin 1R centrifuge (Fisher) to allow for density-based equilibration. Choanoflagellates were retained in the pellet, while the bacteria-enriched supernatant was discarded. Cells were then washed in unenriched sea water.

While the density gradient greatly reduced the bacteria present in the water column, it did not eliminate the bacteria that were trapped on or inside of choanoflagellates that had been feeding. To physically separate *S. rosetta* nuclei from their prey bacteria, we used a gentle lysis protocol (similar to [72]) and only measured the DNA in intact *S. rosetta* nuclei. Briefly, pelleted *S. rosetta* cells were resuspended in a low-salt buffer (50mM Tris pH 8, 140 mM NaCl, 1.5 mM $MgCl_2$) to induce swelling and mild osmotic shock. *S. rosetta* cells were spun down again and the pellet was resuspended in ice-cold low-salt buffer with NP-40 and RNAse A [50 mM Tris pH 8, 140 mM NaCl, 1.5mM $MgCl_2$, 0.9 mg/mL RNAse A (Qiagen), 0.25% (vol/vol) NP-40]. After a 5 minute incubation at 4 °C, the cell lysate was passed through a 40 µm filter and stained with 20 µg/ml propidium iodide (Sigma). Samples were analyzed on a Beckman-Coulter FC-500 cytometer, with gating based on forward and side scatter to differentiate stained *S. rosetta* nuclei from contaminating bacteria and the yeast internal standard (see below and Fig. 2.2). The propidium iodide fluorescence was used as the trigger signal, and approximately 50,000 events were measured per sample. Doublet discrimination was used to eliminate aggregates of nuclei

from the analysis [73]. Flow cytometry data were analyzed using FlowJo from Tree Star Inc. (Ashland, OR).

To enable comparison of the DNA content among different *S. rosetta* cell lines, we used a sample of haploid DDY904 *Saccharomyces cerevisiae* [74] as an internal standard in all experiments. A single yeast colony was inoculated into 2 mL YPD medium and grown shaking overnight at 200 rpm at room temperature. Yeast cells were prepared similar to [75], by fixation in 70% ethanol, treatment with 9 mg/mL RNAse A (Qiagen) at 37 °C for 2 hours, treatment with 0.5 mg/mL proteinase K (Fisher) at 50C for 1 hour, and resuspension in 50mM Tris pH 7.5. A small sample of 2 to 5 µL of prepared yeast was spiked into each *S. rosetta* lysate sample. This internal standard allowed us to verify which peaks in different *S. rosetta* samples represented nuclei with equivalent DNA content. We gained additional information about *S. rosetta* ploidy by comparing the sizes of the yeast and *S. rosetta* peaks to each other. The *S. rosetta* '1n' peaks were 3 to 5.4 times larger than the yeast internal standard peaks, corresponding to the fact that the *S. rosetta* genome is approximately 4.4 times larger than the *Saccharomyces cerevisiae* genome and suggesting that the presumed '1n' *S. rosetta* peaks were indeed haploid. Differences in the fixation and staining methods for the yeast and *S. rosetta* can introduce variability into propidium iodide staining of these different species, but even so the size of the yeast peaks corresponds better to the '1n' peak being from haploid *S. rosetta* (4.4 times larger) than to a diploid *S. rosetta* (8.8 times larger). We therefore conclude that *S. rosetta* alternates between haploid and diploid states as opposed to diploid and tetraploid states. Finally, as an additional control, we also processed a sample of isolate A in all experiments, as isolate A was consistently diploid under standard growth conditions.

**Estimating the percentage of haploid and diploid cells in mixed cultures**

We used the relative sizes of the 1n, 2n, and 4n peaks to estimate the percentage of cells in the culture that were diploid. As the size of the 2n peak is the sum of the haploid G2 peak and the diploid G1 peak, it is not possible to use the height of the 2n peak as a proxy for the number of diploids. Instead, we used the following approach to approximate the number of diploid and haploid cells. First, we assumed that at a given time in an unsynchronized culture, the haploid and diploid cells would have a similar proportion of cells in the G1 and G2 phases of the cell cycle. Thus:

Proportion of cells that are haploid = $x + ax$

Proportion of cells that are diploid = $y + ay$

where x is the proportion of G1 haploids, y is the proportion of G1 diploids, and a is the ratio of cells that are in G1/G2. From the flow cytometry data, we can measure the proportion of cells in the 1n peak (= x), the 2n peak (= ax + y), and the 4n peak (= ay), and solve for a using the quadratic equation. This gives two solutions; a large G1/G2 ratio (with 2n consisting mostly of diploids in G1) or a small G1/G2 ratio (with 2n consisting mostly of haploids in G2). Cultures that have spent several days in unenriched seawater are not

rapidly dividing, and so we use the solution with a large G1/G2 ratio to estimate the proportion of cells that are diploid.

## Genomic DNA sequencing and SNP analysis

We sequenced the genomes of isolates A, B, and C to detect polymorphisms. For each isolate, *S. rosetta* genomic DNA was prepared by phenol chloroform extraction and physically separated from bacterial DNA on a CsCl gradient [68]. Samples were sequenced with 100 bp Illumina paired end reads to a median read depth of 186x, 161x, and 37x respectively for isolates A, B, and C.

Raw reads were trimmed using TrimmomaticPE [76] to remove Illumina adapters and low quality base calls. Trimmed reads were mapped to the *S. rosetta* reference genome [39] using BWA [77]. After removing duplicates using PicardTools (http://picard.sourceforge.net), variants were called using samtools and bcftools with default parameters [78]. All subsequent analyses used only high-quality SNPs with a quality score > 100. The frequency of SNP-containing reads was calculated as the number of reads at a locus that contained the SNP divided by the total number of reads at the position. To generate the map of SNPs across the genome, we binned SNPs into 20 kb sliding windows and plotted their positions semi-transparently, so that higher concentrations of SNPs were displayed as a darker color.

## Determination of SNP linkage

To investigate the linkage between SNPs, we cloned and sequenced a 2.3 kb region of supercontig 8 from isolates A, B, and C. We used PCR to amplify the region with a mixture of Taq (New England Biosciences) and Pfu (Finnzymes) polymerases and these primers: TGGAACAGTGCAGTCCTGTG and AGCTCAAACAGGTACGGACA. The region was cloned into the TOPO TA vector (Invitrogen) and transformed into TOP10 cells (Invitrogen). Multiple clones for each isolate were analyzed by Sanger sequencing to determine the linkage among SNPs. This region of supercontig 8 contains long stretches of GT dinucleotide repeats that hindered Sanger sequencing, but this problem was ameliorated by using elevated concentrations of dGTP in the Sanger sequencing reaction. However, the isolate A copy 2 clone contains an expansion of a GT repeat region that we were unable to sequence across fully; this region was excluded from Fig. 2.4B, but the full alignment (made with FSA [79] and Jalview [80]) is shown in Fig. 2.6.

## Immunofluorescence and time lapse video microscopy

To generate time lapse movies, isolate C cells were cultivated in unenriched sea water for 24 hours before transfer to a 35 mm Fluorodish glass bottomed dish (World Precision Instruments, Sarasota FL). After allowing 5-10 minutes for cells to settle to the

bottom of the dish, multiple fields of view was imaged once per minute with 40x phase contrast microscopy for a period of 10 hours. Movies were prepared from still images using the Animation setting in ImageJ 1.43 [81] at frame rate of 7 frames per second, which corresponds to a speed of 420x real time. For immunofluorescence staining, cells cultured under the same conditions were allowed to settle live on poly-L-lysine coated coverslips (BD Biosciences) for 30 minutes. After a 6% acetone/4% formaldehyde fixation, cells were stained as in [82] with E7 anti-ß-tubulin primary antibody (Developmental Studies Hybridoma Bank, Iowa City, IA), Alexa Fluor 488 goat anti-rabbit IgG (H+L) secondary antibody (Molecular Probes, Carlsbad, CA), 6 U/mL rhodamine phalloidin (Molecular Probes, Carlsbad, CA), and 0.5 µg/mL DAPI (Invitrogen) before mounting in 4 µL ProLong Gold antifade reagent with DAPI (Molecular Probes, Carlsbad, CA). All images were taken using a Leica DMI6000 B inverted compound microscope and Leica DFC350 FX camera. Immunofluorescence images were obtained using a 100x oil immersion objective.

## Results and Discussion

### Discovery of cryptic differences in ploidy among laboratory cultures of *S. rosetta*

We investigated the ploidy of a laboratory culture of *S. rosetta* called Px1 [39, 48] and three derived laboratory strains, isolates A – C (Fig. 2.1A), using flow cytometry on propidium iodide-stained *S. rosetta* nuclei [83]. *S. rosetta* isolates A and B were diploid (Fig. 2.1C and D, Fig. 2.2), with 2n and 4n peaks representing cells in the G1 and G2 stages of the cell cycle, respectively. In contrast, isolate C was haploid (Fig. 2.1E), and Px1 produced a mix of diploid- and haploid-sized peaks, in addition to a small 3n-sized peak (Fig. 2.1B). Because each of the tested isolates was derived from Px1 (Fig. 2.1A), we inferred that *S. rosetta* might be capable of switching ploidy under laboratory conditions, perhaps as part of a sexual life cycle. Nonetheless, after three months under standard culturing conditions, isolate C remained haploid and isolates A and B remained diploid (Fig. 2.3).

### Altered nutrient availability triggers changes in *S. rosetta* ploidy

Because sex and meiosis in many unicellular eukaryotes are triggered by changes in nutrient availability [57, 61-63], we next tested the influence of different culture media on *S. rosetta* ploidy. A haploid isolate C culture that had been maintained in high-nutrient (HN) media was divided in two, with one half continuing in HN media while the other half was transferred to unenriched sea water. After 6 days, isolate C cells grown in HN media remained haploid (Fig. 2.1F), whereas approximately 89% of isolate C cells cultured in unenriched sea water were diploid (Fig. 2.1G). The change in ploidy was reversible; isolate C diploid cells transferred from unenriched sea water back into HN media returned to a haploid state within 3 days (Fig. 2.1H, Experimental Procedures).

Cultures of *S. rosetta* fed only *A. machipongonensis* bacteria (i.e. isolate A, B, and Px1) grew poorly in HN media for unknown reasons. Still, we could induce isolate B to be largely haploid after daily passaging in cereal grass (CG) medium for 3 weeks, which provided a continual replenishment of nutrients in the culture (Fig. 2.3E). We hypothesize that the largely stable differences in ploidy between isolates A, B, and C were due to the relative levels of media richness and/or nutrient availability in the different cultures. Thus, we infer that nutrient limitation induces *S. rosetta* haploid populations to become diploid, perhaps through mating, and nutrient rich conditions promote a transition from diploidy to haploidy, potentially through meiosis.

## Patterns of polymorphism reveal a history of sex and recombination

If *S. rosetta* has a sexual cycle, we would expect to find signatures of historical, meiotic recombination throughout the genome. In the absence of recombination, single nucleotide polymorphisms (SNPs) are inherited in haplotype blocks that span the length of each chromosome, whereas recombination can break up haplotype blocks into smaller chromosomal regions [56, 84-86]. We found that SNPs in isolates A, B, and C were broken up into discrete haplotype blocks on 33 of the 40 largest supercontigs (which range in size from 0.4 to 2.6 Mb; Fig. 2.4A). The presence of haplotype blocks on the majority of large supercontigs suggested a history of genome-wide recombination generated during meiosis. The sharp boundaries at the edges of the haplotype blocks marked the chromosomal regions where we infer there has been genetic exchange between homologous chromosomes. Most SNPs were also shared among the isolates (Fig. 2.4A and B), suggesting that these polymorphisms are segregating in laboratory populations of *S. rosetta*. Furthermore, by interrogating the physical linkage of SNPs in these strains, we observed direct evidence of recent recombination in *S. rosetta* that has resulted in novel combinations of the segregating alleles (Fig. 2.4B, Fig 2.5).

In addition to offering evidence for historical recombination, the segregating polymorphisms in diploid isolates A and B yielded insights into *S. rosetta* sex. We found that 84% of the SNPs in isolate B were homozygous, suggesting that this strain became diploid either by self-fertilization or through auto-diploidization (Fig. 2.4A and B, Fig 2.6). In isolate A, however, 93% of the SNPs were heterozygous and this strain had non-identical homologous chromosomes, suggesting that isolate A arose through the mating (i.e. fusion) of two genetically distinct individuals. Thus, we propose that *S. rosetta* isolates A and B have experienced a history of sex and meiosis, which involved the generation of diploids (either outcrossed or selfed), recombination between genetically distinct chromosomes, and segregation of recombinant chromosomes into different cells.

## Morphological evidence for sex during haploid-to-diploid transitions

To investigate the cell morphology and behavior of putative *S. rosetta* gametes, we monitored cultures of *S. rosetta* during the haploid-to-diploid transition. Twenty-four hours after transferring haploid isolate C cells into unenriched sea water to induce the switch to

diploidy, we monitored the culture using time lapse microscopy for 10 hours. We observed approximately 1,200 total cells and recorded 6 cell fusion events, indicative of mating (Fig. 2.7). All mating events involved a small, rounded, uniflagellated cell (provisionally termed the 'male gamete') fusing with and entering a larger, ovoid, uniflagellated cell (the 'female gamete'). For approximately 10 to 30 minutes prior to fusion, the male gamete cell body contacted the microvillar collar of the female gamete, possibly as part of a recognition process. We also observed many examples of adhesion between cells that did not result in fusion, and adherent clumps of cells often persisted over a period of minutes to hours. Successful cell fusions were always initiated by contact between the basal end of the male gamete (opposite the flagellum) and the base of the collar of the female gamete. Each cell fusion event we observed lasted for 5 to 8 minutes and resulted in a slightly larger, more rounded cell.

Immunofluorescence microscopy of cultures undergoing the haploid-to-diploid transition provided further support for a sexual cycle in *S. rosetta* (Fig. 2.8). Isolate C cells grown rapidly in HN media, tended to have a rounded cell body and a doughnut-shaped nucleoplasm. In contrast, isolate C cells in unenriched sea water were often found in disorganized, adherent clumps consisting of both male and female gamete types (Fig. 2.8B). Contacts between adherent cells were through actin-based microvillar collars and/or filopodia. Strikingly, we also observed a bridge of DNA between the nuclei of adherent male and female gametes (Fig. 2.8B), consistent with a process of genetic transfer.

Choanoflagellates feed by phagocytosing bacterial prey at the base of their collars, but several pieces of evidence indicate that *S. rosetta* cell fusion is sex rather than cannibalism. First, the male gamete appears to take on a consistent orientation prior to fusion, which would not be expected if it were merely prey. Second, starvation induced changes in nuclear morphology that were suggestive of genetic transfer (Fig. 2.8B). Third, fusion was preceded by a long period of cell contact, and only a minority of male-female encounters resulted in cell fusion. This may indicate that signaling between the gametes or another form of recognition might be required for cell fusion, rather than indiscriminate engulfment. Finally, cell fusion occurred as haploid populations were becoming diploid, as would be expected in the case of sex.

## Conclusions and implications

We conclude that the choanoflagellate *S. rosetta* has a sexual phase to its life cycle. Haploid-to-diploid transitions (i.e. sex) occurred in nutrient limiting conditions, while diploid-to-haploid transitions (presumably meiosis) occurred in nutrient rich conditions. Although this sexual response is the reverse of the meiotic response of *S. cerevisiae* to nutrient limitation [62], diverse other unicellular eukaryotes respond to nutrient limitation through sexual differentiation and fusion [61, 63]. During the *S. rosetta* haploid-to-diploid transition, we observed the fusion of previously undocumented, morphologically differentiated cell types that we infer are anisogamous gametes. As both cell fusion and the

haploid-to-diploid transitions occurred within clonal cultures (without mixing different isolates together), we conclude that *S. rosetta* is homothallic (i.e. that it can self-fertilize). This is consistent with the finding that diploid isolate B has genome-wide homozygosity. However, diploid isolate A has genome-wide heterozygosity, implying that *S. rosetta* can also undergo outcrossed mating, which may be harnessed in the future for crosses and genetics in choanoflagellates.

One implication of the homothallism of *S. rosetta* is that a single genotype can produce both male and female gametes. This suggests that mating type is not permanent and instead that a sort of mating type switching may occur, possibly through sex chromosome loss or epigenetic mechanisms. Understanding how *S. rosetta* gamete identity is specified may help to illuminate the ancestry of animal gametogenesis.

Although sexual systems have diversified extensively within the fungal clade [55], sex in animals has remained relatively consistent [87] suggesting that choanoflagellates could shed light on ancestral sexual systems in Opisthokonta, the clade that includes animals, choanoflagellates, and fungi. Morphologically, the fusion of flagellated, anisogamous gametes in choanoflagellates bears a striking resemblance to the flagellated, anisogamous gametes of *Allomyces macrogynus*, a basal-branching fungal species [88, 89]. In addition to the morphological similarities of the gametes (uniflagellated female cells that are 2-4 times larger than the male cells), both species also appear to have a biflagellated intermediate that forms after cell fusion, with nuclear fusion occurring while the cell retains multiple flagella [88] (Fig. 2.8 C and D). If these features are shared among the gametes of other Opisthokont groups, the last common ancestor of fungi and animals may have had sex that resembled mating in *S. rosetta*. Finally, the discovery of sex in choanoflagellates represents a previously unreported mode of cell-cell recognition and adhesion in the closest relatives of animals. Mechanisms used for gamete recognition and fusion in the ancestors of animals may have been co-opted for somatic cell adhesion during the evolution of animal multicellularity [1, 34, 90].

**FIGURES**

**Figure 2.1. *S. rosetta* cultures transition between haploid and diploid states in response to changing nutrient availability. A)** The *S. rosetta* cultures used in this study were all derived from a Px1 culture [39] through clonal isolation (*), during which one cell was isolated and propagated to establish the new culture line. For isolate A, a Px1 culture was propagated for 500-1000 generations before the new culture was established through clonal isolation. For isolate B, clonal isolation occurred after EMS mutagenesis (+). For isolate C, clonal isolation occurred after replacing the feeder bacteria from Px1 with a new species, *Echinicola pacifica* (~). During preparation for genome sequencing, Px1 was passaged rapidly, which may have resulted in changes in ploidy (see Supplemental Experimental Methods). **(B– E)** *S. rosetta* laboratory cultures can be haploid, diploid, or have a mix of haploid and diploid cells, as revealed by flow cytometry of propidium iodide-stained nuclei from unsynchronized cultures of Px1 **(B)**, isolate A **(C)**, isolate B **(D)**, and isolate C **(E)**. The expected locations of the 1n, 2n, and 4n peaks were based on an internal standard of *S. cerevisiae* (Fig. 2.2). **(F – H)** Nutrient limitation induces haploid cultures to become diploid. Flow cytometry showed that isolate C was haploid after growth in HN medium for six days **(F)**. Following growth of isolate C in unenriched sea water for six days, most cells were diploid **(G)**. Transferring diploid isolate C cells back to HN medium for three days resulted in a return to the haploid state **(H)**. Haploid-to-diploid transitions under nutrient limitation occurred within clonal cultures and did not require mixing of multiple clones.

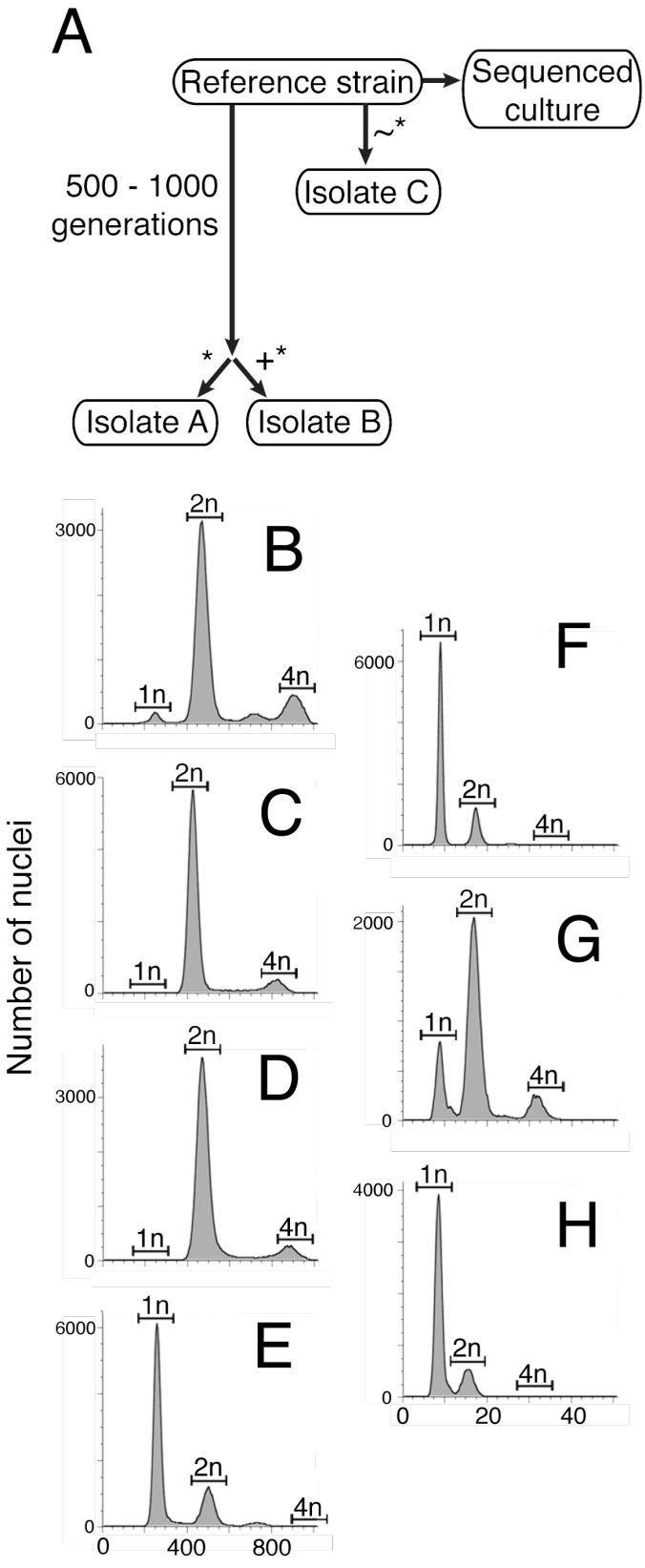

A
Reference strain
Sequenced culture
~*
Isolate C
500 - 1000 generations
*
+*
Isolate A
Isolate B
B
C
D
E
F
G
H
1n
2n
4n
Number of nuclei
Propidium iodide intensity (arbitrary units)

**Figure 2.2. Flow cytometry analysis of *S. rosetta* ploidy.** Flow cytometry analysis to differentiate propidium iodide signals from *S. rosetta* nuclei, co-cultured bacteria, and yeast internal standard. *S. rosetta* samples were analyzed before (not shown) and (**A**) after adding a yeast internal standard. This before-and-after analysis allowed us to clearly identify the yeast signal and to draw the gates separating *S. rosetta* and yeast populations, based on forward and side scatter. As forward scatter separates particles roughly based on size, the small, isolated choanoflagellate nuclei are to the left of the larger, fixed, intact yeast cells. **B)** Histogram of propidium iodide intensities from the yeast internal standard, gated in (**A**). These peaks provided the internal standard, allowing for comparison of DNA content among *S. rosetta* samples. **C)** Clumps of nuclei can yield aberrantly large propidium iodide signals, but these clumps can be eliminated by doublet discrimination. To eliminate clumps of nuclei, we plotted peak height by peak area in the propidium iodide channel. Data points off the diagonal represent clumps of nuclei, and these were excluded from the analysis. Bacterial signals may be discriminated from choanoflagellate signals due to their small DNA content, so we also eliminated the residual bacterial signal in the lower left corner during this step. **D)** Final histogram of propidium iodide intensities from *S. rosetta* nuclei, after eliminating bacteria and clumps of nuclei in (**C**). Horizontal lines separate groups of samples that were run on different days, with slightly adjusted voltages and gates on each day. In particular, after the first four samples were obtained, the forward scatter parameter was later adjusted to bring the yeast data points off axis.

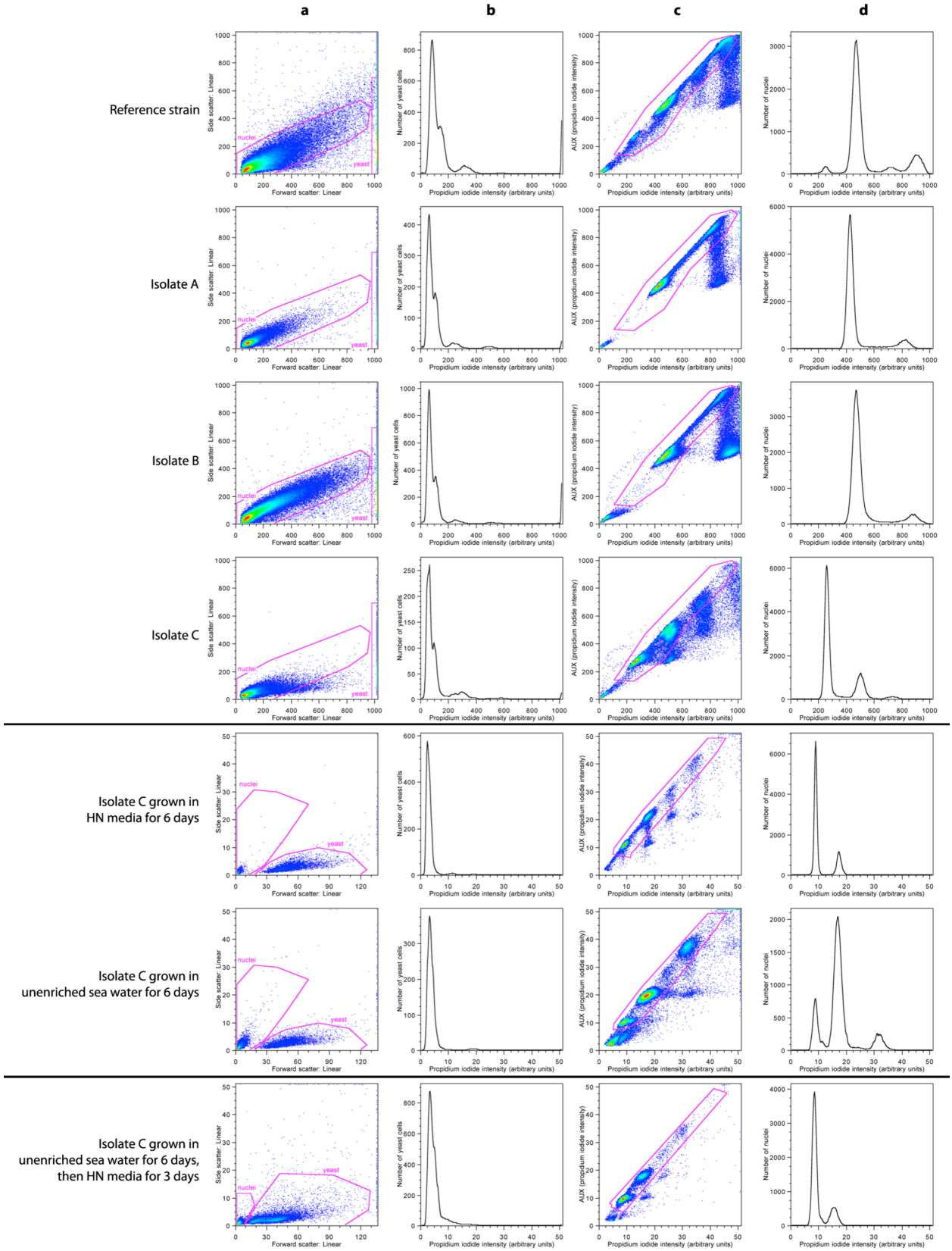

a
b
c
d
Reference strain
Isolate A
Isolate B
Isolate C
Isolate C grown in HN media for 6 days
Isolate C grown in unenriched sea water for 6 days
Isolate C grown in unenriched sea water for 6 days, then HN media for 3 days
Side scatter: Linear
Forward scatter: Linear
nuclei
yeast
Number of yeast cells
Propidium iodide intensity (arbitrary units)
AUX (propidium iodide intensity)
Number of nuclei

**Figure 2.3. Stability of *S. rosetta* ploidy.** Splitting cultures into fresh growth media every 3 days for 3 months did not induce changes in ploidy in isolate A (**A**), isolate B (**B**), isolate C (**C**) or Px1 (**D**). Fig. 2.1 B – D shows isolates A – C and Px1 three months earlier than the samples analyzed here. Isolate C was propagated in HN medium while the other cultures were propagated in CG medium. **E**) An isolate B culture propagated daily in CG medium for 3 weeks had at least as many haploid cells as diploid cells, depending on the relative abundance of cells in G1 vs. G2.

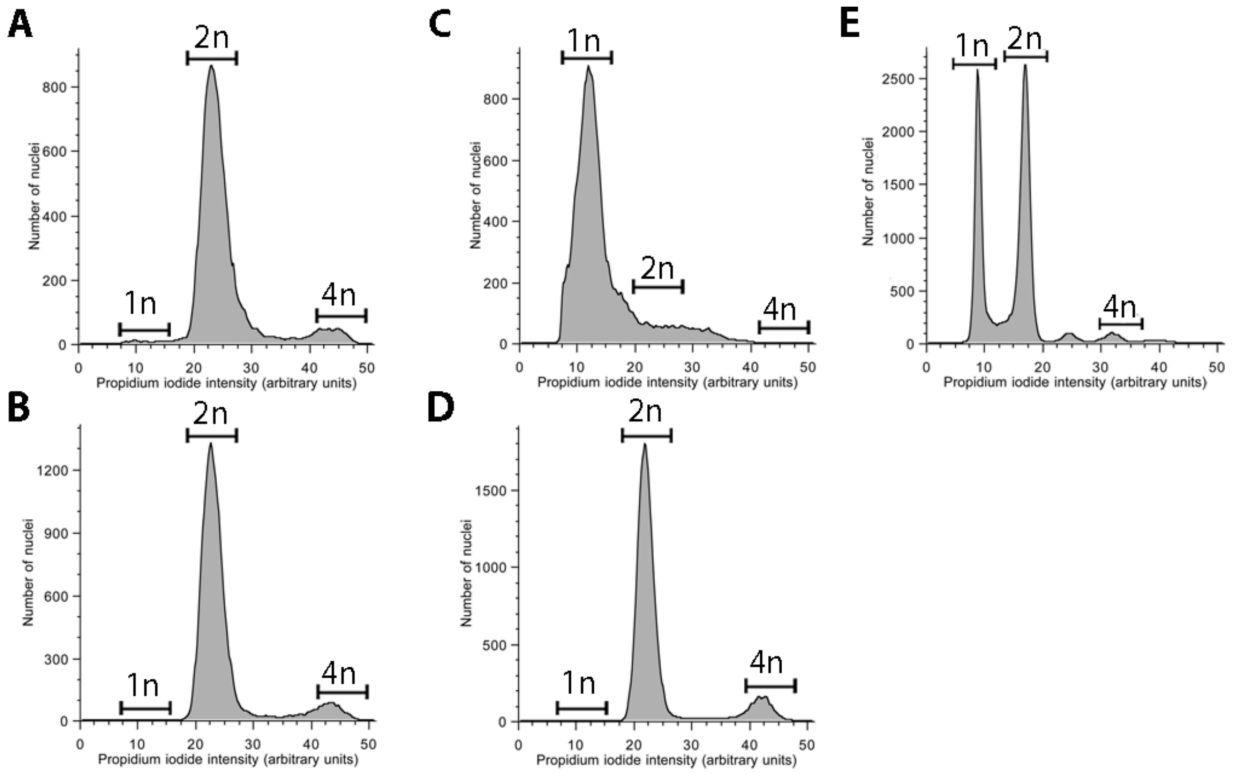


**Figure 2.4. Patterns of polymorphism reveal a history of sex and recombination in *S. rosetta*. A)** Mapping of SNP positions in isolates A, B, and C on the ten largest supercontigs suggests a history of recombination in *S. rosetta*. Sites that match the reference genome sequence are indicated in grey, homozygous SNPs in red and heterozygous SNPs in blue. Most SNPs are concentrated in large, contiguous haplotype blocks with sharp cut-offs that mark inferred sites of historical recombination. **B)** SNPs in segregating haplotype blocks are shared among isolates. The boxed region of supercontig 8 indicated in **(A)** was cloned and sequenced from isolates A, B, and C to reveal the physical linkage among SNPs (see Fig. 2.6 for full alignment). Two distinct sequences were obtained from the heterozygous diploid isolate A; copy 1 (grey) matched isolate C and the reference genome, whereas copy 2 (black) contained many SNPs relative to the reference genome. In isolate B, the 5' end of the region shares the copy 2 (black) haplotype, while the 3' end is more similar to the reference sequence (grey). The transition from black to grey in isolate B indicates the inferred region of recombination. Isolate B also has a single SNP (*) that was not detected in any other sequenced sample and likely arose within isolate B.

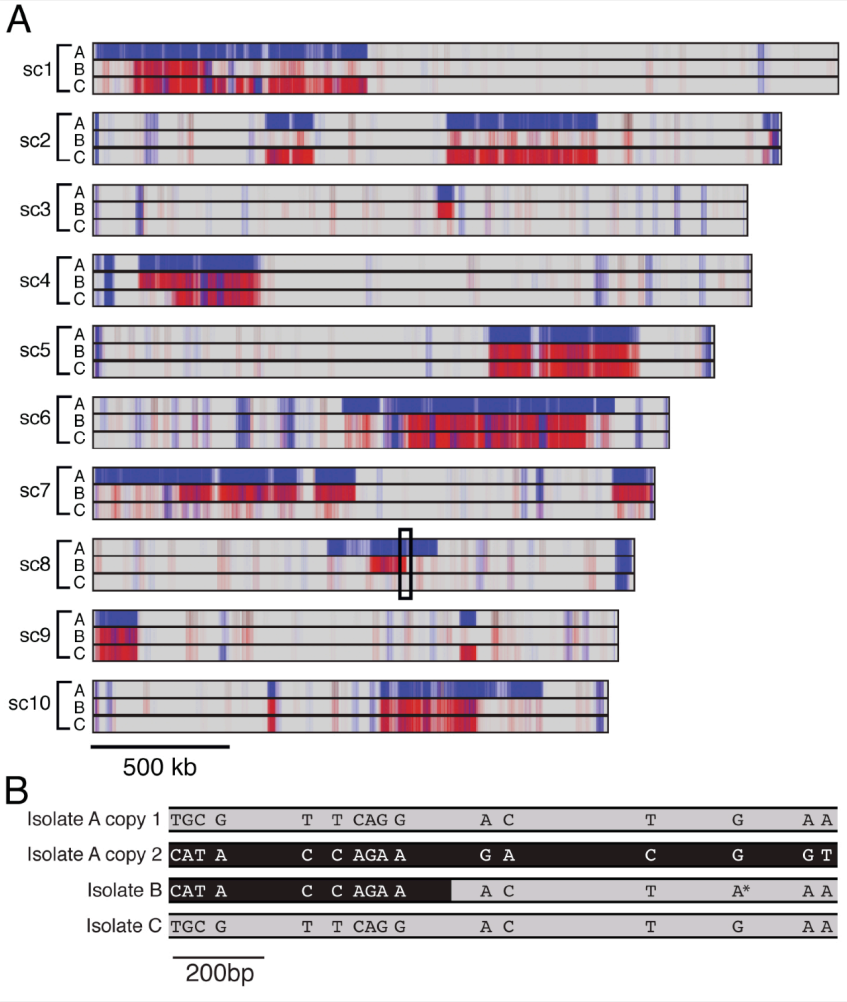


**Figure 2.5. Homozygosity and heterozygosity of sequenced isolates.** Because all of the sequenced isolates were clonally derived, we expect there to be only one or two copies of each homologous chromosome in each population. To determine for each isolate whether it was heterozygous or homozygous, we analyzed the frequency of alleles in SNP-containing reads generated by short-read genome sequencing. For heterozygous loci in a diploid organism, we expect on average 50% of the reads to contain a SNP (with the other 50% carrying the same allele as the reference sequence), whereas homozygous loci (either in haploids or in diploids with the same allele on both homologous chromosomes) should have a SNP in 100% of the reads. The majority of SNPs in isolate A were heterozygous whereas the majority of SNPs in isolates B and C were homozygous. As isolate C is haploid, we expected to detect only homozygous SNPs, yet there were a small number of SNPs called as heterozygous. Read mis-mapping and/or read mapping in mis-assembled regions of the genome may explain these infrequent "heterozygous" SNP calls.

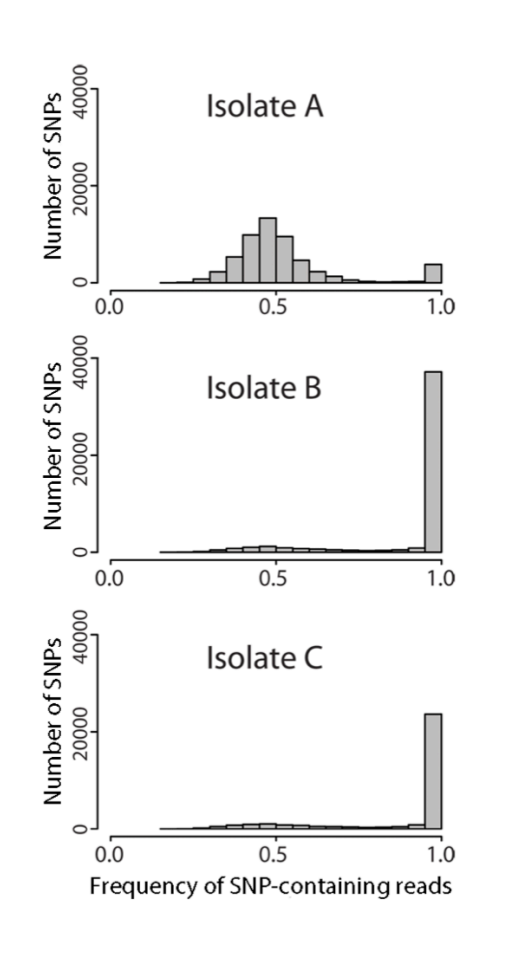


**Figure 2.6. Full alignment of a portion of supercontig 8 that was cloned and sequenced from isolates A, B, and C.** The two different clones from isolate A are indicated as A1 and A2. Of the multiple clones analyzed from isolate B, all sequences were identical, consistent with the predicted homozygosity of SNPs in this sample. Similarly, all clones from haploid isolate C were identical. The alignment is colored according to percent sequence identity; positions that are identical in all samples are dark blue, regions with shared insertions or deletions are light blue, and SNPs and unique insertions are in white. Dots indicate a section of Isolate A2 that we were unable to sequence due to the long GT repeat region, which was excluded from Fig. 2.4C as we could not determine the genotype of all four clones in this region. Fig. 2.4C in the main text shows the SNPs present from position 684 to 2070 (numbered according to Isolate A1) in this alignment.


Isolate_A1/1-2076
Isolate_A2/1-2085
Isolate_B/1-2055
Isolate_C/1-2073

**Figure 2.7. Cell differentiation and cell fusion during haploid-to-diploid transitions.** Time-lapse microscopy shows the fusion of morphologically distinct cells in an isolate C culture following 24 hours of nutrient limitation. A small, rounded cell (the 'male gamete', black arrow) entered the field of view (minute 2) and adhered to a larger, ovoid cell ('female gamete', white arrow; minutes 3 – 22) before cell fusion (gray arrow; minutes 23 – 27). The bracket in the first panel indicates the location of the microvillar collar of the female gamete. Cells were imaged once per minute using 40x phase contrast microscopy. Scale bar is 10 μm.

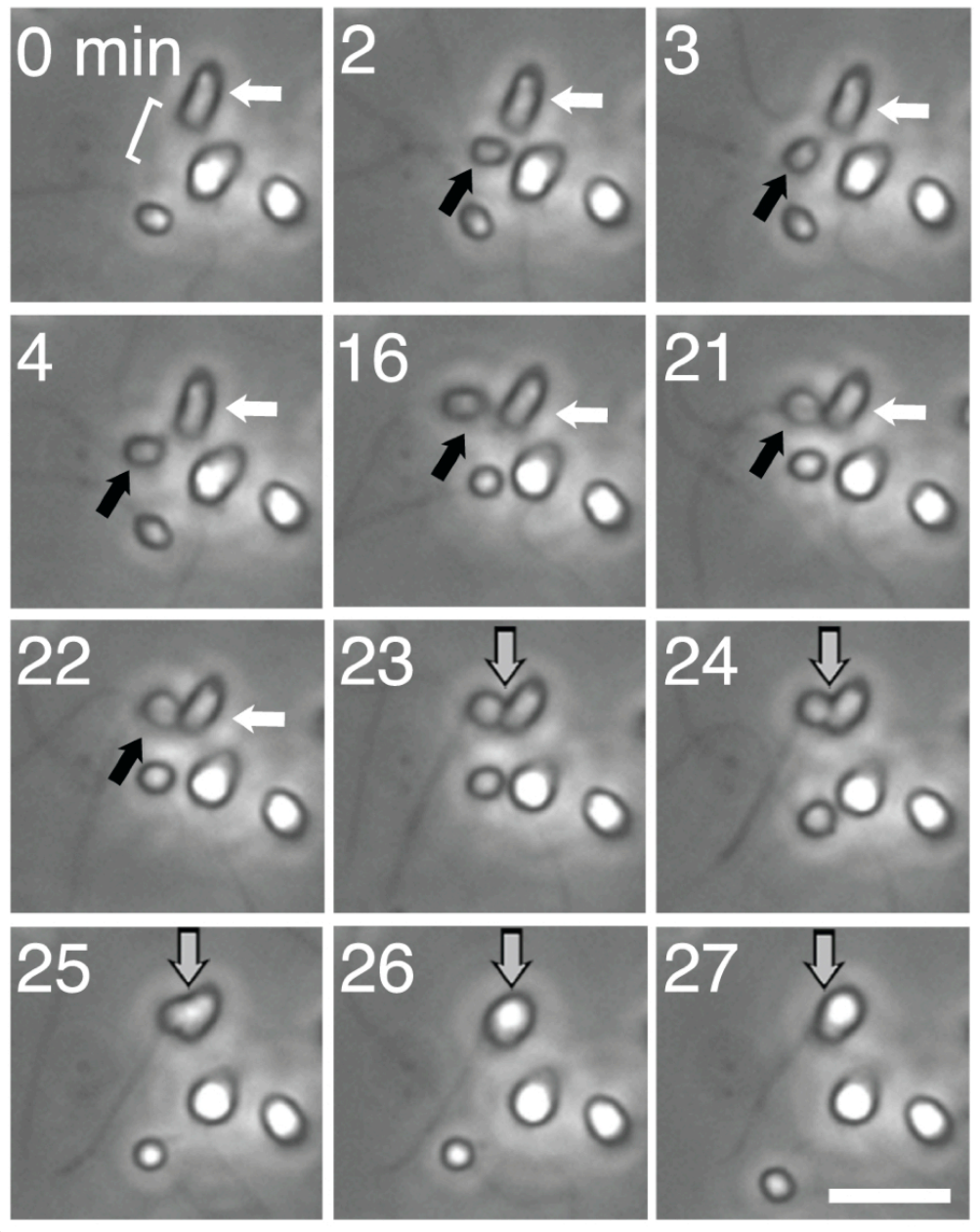
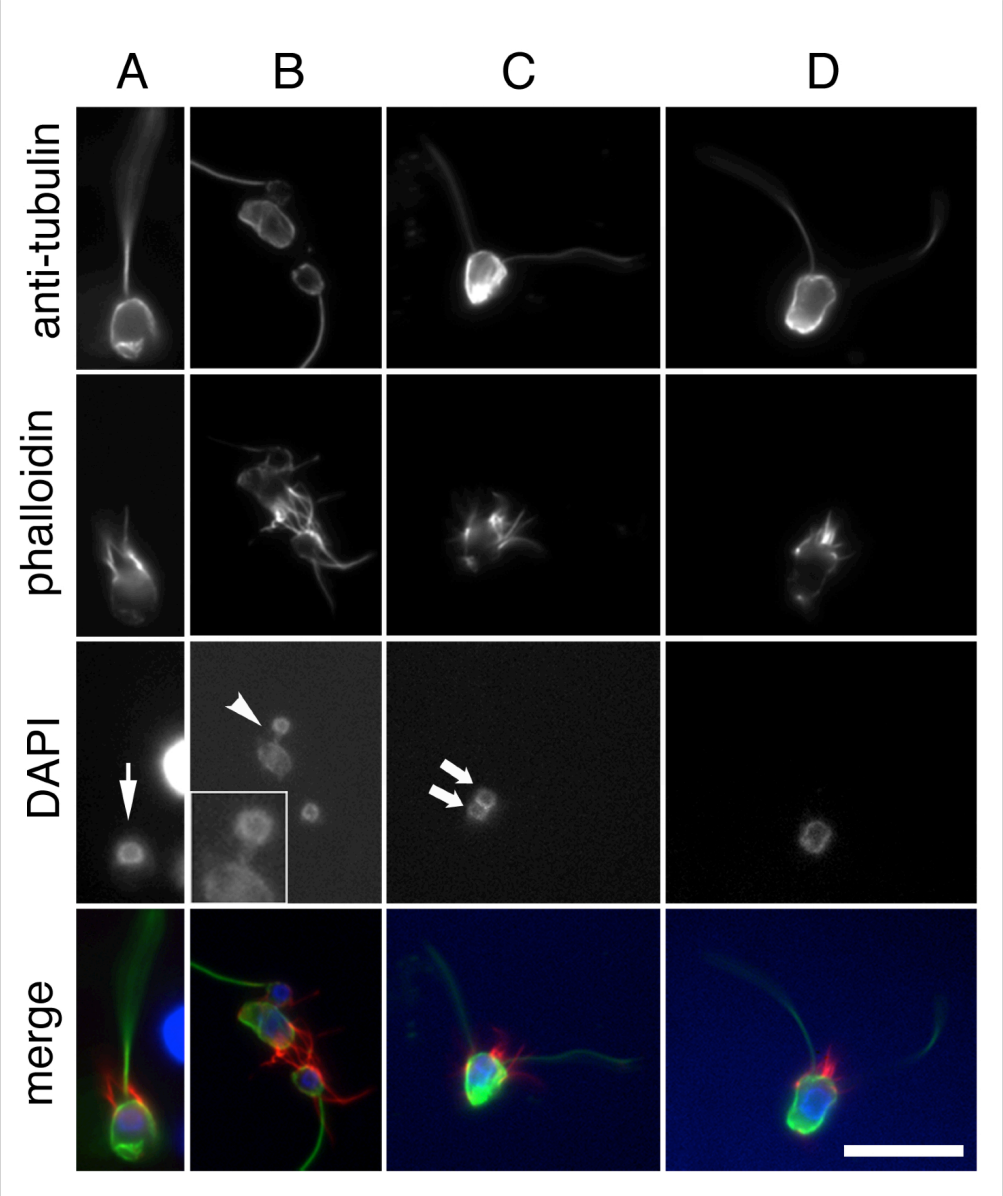


**Figure 2.8. Immunofluorescence of cells during haploid-to-diploid transitions. A)** Haploid isolate C cells grown in HN media have one flagellum and one doughnut-shaped nucleoplasm (indicated by an arrow), as revealed by staining with anti-tubulin antibodies and DAPI. **B)** Nutrient deprived cells differentiate into male and female gamete morphologies and often adhere to form loose clumps. The arrowhead in (**B**) indicates DNA that bridges between the nuclei of two cells, possibly genetic transfer. **(C – D)** Nutrient deprived cultures also contain cells with multiple flagella and either one (**D**) or multiple (**C**) nuclei. These cells may represent intermediate stages of the cell and nuclear fusion involved in sex. Scale bar is 10 μm.



# Chapter 3

# The first choanoflagellate genetic screen links a C-type lectin to rosette formation

*The results presented here were published as part of the following paper:*

Levin, T. C., Greaney, A. J., and N. King. 2014. *A genetic screen implicates a C-type lectin in the regulation of choanoflagellate multicellular development. In review.*

*A.G. helped with screening and mutant characterization. I performed all other experiments and analyses.*

## Summary

The origin of animal multicellularity and development may be reconstructed by comparing animals with their closest living relatives, the choanoflagellates. The choanoflagellate *Salpingoeca rosetta* forms multicellular, rosette-shaped colonies in a process reminiscent of the earliest stages of animal embryogenesis. Here, we establish forward genetics in *S. rosetta* to investigate rosette development. From 15,344 clonal lines screened, we isolated nine mutants with distinct defects in rosette development, including mutants with altered cell packing within rosettes, others that developed into large, amorphous multicellular structures, and one lacked multicellularity entirely. In the study of one mutant, called Rosetteless, a combination of genome re-sequencing and the first choanoflagellate cross revealed that the rosette defect phenotype was perfectly linked to a mutation that causes misregulated splicing in a predicted C-type lectin. The availability of crosses and forward genetics promises to transform the study of choanoflagellate biology and enable a more detailed molecular understanding of animal origins.

## Introduction

Animal origins may be partially reconstructed through the study of choanoflagellates, the closest living relatives of animals [16-22, 38]. Some choanoflagellate species can develop from single cells into organized multicellular colonies in a process reminiscent of the hypothesized transition to multicellularity in the first animals [2, 6, 7, 24, 44, 47, 49, 51, 91]. Thus, understanding the genetic basis of colony development may help to illuminate the origin of animal multicellularity and development.

The choanoflagellate *Salpingoeca rosetta* has recently emerged as a phylogenetically-relevant model organism for testing hypotheses about animal origins. *S. rosetta* was isolated from nature as a "rosette" colony [44], a spherical form composed of anywhere from four to fifty cells (Fig. 1.2). Rosettes develop through repeated rounds of

cell division in which sister cells remain attached through fine cytoplasmic bridges [44, 47]. The development of rosettes from single cells can be induced by specific lipids produced by the prey bacterium, *Algoriphagus machipongonensis* [48]. In the absence of the inducing signals, *S. rosetta* cell division results in the production of single cells or linear, "chain" colonies instead. Additional, as-yet-unidentified environmental cues regulate the ability of *S. rosetta* to differentiate into diverse other cell types and morphologies, including thecate cells, fast swimmer cells, slow swimmer cells, and morphologically differentiated gametes (Fig. 1.2) [44, 47, 92].

Understanding choanoflagellate cell differentiation may help to illuminate animal origins. For example, it is hypothesized that mechanisms for regulating cell differentiation during animal development were likely co-opted from pre-existing mechanisms for regulating transient cell differentiation in pre-metazoans [1, 7, 45, 46]. Indeed, many genes required for animal multicellularity and the regulation of bilaterian development are conserved in protozoan relatives of animals (e.g. choanoflagellates, filastereans, and ichthyosporeans), including Tbx/Brachyury, Myc, homeobox genes, receptor tyrosine kinases, cadherins, integrins and C-type lectins [1, 9, 33-37, 93, 94]. The existence of such genes in animal sister groups suggests that they evolved prior to the origin of animals, although their pre-metazoan functions are largely unknown.

Likewise, although the *S. rosetta* genome has been sequenced and several cell-type specific transcriptomes are available [39], nothing is yet known about how it regulates cell differentiation during its life history. Furthermore, there have been no published reports of successful gene knockdowns, knockouts, or transgene expression in any choanoflagellate. Despite the potential importance of choanoflagellates for reconstructing animal origins, no direct, functional links between genotype and phenotype have been made for any choanoflagellate gene or trait.

Therefore, to identify genes underlying rosette development, we established classical, forward genetics in *S. rosetta* and isolated nine mutants with rosette defects. We focused on a single mutant, Rosetteless, that fails to form rosettes but appears to be wild type in all other respects. To determine which mutation(s) caused the Rosetteless phenotype, we established an approach for performing mapping crosses in choanoflagellates. Strikingly, a single mutation in a C-type lectin gene was perfectly linked to the Rosetteless phenotype. This study reports on the first use of forward genetics in choanoflagellates and provides evidence linking a C-type lectin to the regulation of rosette development.

## Materials and Methods

### Culture media

HN media (250 mg/L peptone, 150 mg/L yeast extract, 150 µL/L glycerol in unenriched sea water) was made by diluting Sea Water Complete Media [65] to 5% (vol/vol) in ASW. *A. machipongonensis* conditioned media (ACM) was made from the sterile

supernatant of a liquid *A. machipongonensis* culture (ATCC BAA-2233) [66] grown shaking for 48 hours in HN media at 30° C to an $OD_{600}$ of 0.30-0.39.

The above conditions were used for the isolation of all mutants except Rosetteless. For this mutant, ACM was prepared in CG media and was grown for 24 hours to an $OD_{600}$ of 0.1. Rosetteless clonal isolation steps used a mixture of 20% *Algoriphagus* conditioned CG media, 5% fresh CG media, and 75% ASW (vol/vol).

**Parental strain for the screen**

The SrEpac parental strain is consistently haploid when cultured in HN media [92] and it includes *S. rosetta* grown with a single bacterial food source, *E. pacifica* (DSM 19836) [69], that does not induce rosette colony development. SrEpac was generated through serial clonal isolation [92] to ensure a genetically homogeneous background for the screen, and frozen stocks of SrEpac were thawed prior to each mutagenesis to limit the accumulation of random mutations. During each mutagenesis treatment, an SrEpac culture was divided in two; one half was mutagenized and the other half underwent all incubations, washes, and clonal isolation steps of the protocol except for the mutagenesis.

**Mutagenesis**

For EMS mutagenesis, approximately $10^6$ cells were washed and resuspended in 1 mL ASW. Liquid EMS (ethyl methanesulfonate, Sigma) was added to 0.3% (vol/vol) and cells were incubated 1 hour at room temperature. The EMS was subsequently removed and neutralized by washing the cells three times in 5% sodium thiosulfate in ASW (wt/vol) before returning the cells to their initial media (HN or CG media) for 24 hours of recovery. The co-isolated wild type strain (C2E5) was an unmutagenized wild type control that was isolated in parallel with Rosetteless.

Fox X-ray mutagenesis, approximately $10^6$ cells were transferred into 35mm diameter tissue culture dishes (Falcon) and placed in an X-ray cabinet (Faxitron 43855C) 30.3 cm from the X-ray source with the lids of the dishes removed. Cultures were irradiated at the 125V 3mA setting for 3 hours, which corresponded to a dose of approximately 6,300 rems. Although we observed only mild choanoflagellate death from the mutagenesis treatments (Fig. 3.3A), there was significant death and/or growth inhibition of the *E. pacifica* bacteria following X-ray mutagenesis. Therefore, to avoid *S. rosetta* starvation, we added 500 µL of an unmutagenized, liquid culture of *E. pacifica* bacteria to the *S. rosetta* after X-ray mutagenesis and resuspended the cells in 10 mL HN media before a 24 hour recovery.

To determine the X-ray dose delivered under these conditions, we placed ring dosimeters at the same position and exposed them for 1 min, 1.5 min, or 1.75 min to generate a standard curve. By linear regression, we obtained the following formula with a

fit of $R^2 = 0.997$: Millirems of exposure = 35091 * (minutes exposure) – 4531.1. Given this equation, we calculated that the X-ray mutagenesis dose corresponded to approximately 6,300 rems.

**Screen for rosette defect mutants**

Twenty-four hours after mutagenesis, control and mutant clones were isolated by dilution-to-extinction as in [92] into 150 µL Screen media (20% ACM, 40% HN media, 40% ASW (vol/vol)) in 96-well plates. Cells were plated at an approximate density of 1 cell/150 µL (aka 1 cell per well). The probability that each isolate underwent a clonal bottleneck during this step was 0.70 to 0.89, calculated using the Poisson distribution and the number of choanoflagellate-free wells per plate as in [92]. After 5-7 days, clonal populations were visually screened for mutants deficient in rosette colony formation. Selected controls and rosette defect mutants were expanded in 3 mL 10% ACM in 6-well plates to verify the phenotype.

In total, we isolated 19 candidate mutants. Nine were later verified as rosette defect mutants through repeated testing of rosette induction (Fig. 3.1B). One mutant had a mild growth defect. Of the remaining nine candidate mutants, most were isolated as thecate cells, a cell type that is not competent to form rosettes (Fig. 1.2D, [44]), but upon further passaging the cells in these cultures began to form rosettes. We concluded that the rosette phenotypes of the thecate isolates were likely a result of epigenetic rather than genetic heritability, and we focused instead on the nine verified rosette defect mutants.

To ensure that each mutant and control isolate was truly clonal, a second clonal isolation step was performed into 96-well plates to an average of 1 cell/1500 µL (aka 1 cell / 10 wells). The probability that each isolate underwent a clonal bottleneck during this step was 0.935 to 0.997, to an overall probability of 0.991-0.999 that each isolate underwent a clonal bottleneck at least once.

**Quantification of mutant rosette defects**

*S. rosetta* cultures were exposed to either ACM or live bacteria. For the live bacteria treatments, *A. machipongonensis* liquid cultures were grown shaking in HN media at 30 C for 24 hours. To begin the induction, *S. rosetta* cells were diluted to $1*10^4$ cells/mL in 3 mL HN media with either 20% ACM or 4 µL/mL of liquid *A. machipongonensis* culture. Forty-eight hours after induction, we pipetted the culture rigorously to break up chain colonies, concentrated the cells 5-fold by centrifugation, fixed an aliquot of the culture with formaldehyde, and assessed rosette formation by counting on a hemacytometer. Thus, our operational definition for rosette colonies only included those rosette colonies that were robust to pipetting.

**Imaging mutant rosette phenotypes**

For all experiments to visualize rosette phenotypes, cells were plated at a density of $10^4$ cells/mL in 3 mL of either HN media or 10% ACM in HN media (vol/vol). Cultures were imaged 48 hours after induction.

For the high magnification DIC images (Fig. 3.2A and 3.2C), 96-well μclear flat bottom plates (Greiner) were coated with 0.1 mg/mL poly-D lysine (Sigma) for 5 minutes and allowed to air dry for 5 minutes before gently transferring 100 μL of culture to the well with a cut-off pipet tip. Cells were allowed to settle for 5 minutes and imaged live at 63X oil immersion with a Leica DMI6000B microscope equipped with a Leica X-Cite 120 camera.

To visualize low magnification fields of view of the mutants following pipetting (Fig. 3.3D), cells were pipetted rigorously to break up chain colonies and concentrated 30-100-fold by centrifugation. Ten to twenty microliters of concentrated cells were imaged live on a slide at 10X on a Leica DMIL LED inverted compound microscope with a Leica DFC 300FX camera.

For the confocal slices through rosette colonies (Fig. 3.2D), sterile, 8-well μ-slides (Ibidi) were coated with 0.1 mg/mL poly-D lysine (Sigma) for 5 minutes and allowed to air dry for 5 minutes before gently transferring 250 μL of culture to the well with a cut-off pipet tip. Cells were fluorescently stained with 1 μL of 2.5 μg/mL FM 1-43X dye (Molecular Probes), fixed with 1 μL 25% gluteraldehyde (Electron Microscopy Services), allowed to settle for 5 minutes, and imaged at 63X using a Zeiss LSM 700 confocal microscope. Single confocal slices are shown.

**Imaging mutant chain phenotypes (Fig. 3.2E and 3.4)**

Because chain colonies break up upon pipetting and because some of the mutants formed chains with very large clusters of cells, we attempted to visualize the chain phenotypes while manipulating the cells as little as possible. Cells were diluted at a 1:10 ratio into 10 mL of HN media in 25 $cm^2$ culture flasks (Corning). Twenty-four hours later, we imaged the chain colonies at the bottom of the flask at 10X using a Leica DMIL LED inverted compound microscope with a Leica DFC 300FX camera. Images were manually false colored to highlight the chain colonies that were in focus.

**Genome sequencing**

We prepared genomic DNA from mutant and wild type *S. rosetta* cultures by phenol chloroform extraction and used a CsCl gradient to separate *S. rosetta* and *E. pacifica* DNA by GC content [68]. Multiplexed, 100 bp paired-end libraries were sequenced on an Illumina HiSeq 2000. Raw reads were trimmed with TrimmomaticPE [76] to remove low quality base calls. Trimmed reads were mapped to the *S. rosetta* reference genome [39] using Burrows-Wheeler Aligner [77]. We removed PCR duplicates with Picard (http://picard.sourceforge.net), realigned reads surrounding indel calls using GATK [95]

and called variants using SAMtools and bcftools [78]. To obtain the high quality variant calls (Table 2), we removed all variants that were called with a quality score below 100 in addition to all variants that were called as heterozygous, as we expected these haploid genomes to yield homozygous calls. The raw reads for the SrEpac parental strain, the C2E5 wild type co-isolate, and Rosetteless are publicly available (accession numbers SRX365844, SRX476076, and SRX476075, respectively). All alignments were made using fast statistical alignment [79].

## Validating SNVs and genotyping

To investigate whether additional mutants isolated in this screen (i.e. mutant classes B-G; Table 1) bore mutations in *EGD82922*, we used Phusion polymerase (New England Biosciences) to clone the coding region of the *EGD82922* from each mutant into the pCR 2.1 vector (Invitrogen). The coding region was divided into three regions for each mutant, using the following primer pairs: Rtls_L1/Rtls_R3, Rtls_L5/Rtls_R4, and Rtls_L3/Rtls_R2 (Table 3). The full insert of each clone was analyzed by Sanger sequencing. No mutations were found in *EGD82922* in any of the eight remaining rosette defect mutants.

Short regions of genomic DNA flanking the predicted unique Rosetteless SNVs were amplified by PCR using a 1:1 mix of Taq (New England Biosciences) and Pfu (Finnzymes), gel extracted using the GeneClean II kit (MP Biomedicals), and analyzed by Sanger sequencing. This process was also used to genotype the gt2 and gt3 markers in the progeny of the cross (Table 3). SNVs were considered 'verified' if they were present in Rosetteless gDNA but absent from gDNA from the parental strain. To genotype microsatellites with size polymorphisms larger than 30 bp (e.g. the indel1 marker), we separated PCR products on a 2% agarose gel. To genotype smaller microsatellites, we fluorescently labeled PCR products as in [96] and analyzed the size polymorphisms by fragment analysis on a 3730XL DNA Analyzer (Applied Biosystems). The gt_indel_2 and gt_indel_7 primer sets included an M13 site on the left primer to enable fluorescent labeling in a 3-primer reaction, while the gt_indel_9 left primer was directly fluorescently labeled (Table 3). To generate a preliminary *S. rosetta* linkage map, the cross progeny were further genotyped at 60 markers using KASP technology by LGC Genomics (Beverly, MA; Table S3). The linkage map was generated in two steps: first, a map based on genetic linkage was constructed using R/qtl, with genotyping data coded as a backcross [97]; second, markers that were known to be physically linked in the *S. rosetta* genome were manually added to the linkage map.

## RT-PCR and cloning of EGD82922 splice isoforms

We isolated RNA from the SrEpac parental strain and Rosetteless using the RNAqueous kit (Life Technologies) and prepared cDNA using oligo dT primers and Superscript II Reverse Transcriptase (Invitrogen) according to the manufacturer's instructions. After first strand synthesis, we performed PCR using primers flanking the candidate splice donor mutation (Table 3) and cloned the resulting bands into the pCR 2.1 vector (Invitrogen). Splice isoforms were determined by Sanger sequencing the full insert

of each clone. From Rosetteless, we sequenced 11 clones that had retained intron 7 (isoform b), 3 clones that had a late splice donor (isoform c), and 21 clones that had an early splice donor (isoform d), but no clones with the wild type isoform. From both wild type and Rosetteless cDNA, no bands were observed from the negative control PCRs, which used as a template RNA that was not reverse transcribed.

## Performing a choanoflagellate cross

### Part 1: Isolation of a haploid strain of Isolate B

As Rosetteless and its SrEpac parental strain had very few genetic variants that could be tracked in a backcross (Fig. 3.5B), we opted instead to perform a cross between the Rosetteless mutant and Isolate B, which was previously sequenced [92] and was predicted to have 39,451 polymorphic markers relative to Rosetteless (Fig. 3.8A). Isolate B had been maintained as a diploid culture [92], which was not suitable for crossing to the haploid Rosetteless strain. However, as Isolate B exhibited genome-wide homozygosity, we reasoned that if we could induce Isolate B to undergo meiosis and generate an Isolate B haploid strain, this haploid strain would inherit the same predicted markers as the sequenced, diploid Isolate B strain. Thus, our first goal was to isolate a haploid strain from Isolate B.

Isolate B consists of *S. rosetta* cells that are fed *A. machipongonensis* bacteria and cultivated in CG media [92]. We induced Isolate B to become haploid by passaging the culture with a 1:2 or 1:5 dilution every 2 to 3 days in CG media for several weeks. Although Isolate B is typically thecate when diploid, this passaging regime resulted in a culture consisting mostly of rosettes. We measured the ploidy of the culture by flow cytometry as in [92] and found that approximately 51% of the population was haploid. To establish a clonal, haploid line of Isolate B, we isolated cells by limiting dilution into 96-well plates containing 10% CG media in ASW. The probability of clonal isolation during this step was 0.93. We selected 12 isolates to expand into larger volumes and measured the ploidy of each clonal population by flow cytometry. We selected one isolate that consisted almost entirely of haploid cells to proceed.

### Part 2: Induction of mating

Our next goal was to induce mating between Rosetteless and the haploid Isolate B culture. At the beginning of the cross, the haploid Rosetteless strain consisted mostly of chain colonies and the haploid Isolate B strain consisted mostly of rosette colonies. *S. rosetta* mating can be induced by transferring a stationary phase culture to nutrient poor media for several days [92]. As *S. rosetta* can undergo both self-fertilization and outcrossed mating [92], we expected this procedure would generate a mixed population of cells that would include: 1) outcrossed heterozygous diploids containing both mutant and wild type alleles, 2) homozygous diploids generated from the self-fertilization of either parental type, and 3) haploid, parental-type cells that never underwent mating. For the purposes of the cross, we were interested in only the outcrossed, heterozygous diploids. To enrich for these cells, we needed to both maximize the proportion of the population that was induced to

mate and attempt to have an equal mix of Rosetteless and Isolate B cells present when mating occurred.

We first attempted to grow the Rosetteless and Isolate B cultures at similar rates through similar passaging regimes. The *S. rosetta* in Rosetteless and Isolate B are cultured with two different species of bacteria, so we added 1mL of an *A. machipongonensis* liquid culture to Rosetteless and 1 mL of an *E. pacifica* liquid culture to Isolate B to ensure that both choanoflagellate cultures were fed both bacterial species. The two isolates were passaged daily for 8 days to a starting cell density of $5*10^4$ cells/mL in 10 mL CG media. This was continued for 8 days. On the first two days, 1 mL of liquid *E. pacifica* culture in HN media was added to each isolate to encourage rapid growth.

On the ninth day, we set up the starvation conditions for the cross. We mixed the cells together by adding $5*10^5$ cells from each culture to 9 mL CG media and 1 mL of liquid *E. pacifica* culture. The next day, the mixed culture was pelleted and resuspended in 10 mL ASW to starve the cells and induce mating [92]. After 11 days of starvation in ASW, we measured the ploidy of the culture and found that approximately 75% of the cell population was now diploid, suggesting that mating had occurred. Three mL of the starved culture was then added into 10 mL CG media and cells were subsequently passaged every 1-3 days for eight days to revive the cultures from their starved state.

**Part 3: Identification of outcrossed diploids**

Following mating, we isolated clones and identified cells that had undergone outcrossed mating through genotyping. We isolated clones by limiting dilution into twenty 96-well plates containing 10% CG media in ASW (vol/vol). The probability of clonal isolation in this step was 0.86. After 1 week of growth, 384 isolates were expanded into 4 mL 50% CG media in 6-well plates to accumulate enough biomass for genotyping. At this stage, nearly all the clonal isolates had differentiated into thecate cells.

After 4-7 days growth, we pelleted 2mL of the culture and extracted DNA using a base/Tris method as follows. We resuspended the pellet in 20 µL base solution (25mM NaOH, 2mM EDTA), transferred the sample into PCR plates, and boiled at 100C for 20 min, followed by cooling at 4C for 5 min. We then added 20 µL Tris solution (40mM Tris-HCl, pH 7.5), and used 1 µL of this sample as the DNA template for each of our genotyping reactions.

To identify which clonal populations were the result of outcrossed mating, we genotyped each of the clonal isolates at three microsatellite markers that were polymorphic between the Rosetteless and Isolate B parental strains (gt_indel_2, gt_indel_7, and gt_indel_9; Table 3). Any clonal populations that had undergone outcrossed mating were expected to be heterozygous at all three markers, while those that did not mate or self-fertilized were expected to be homozygous at all three markers. Of the 384 genotyped clones, 7 were heterozygous at all three markers, suggesting that these clones were the product of outcrossed mating. All remaining clones were homozygous at all three markers.

**Part 4: Induction of meiosis**

We next induced meiosis in the outcrossed heterozygotes to complete the sexual life cycle and obtain recombinant, haploid progeny from the cross. We expanded the heterozygotes in 10 mL CG media and measured the ploidy of the cultures, expecting that the heterozygotes would form largely diploid populations of cells [92]. In four of the seven cultures, the population remained mostly diploid, whereas there was a substantial haploid population in the remaining three cultures, suggesting the meiosis had already occurred for a large subset of these cells. To ensure that we could isolate the products of independent meioses, we divided the four, mostly diploid cultures into ten total flasks and then passaged these cultures rapidly to induce meiosis [92]. One mL of liquid *E. pacifica* culture was added to each flask. The ten flasks were passaged every 1-2 days for 7 days, scraping to dislodge thecate cells and adding liquid *E. pacifica* each time.

**Part 5: Identification of the haploid, meiotic progeny of the cross**

We next repeated the clonal isolation and genotyping steps, as above, to identify the haploid products of meiosis. We genotyped 288 clonal populations and identified 32 isolates that were now homozygous at all three genotyped markers. After several passages, all 32 isolates differentiated into either rosette or chain colonies. We also identified 17 clonal cultures that were homozygous at some markers but heterozygous at others; these apparently diploid clones were presumably generated from the products of meiosis that later underwent a second round of mating before clonal isolation.

In our prior study [92], meiosis was induced in high nutrient (HN) medium and resulted in a more rapid and complete switch from diploid-to-haploid than we were able to achieve during the Rosetteless/Isolate B cross reported here. Because Isolate B is fed on *A. machipongonensis* bacteria instead of *E. pacifica*, it can only be cultured in the less nutrient-rich CG medium, which apparently is less effective at inducing meiosis.

**Linkage map construction**

The program R/qtl was used to construct a preliminary linkage map from the Rosetteless-Isolate B cross progeny [97]. For each genotyping reaction, Rosetteless alleles were coded as “A” and Isolate B alleles were coded as “H” to emulate a backcross and allow for haploid genetics to be analyzed. A single marker, sc015_03, was excluded from the linkage map due to apparent segregation distortion. With 30 independent progeny isolated from the cross, we were limited in our statistical power to detect significant linkage. Therefore, for this preliminary linkage map we chose to use a maximum recombination fraction of 0.4 and a minimum LOD of 1.5. Following linkage group construction by R/qtl, we used the physical linkage known from the *S. rosetta* genome assembly to manually link markers on the same supercontig into a single linkage group.

### Guidelines for choanoflagellate gene naming

To date, no official rules have been established for naming choanoflagellate mutants or genes. Thus, we outline here proposed guidelines for naming choanoflagellate genes. Any genes with clear homology to named genes in other organisms should be referred to by the pre-existing name (e.g. *hsp90*). Any genes without clear homology to named genes should be given names that allude to the gene's function or the phenotype of the first mutant allele isolated. Before mutant genes are cloned, each mutant is given its own name, but renaming may be necessary once the causative mutation is identified. Gene names should be written in lower case and in italics. Mutant names and protein names are written in upper case and non-italics. As many three-letter abbreviations have already been used for genes in other organisms, we propose the use of four-letter abbreviations.

## Results

### A screen for mutants with defects in rosette formation

To identify *S. rosetta* genes required for rosette development, we set out to establish an approach for classical, forward genetics. As this was to be the first genetic screen to be performed in a choanoflagellate, we began with a relatively straightforward screen design (Fig. 3.1A). To induce mutations, cultures of haploid cells were exposed either to EMS or X-rays at standard doses [98-100]. Clonal lines of potential mutants were generated by isolating individual cells through limiting dilution (i.e. on average, plating less than one cell/well) into 96-well plates containing *A. machipongonensis* conditioned media (ACM). After five to seven days of proliferation in ACM, wells seeded with wild type cells were filled with rosette colonies. In contrast, wells seeded with mutant cells defective in rosette development were expected to produce cultures of solitary cells and/or chain colonies, but few to no rosette colonies, even in the presence of ACM (Fig. 1.2D).

In total, we screened 15,344 clonal cultures for the presence or absence of rosette colonies. Each culture containing a putative rosette defect mutant was subjected to a second round of limiting dilution to ensure clonality and was re-tested for the ability to form rosette colonies in response to live bacteria or ACM. Cultures that reverted to wild-type levels of rosette colony development were discarded. Nine mutants with validated rosette defects were isolated, each of which showed a dramatic reduction in rosette colony development relative to wild type following treatment with ACM (Fig. 3.1B and 3.2B).

Several features of the screen design enhanced our ability to identify mutants with defects that were specific to rosette development. First, mutants with general growth defects, which we expected would result in secondary and non-specific defects in rosette development, would not be recovered through this screen as they would fail to proliferate following clonal isolation. In addition, the use of sterilized ACM rather than live *A. machipongonensis* bacteria during primary screening meant that the mutant cultures could

later be examined separately in the presence or absence of live *A. machipongonensis* or ACM. Finally, secondary screening revealed that some of the isolated mutants that were defective in rosette development when grown in the presence of ACM formed rosette colonies more readily in the presence of live *A. machipongonensis* (Table 1, Fig. 3.3C). Thus, the use of ACM served to sensitize the primary screen and allowed us to isolate mutants with both mild and severe defects in rosette development.

**Phenotypes of rosette defect mutants**

We categorized the nine validated rosette defect mutants into six broad phenotypic classes by their ability to form rosette colonies, their swimming behavior as solitary cells, and the morphology of the chain colonies when grown in the absence of ACM (classes A-G; Table 1, Fig. 3.2). Class A consisted of a single mutant, Rosetteless, that was isolated after EMS treatment. In the presence of either ACM or live *A. machipongonensis,* Rosetteless showed a complete lack of rosette formation, but was otherwise indistinguishable from wild type (Fig. 3.2A and 3.3B). Mutants from classes B, C, and D were capable of occasionally forming rosette colonies in response to ACM or live bacteria, although Uptight and Slacker exhibited aberrant cell packing within the mutant rosettes (Fig. 3.2B, 3.2D, and 3.3C). As rosettes form through incomplete cytokinesis and not aggregation [44, 47, 92], we infer that these cell packing phenotypes arose from altered cell-cell or cell-extracellular matrix interactions following cell division. Meanwhile, rosette development was never observed in mutants from classes E, F, or G (Fig. 3.2B, Table1).

As rosette colonies represent only one of two multicellular stages in the *S. rosetta* life history (Fig. 1.2D), we also investigated whether the rosette defect mutants had co-occurring defects in their ability to form linear, chain colonies when grown without *A. machipongonensis*. Three mutants (Rosetteless, Insensate, and Slacker; classes A, B, and C) formed apparently wild type chain colonies consisting of three to twelve cells. In contrast, Uptight, Jumble, and Branched (classes D and E) displayed slightly more branched or clustered chain morphologies than wild type cultures, while Seafoam and Soapsuds (class F) grew into giant, disorganized clusters, sometimes including over a hundred cells (Table 1, Fig. 3.2C, 3.2E, and 3.4). The large clusters and branched chains of mutants in classes D, E, and F were distinguished from rosette colonies by: 1) their amorphous rather than spherical shapes, 2) their occurrence even in the absence of ACM, and 3) their sensitivity to mechanical disruption, as compared to the robustness of rosette colonies (Fig. 3.3E). In contrast to these altered chain morphology phenotypes, Solo mutants (class G) existed primarily as single cells and rarely formed chain colonies (Fig. 3.4). Solo mutant cells also swam slowly with pronounced lateral shaking and differentiated into thecate cells more readily than wild type. The ability to isolate mutants with a range of rosette formation phenotypes demonstrates the potential of forward genetics to illuminate diverse aspects of rosette colony development in *S. rosetta*.

**Identifying candidate mutations in the Rosetteless genome**

Prior to this study, there had not to our knowledge been any attempt at choanoflagellate genetics, so it was initially unclear how best to identify the causative mutations underlying the rosette defect mutants isolated in this screen. We therefore focused our efforts on a single mutant – the Rosetteless mutant – selected because of the high penetrance of its rosette development phenotype (i.e. a complete lack of rosette formation; Fig. 3.2) and the absence of any additional obvious defects. A classical approach would involve a series of mapping crosses to strains with phenotypic markers, but crosses had not previously been reported in choanoflagellates and no validated markers were available. An alternative, increasingly common approach uses whole genome sequencing to identify mutations [101-108]. However the main challenge in this approach lies in differentiating the causative mutations from other mutations, particularly if a high mutagen dose has been used. We opted to use a combination of the two approaches to map the genetic lesion(s) underlying the Rosetteless mutant. Whole genome sequencing allowed us to generate a list of candidate mutations and identify polymorphic markers, while a classical mapping cross narrowed down the locus of the causative mutation.

We began by sequencing the genomes of the Rosetteless mutant, the parental strain from which it was derived, and an unmutagenized wild type strain (C2E5) that was isolated and cultured in parallel with Rosetteless. Because each of these strains was grown in the presence of the bacterium *Echinicola pacifica* (44% GC) [69], we were able to separate the genomic DNA of *S. rosetta* (56% GC) [39] from the genomic DNA of *E. pacifica* over a cesium chloride gradient prior to Illumina sequencing. Rosetteless was sequenced to a median coverage of 71X and over 93% of the reference genome had at least 10X coverage, while the parental strain and C2E5 were each sequenced to a median coverage of 50-60X and over 91% of the genome had at least 10X coverage.

We next searched for potential mutations by comparing Rosetteless to the *S. rosetta* reference genome [39], focusing on detecting single nucleotide variants (SNVs) because Rosetteless was isolated following EMS treatment. After filtering the SNVs by quality, we found that Rosetteless contained 25,186 high-quality SNVs, 25,169 of which were shared among Rosetteless and at least one of the wild type strains (the parental strain and C2E5), meaning they were segregating polymorphisms unlikely to contribute to the Rosetteless phenotype. That left only 17 SNVs that were unique to Rosetteless and therefore plausible candidates for causing the Rosetteless phenotype (Fig. 3.5A and Table 2).

We were initially surprised to find such a small number of unique mutations in Rosetteless. However, it is possible that the EMS mutagenesis was ineffective, which is consistent with both the low rate of mutant recovery following EMS treatment and the fact that the mutagenesis did not result in substantial cell death (Fig. 3.1B and 3.3A). Thus, despite the fact that Rosetteless was derived from a culture treated with EMS, it may in fact be a spontaneous mutant.

Of the 17 Rosetteless-specific SNVs, only a single T-to-C mutation at position 427,804 on supercontig 8 was predicted to alter a coding region, by changing a predicted splice donor in gene *EGD82922* from GT to GC. We confirmed that this mutation was

present in Rosetteless and absent from the parental strain by PCR and Sanger sequencing (Fig. 3.5C). In contrast, when we attempted to verify the other 16 detected Rosetteless SNVs, only three were verified as polymorphic between Rosetteless and the parental strain (Fig. 3.5B). Of the remaining variants, three were false-positive variant calls in Rosetteless, two lay within regions of the reference genome that were misassembled, and eight were false-negative variant calls, where shared, segregating polymorphisms were not identified in the parental strain or the C2E5 wild type strain (Table 2). This meant that of the 17 potential unique SNVs originally identified in the Rosetteless genome, there remained only four verified SNVs, including the predicted splice donor mutation at supercontig 8: position 427,804 (Fig. 3.5B).

**Misregulated splicing of a C-type lectin in the Rosetteless mutant**

We hypothesized that the candidate splice donor mutation would alter the splicing of gene *EGD82922* in Rosetteless cells (Fig. 3.6). To test this hypothesis, we used RT-PCR to amplify the region flanking the mutation from Rosetteless and wild type cDNA. From wild type cDNA, we obtained a single band of the size predicted from the gene model (516 bp). In contrast, from Rosetteless mutant cDNA we obtained RT-PCR products of diverse sizes (Fig. 3.6D). By cloning and sequencing 35 of the RT-PCR products, we found that there were three major splice isoforms, each of which differed from wild type. The major splice isoforms were a 736 bp transcript in which the entirety of intron 7 was retained and two smaller transcripts (502 bp and 543 bp) in which another splice donor either 14 bp upstream or 27 bp downstream of the mutation was used instead (Fig. 3.6E). For two of the major splice isoforms, the altered splicing led to a frame shift and an early stop codon downstream of the mutation.

The *EGD82922* gene encodes a predicted 119 kDa C-type lectin protein. Transcriptome profiling performed as part of the *S. rosetta* genome project [39] had previously shown that *EGD82922* is expressed approximately five times higher in swimming cells, chain colonies, and rosette colonies than in thecate cells. The predicted wild type EGD82922 protein contains a N-terminal signal peptide and two C-type lectin-like domains (CTLDs; Fig. 3.6A). Alignment of the EGD82922 CTLDs with CTLDs from animals revealed that the cysteines required for two disulfide bonds are conserved, as is the mannose-type binding motif Glu-Pro-Asn (EPN; Fig. 3.6B). The gene also encodes several low complexity regions, each containing as many as 50-60 consecutive threonines and serines, and two high-complexity internal repeat regions (RP1 and RP2) of unknown function near the C-terminus of the protein (Fig. 3.6A and 3.6C). The serine-threonine-rich regions resemble mucin-like domains found in some animal C-type lectins and are likely sites of O-linked glycosylation [109, 110]. The candidate splice donor mutation, a T-to-C mutation in the splice donor of intron 7, falls 3' of the sequences encoding the two CTLDs and 5' of the RP1 and RP2 sequences. Thus, we hypothesize that the early stop codons in two of the three major splice isoforms either lead to a truncation of the EGD82922 protein or to degradation of the transcript by nonsense mediated decay in the Rosetteless mutant [111].

**Rosetteless maps to the C-type lectin gene *EGD82922***

The altered splicing of *EGD82922*, although compelling, did not necessarily mean that this mutation was more likely than the other four Rosetteless SNVs to cause the rosette defect phenotype. It also remained possible that the Rosetteless causative mutation(s) could reside in a poorly covered or unassembled region of the genome in which we were unable to call variants, or that the genetic lesion underlying the Rosetteless phenotype could have been a large-scale deletion or chromosomal rearrangement that would have been challenging to detect by Illumina sequencing. We therefore pursued an independent approach to determine the causative mutation.

Our recent discovery of the sexual cycle of *S. rosetta* [92] suggested that it might be possible to perform choanoflagellate mapping crosses. Although we initially considered backcrossing Rosetteless to its parental strain, the high level of similarity between these genomes left few informative genetic markers that could be tracked in such a cross (Fig. 3.5B). Instead, we opted to cross Rosetteless with a different, previously sequenced strain, "Isolate B" [92] that had 39,451 putative polymorphic markers relative to Rosetteless after filtering out poorly supported variants (Fig. 3.8).

As it is not currently possible to morphologically differentiate haploid and diploid *S. rosetta* cells and there are no available cell surface markers, we relied on clonal isolation and genotyping to monitor the cross. We reasoned that genotyping each clone at three unlinked markers could provide evidence for meiosis through independent assortment, while also allowing for multiple, genetically distinct, progeny to be isolated from each flask. Because the chromosome number of *S. rosetta* has not been determined, we selected three markers, each on one of the three largest assembled supercontigs (2.5 Mb, 2.0 Mb, and 1.9 Mb in size), to help ensure that the markers would either be on different chromosomes or far enough apart to be unlinked.

To cross Rosetteless with Isolate B, we used a two-phase procedure (Fig. 3.7A). First, to identify the diploid products of successful matings, we mixed the two haploid cultures together, induced mating through starvation, isolated clones by limiting dilution into 96-well plates, and genotyped the clones at the three microsatellite markers (Experimental Procedures). Prior to clonal isolation, we measured the ploidy of the starved culture and found that approximately 75% of the starved population had become diploid, suggesting that mating occurred for the majority of the cells. However, as *S. rosetta* is capable of both out-crossing and self-mating, only a subset of mating events were useful for genetic mapping. Indeed, of the 382 clones genotyped, only seven (1.8%) were heterozygous at all three markers, demonstrating that a minority of the clones underwent outcrossed mating. Of the remaining clones, we note that 372 were homozygous at each of the three microsatellite markers for the Isolate B alleles, while only three were homozygous for the Rosetteless alleles, raising the possibility that the Rosetteless mutant was less viable than Isolate B under the starvation conditions used to induce mating. Such differential viability may also explain the low rate of outcrossed mating, if few Rosetteless

haploids survived to mate with Isolate B haploids. Despite the low frequency of outcrossed mating, we were able to proceed with the cross using the seven heterozygous isolates.

We attempted to assess the phenotypes of the heterozygous diploids to determine if the Rosetteless mutation was dominant or recessive. However, the starvation procedure of the cross caused the diploids of all genotypes to differentiate into thecate cells, a cell type that is not competent for rosette formation (Fig. 1.2D) [44]. Therefore, it was not possible to assess the rosette phenotypes of the isolates at this stage.

In the second phase of the cross, we induced meiosis in the heterozygous diploid isolates. Because *S. rosetta* haploid cells can clonally replicate, there was a possibility of recovering multiple instances of the same clone following meiosis. Therefore, to enhance our ability to isolate the products of independent meioses, we divided the heterozygous cultures into ten flasks prior to inducing meiosis. After inducing meiosis, we isolated clones from each flask by limiting dilution and genotyped the clonal isolates. Of 288 clonal cultures genotyped, 32 (11%) were haploid progeny of the cross, as evidenced by their homozygosity at all three microsatellite markers.

We next focused on the genotypes and phenotypes of the cross progeny. Two of the thirty-two isolates were genetically identical to others in the set and were excluded from further analysis. Of the remaining 30 haploid progeny isolated, 57% (17/30) had the Rosetteless phenotype while the remaining progeny readily formed rosette colonies, suggesting that the Rosetteless phenotype behaves as a Mendelian trait and is linked to a single genetic locus (Fig. 3.7B). In addition, each allele of the three microsatellite markers tracked in the cross was present in approximately 50% of the progeny and exhibited independent assortment, as predicted by Mendel's first and second laws of inheritance [112]. Thus, we infer that the haploid cross progeny indeed underwent meiosis, and that choanoflagellate inheritance follows the rules of classical genetics.

To simultaneously map the Rosetteless phenotype and generate the first linkage map for *S. rosetta*, we genotyped each of the 30 cross isolates at 66 sites across the genome, including all four previously identified SNVs that were unique to the Rosetteless mutant. In the resulting linkage map, we linked markers on 30 supercontigs into 13 preliminary linkage groups, comprising approximately 60% of the *S. rosetta* genome (Fig. 3.9), Remarkably, we found that the candidate splice donor mutation in gene *EGD82922* on supercontig 8 (position 427,804) was perfectly linked to the Rosetteless phenotype (30/30; Fig. 3.7B). Furthermore, the only other markers that were significantly linked to the phenotype were also found on supercontig 8 (Fig. 3.7C). Thus, the defect underlying the Rosetteless phenotype is either caused by the splice donor mutation or another mutation that is very tightly linked to this position.

Rosetteless genome sequencing did not reveal any other closely linked SNVs or microsatellites on supercontig 8 that could either represent alternative candidate mutations underlying the Rosetteless phenotype or that could be used as markers to further test the linkage of supercontig 8: 427,804 to the Rosetteless phenotype (Fig. 3.5B). Indeed, the gene *EGD82922* sits in the midst of a 'marker desert' on supercontig 8, a haplotype block that – with the exception of the T-to-C mutation at position 427,804 in

Rosetteless – is otherwise seemingly identical between Rosetteless and Isolate B (Fig. 3.7C and 3.8). To investigate whether our variant calling method was too restrictive, we genotyped the heterozygous diploids at 20 additional putative polymorphisms at the 5' end of supercontig 8 that were called below our quality cutoff, but we confirmed that none of the markers were polymorphic in this cross.

An additional strategy for approximating the location of the causative mutation would be to convert the genetic distance from linked markers into physical distance using the recombination rate. The preliminary linkage map constructed here allowed us to estimate the *S. rosetta* genome-wide mean recombination rate as 29 kb/cM, with a 95% confidence interval of 22 to 40 kb/cM. Combining the 95% confidence interval with information from the three significantly linked supercontig 8 markers gives an estimate that the causative mutation should lie on supercontig 8 between 58,597 and 564,810. Similarly, the local supercontig 8 recombination rate gave an estimated position of 376,211 for the causative mutation, only 2 cM away from the candidate splice donor mutation. These results suggest that the region of supercontig 8 that is perfectly linked to the phenotype likely lies entirely within the supercontig 8 'marker desert'. Therefore, we favor the interpretation that the candidate splice donor mutation in gene *EGD92922* is responsible for the rosette defect phenotype, and we propose that *EGD92922* should be renamed *rosetteless* (*rtls*).

## Discussion

### Forward genetics in a new model organism

As the closest living relatives of animals, choanoflagellates can offer insights into animal origins that are unobtainable from other organisms. Here, we establish an approach for classical, forward genetics in *S. rosetta* by combining random mutagenesis with mutant genome re-sequencing and a single mapping cross, an approach that may be applicable to diverse emerging model organisms. We expect forward genetic screens to be useful for understanding *S. rosetta* colony development because similar screens have been vital tools for uncovering the fundamental mechanisms driving development in many model organisms, including *Drosophila*, *Caehorhabditis*, zebrafish, and *Arabidopsis* [98, 113-115]. Furthermore, although some *S. rosetta* genes have clear homology to genes in animals, approximately 25% of genes in the *S. rosetta* genome have no predicted protein domains and 79% are annotated as "hypothetical proteins", making candidate gene approaches a challenge in this organism.

A number of features of choanoflagellate biology make them well suited to forward genetics. First, *S. rosetta* cells can be cultured and clonally propagated as haploids, eliminating the need for backcrossing or multi-generational screening strategies. Second, these organisms are amendable to cryopreservation and their relatively small genome (55 Mb) makes mutant whole genome re-sequencing possible. Third, haploid strains of *S.*

*rosetta* produce both male and female gametes [92], apparently allowing for any two strains to be crossed without limitations of mating type. Finally, up to $10^{10}$ *S. rosetta* cells can be simultaneously cultivated in the laboratory, making it possible to isolate minimally-mutagenized or spontaneous mutants, thereby greatly facilitating mapping.

This work provides a number of additional tools for forward genetics in *S. rosetta*. First, we outline a simple approach for isolating mutants and identifying genes underlying rosette development. Second, 66 validated markers across the *S. rosetta* genome and a preliminary linkage map are now available for future mapping efforts. Finally, we present a procedure to cross *S. rosetta* isolates in the lab through the use of different culture media, clonal isolation, and genotyping.

## Potential roles for the Rosetteless C-type lectin

In this work, we mapped the Rosetteless phenotype to a predicted C-type lectin, the first gene to be implicated in the rosette development pathway. Through their specific and high-affinity binding to carbohydrates, proteoglycans, lipids, and other ligands, C-type lectins in animals play diverse roles, including cell-cell adhesion, cell-extracellular matrix adhesion, cell signaling, and innate immune recognition of pathogens [116-120]. Thus, there are a number of reasonable hypotheses for the function of the Rosetteless C-type lectin in rosette colony development. First, Rosetteless could be part of the *S. rosetta* extracellular matrix, which fills the center of rosette colonies and may play a role in rosette structural stability [44]. Alternatively, Rosetteless could serve as a secreted receptor for rosette-inducing bacterial lipids, similar to the Mincle C-type lectin [121]. Finally, as the inducing lipids are hydrophobic, *S. rosetta* may perceive these signals in the context of bacterial vesicles, which are prevalent in marine ecosystems [122]; rather than having a role in binding the inducing lipids directly, Rosetteless could play a role in generic vesicle recognition.

## Toward understanding the rosette development pathway

The rosette development pathway may be further elucidated through the characterization and mapping of the eight additional rosette defect mutants described here (Fig. 3.2, Table 1). Some insights into the pathway are offered through the analysis of the mutant phenotypes alone. For example, six of the nine rosette defect mutants displayed co-occurring chain morphology phenotypes, suggesting that rosette colony formation and chain colony formation are mechanistically overlapping developmental processes, an inference that is supported by their similar transcriptional profiles [39] and ultrastructural features [44]. At the same time, three of the mutants (Rosetteless, Insensate, and Slacker) showed essentially wild type chains, indicating that the rosette and chain formation pathways can be separated. In addition, the Uptight and Slacker mutants each developed into malformed rosettes that were less stable than wild type, but the lack of stability was not correlated with cell proximity within the rosette (Fig. 3.2D and 3.3D), perhaps suggesting that direct cell-cell interactions are less important than cell-extracellular matrix

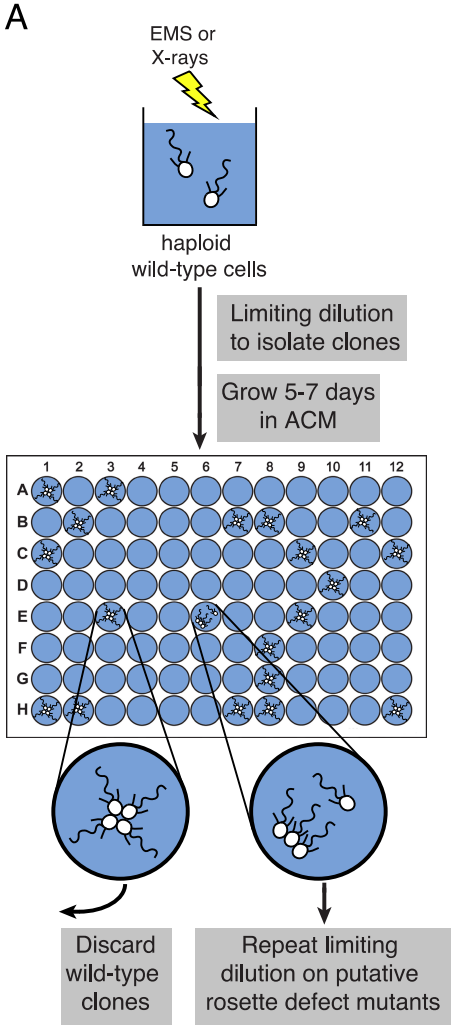
interactions in rosette stabilization. Meanwhile, the Solo mutant formed neither rosettes nor chains, but it differentiated more readily into thecate cells and fast swimmer cells that are not competent to form rosettes. We hypothesize that the Solo mutant cells contain a differentiation defect that restricts them to the part of the *S. rosetta* life history (e.g. thecate cells, fast swimmer cells, and a subset of slow swimmer cells; Fig. 1.2D) that is not competent to produce rosette or chain colonies [44].

As none of the remaining mutants had mutations within the coding region of *rosetteless* (Experimental Procedures), they likely have mutations in different genes in the rosette formation pathway. Although it remains unknown how many complementation groups are represented among the eight other mutants, it is unlikely that the screen has been performed to saturation and thus further screening is likely to yield additional informative mutants. Moreover, the establishment of classical genetics in *S. rosetta* means that we can search for suppressors of the *rosetteless* mutation and, thereby, potentially identify genetic interactors. With these genetic approaches, we can begin to gain mechanistic insight into rosette development and diverse other aspects of choanoflagellate biology. The information uncovered will enable an understanding of how these organisms achieve multicellularity while also clarifying whether molecular mechanisms underlying multicellular development are conserved between animals and choanoflagellates.

## FIGURES AND TABLES

**Figure 3.1. A screen for rosette defect mutants**. **A)** Haploid *S. rosetta* cultures were mutagenized by EMS or X-rays before clonal isolation by limiting dilution into liquid media. After five to seven days of growth in *A. machipongonensis* conditioned media (ACM), clones were visually screened for defects in rosette formation. Putative rosette defect mutants underwent a second round of limiting dilution to ensure clonality and their phenotypes were retested. **B)** Results of screen. After screening 15,344 clones, nine mutants were recovered with rosette defects.

A




B

| mutagen | # clones screened | # rosette mutants isolated |
|---|---|---|
| EMS | 8,645 | 2 |
| X-rays | 6,699 | 7 |

**Table 1: Classification of rosette-defect mutant phenotypes**

| | | Observed rosette induction* | | Other phenotypes | |
|---|---|---|---|---|---|
| Strain | Mutagen used | ACM | Live bacteria | Swimming** | Chain morphology |
| Wild type | N/A | 86% | 88% | Wild type | Primarily linear |
| *Mutant class A* | | | | | |
| Rosetteless | EMS | 0 | 0 | Wild type | Primarily linear |
| *Mutant class B* | | | | | |
| Insensate | X-ray | 0 | 5 | Wild type | Primarily linear |
| *Mutant class C* | | | | | |
| Slacker | X-ray | 20 | 42 | Wild type | Primarily linear |
| *Mutant class D* | | | | | |
| Uptight | X-ray | 33 | 56 | Wild type | Branched |
| *Mutant class E* | | | | | |
| Jumble | EMS | 0 | 0 | Wild type | Branched |
| Branched | X-ray | 0 | 0 | Wild type | Branched |
| *Mutant class F* | | | | | |
| Seafoam | X-ray | 0 | 0 | Wild type | Large clusters |
| Soapsuds | X-ray | 0 | 0 | Wild type | Large clusters |
| *Mutant class G* | | | | | |
| Solo | X-ray | 0 | 0 | Slow, shaking | Primarily solitary, i.e. rarely forms chains |

* Rosette induction was quantified as the percentage of cells in rosettes following induction with either *A. machipongonensis* conditioned media (ACM) or live bacteria. For mutants that produced no observable rosette colonies, the percentage of rosette induction is indicated as 0%.

** Swimming phenotypes were observed in single cells.

**Figure 3.2. Phenotypes of rosette defect mutants**. **A)** Wild type *S. rosetta* grown in the absence of *A. machipongonensis* conditioned media (ACM) formed single cells (Fig. 1.2A) or flexible, linear chains. When exposed to conditioned media from *A. machipongonensis* bacteria (ACM), wild type *S. rosetta* formed rosette colonies (arrowheads). The Rosetteless mutant did not form rosette colonies in ACM, but appeared to be otherwise wild type. Scale bar = 10 μm. **B)** Quantification of rosette induction in 20% ACM, shown as mean +/- SEM. After 48 hours in ACM, over 80% of wild type cells were found within rosette colonies. All mutant cultures showed a significantly reduced number of cells of rosette colonies (one-tailed Mann-Whitney test, $p<0.005$). Ø represents cultures in which no rosettes were observed (limit of detection = 0.03%). **C)** Morphologies of the remaining eight rosette defect mutants in the presence and absence of ACM. Although a small percentage of Slacker and Uptight cells were found in *bona fide* rosette colonies (as defined by their organization and resistance to mechanical shear; arrowheads), most remained as single cells or chain colonies that were easily disrupted when exposed to shear (Fig. 3.3E). Seafoam and Soapsuds formed large, disorganized clusters of cells that were easily disrupted when exposed to shear (Fig. 3.3E) and were thus not rosette colonies. Scale bar = 10 μm. **D)** Single confocal slices through rosette colonies stained with FM 1-43X dye showed differences in cell packing within wild type and mutant rosettes. Bottom: Inverted images were false colored to mark the space between cells in the center of the rosette. Scale bar = 5 μm. **E)** Fields of view containing *S. rosetta* chain colonies from wild type and mutant cultures in the absence of *A. machipongonensis* signals. Each white circle is a choanoflagellate cell and dark specks are bacteria. Bottom: Choanoflagellate cells were false colored to emphasize the morphology of the chain colonies. Scale bar = 50 μm.

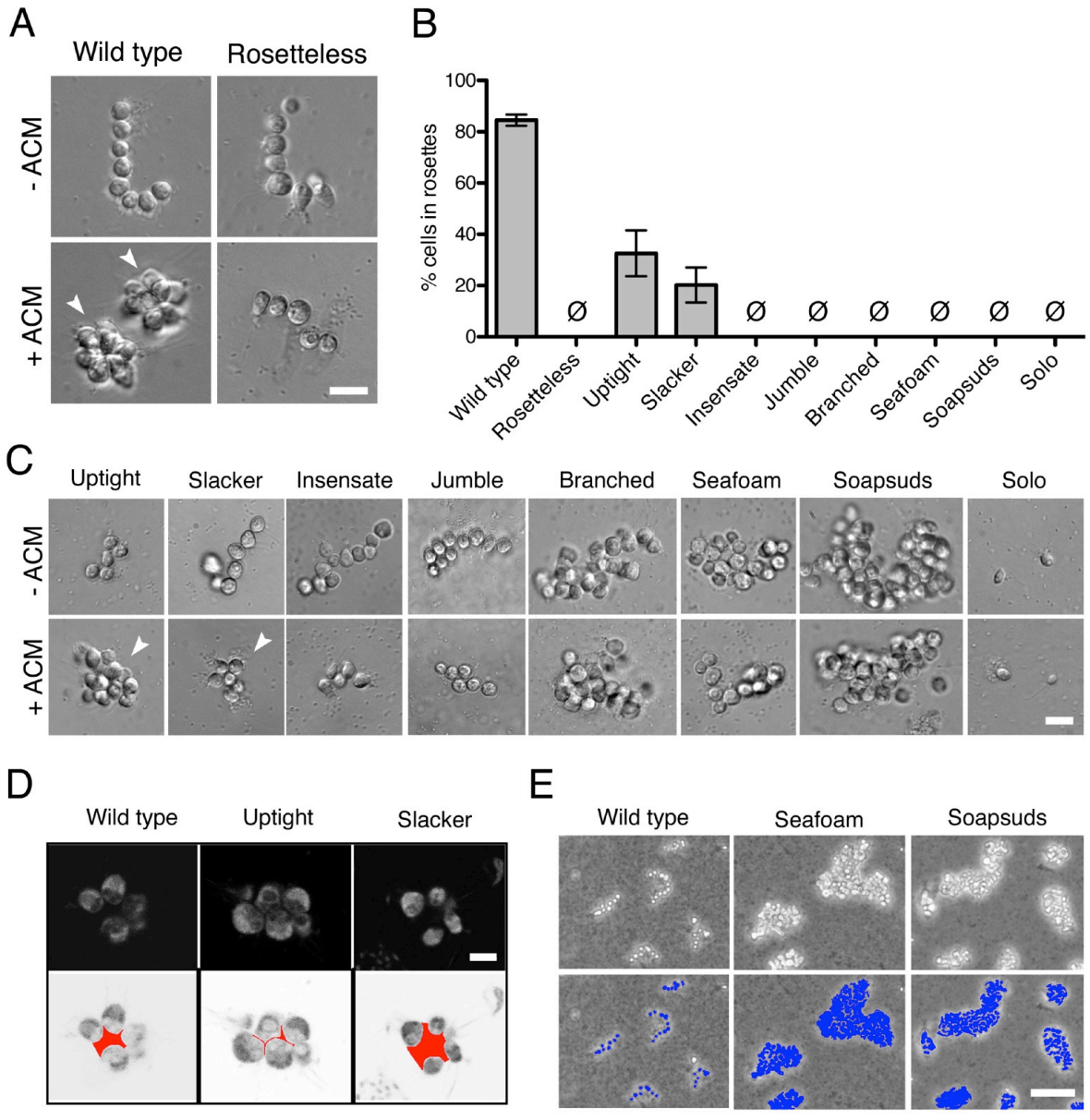

A
Wild type
Rosetteless
- ACM
+ ACM
B
% cells in rosettes
100
80
60
40
20
0
Ø
Wild type
Rosetteless
Uptight
Slacker
Insensate
Jumble
Branched
Seafoam
Soapsuds
Solo
C
Uptight
Slacker
Insensate
Jumble
Branched
Seafoam
Soapsuds
Solo
- ACM
+ ACM
D
Wild type
Uptight
Slacker
E
Wild type
Seafoam
Soapsuds

**Figure 3.3. Mutagenesis and mutant phenotypes. A)** Vertical scatter plot showing the effect of mutagenesis on cell number, as the concentration of mutagenized cells divided by the concentration of cells in a paired, unmutagenized culture at 24 hours post-mutagenesis. Each dot represents one mutagenesis experiment and the mutagenesis from which the Rosetteless mutant was isolated is highlighted (red). Horizontal lines show the mean values for EMS and X-ray mutagenesis. The dotted line at 1.0 represents no effect. Both mutagenesis treatments resulted, on average, in a mild decrease in cell number ($p = 0.01$ EMS, $p = 0.001$ X-rays, one-tailed t-test), although the decrease in cell concentration following X-ray treatment may have been due, in part, to *S. rosetta* starvation caused by the death of the *E. pacifica* prey bacteria. **B)** Growth curve of wild type (open circles, dotted line) and Rosetteless (filled circles, solid line) shows that the Rosetteless phenotype is not due to a growth defect. Error bars show standard deviation. **C)** Quantification of rosette induction in the presence of live *A. machipongonensis*, shown as mean +/- SEM. Ø represents cultures in which no rosettes were observed (limit of detection = 0.03%). Compare to Fig. 3.2B to see rosette induction from *A. machipongonensis* conditioned media (ACM) rather than live bacteria. Notably, the Insensate mutant shows a low level of rosette induction when exposed to live bacteria but not when exposed to ACM. **D)** Phenotypes of rosette defect mutants after rigorous pipetting. Rosettes (red arrowheads) form in the presence of ACM and are robust to pipetting, whereas chain colonies break up into single cells. The Uptight mutant occasionally forms rosettes, but none were visible in this field of view. Scale bar = 20 μm.

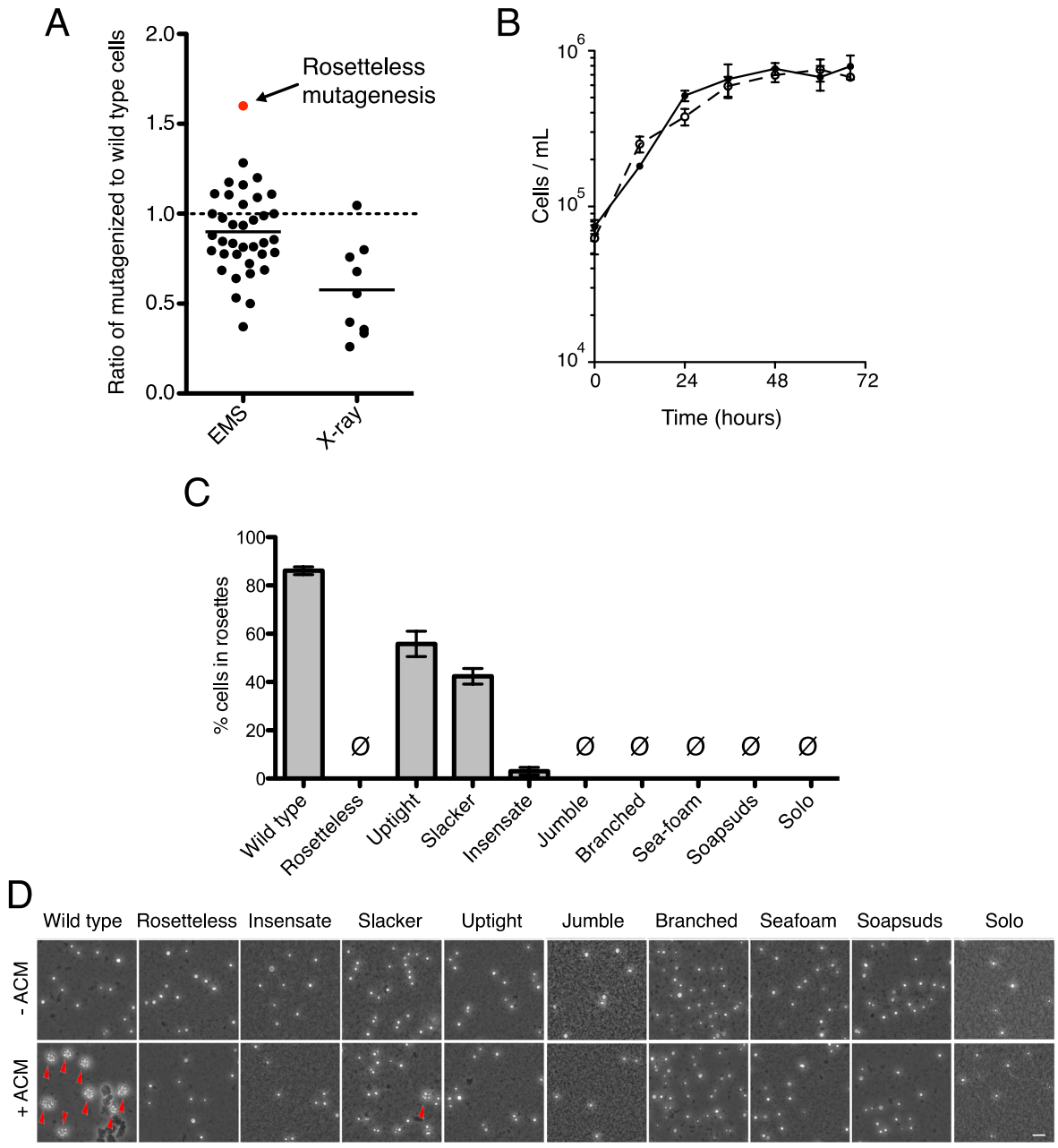

A
Ratio of mutagenized to wild type cells
2.0
1.5
1.0
0.5
0.0
Rosetteless
mutagenesis
EMS
X-ray
B
Cells / mL
$10^6$
$10^5$
$10^4$
0
24
48
72
Time (hours)
C
% cells in rosettes
100
80
60
40
20
0
Ø
Wild type
Rosetteless
Uptight
Slacker
Insensate
Jumble
Branched
Sea-foam
Soapsuds
Solo
D
Wild type
Rosetteless
Insensate
Slacker
Uptight
Jumble
Branched
Seafoam
Soapsuds
Solo
- ACM
+ ACM

**Figure 3.4. Chain colony morphologies of diverse mutants.** *S. rosetta* chain colonies from wild type and mutant cultures grown without exposure to *A. machipongonensis* signals are shown at low magnification to document the morphologies of numerous chain colonies. Each phase-bright circle is a choanoflagellate cell, while the dark specks are bacteria. To the right of each image, the cells of each in-focus chain colony have been false colored blue to identify chains and emphasize chain morphology. Three mutants (Rosetteless, Insensate, and Slacker) exhibit essentially wild type chains. Solo is largely single-celled and rarely forms chains. The five remaining mutants all form chains that are more branched or highly clustered than wild type chains. Scale bar = 50 μm.

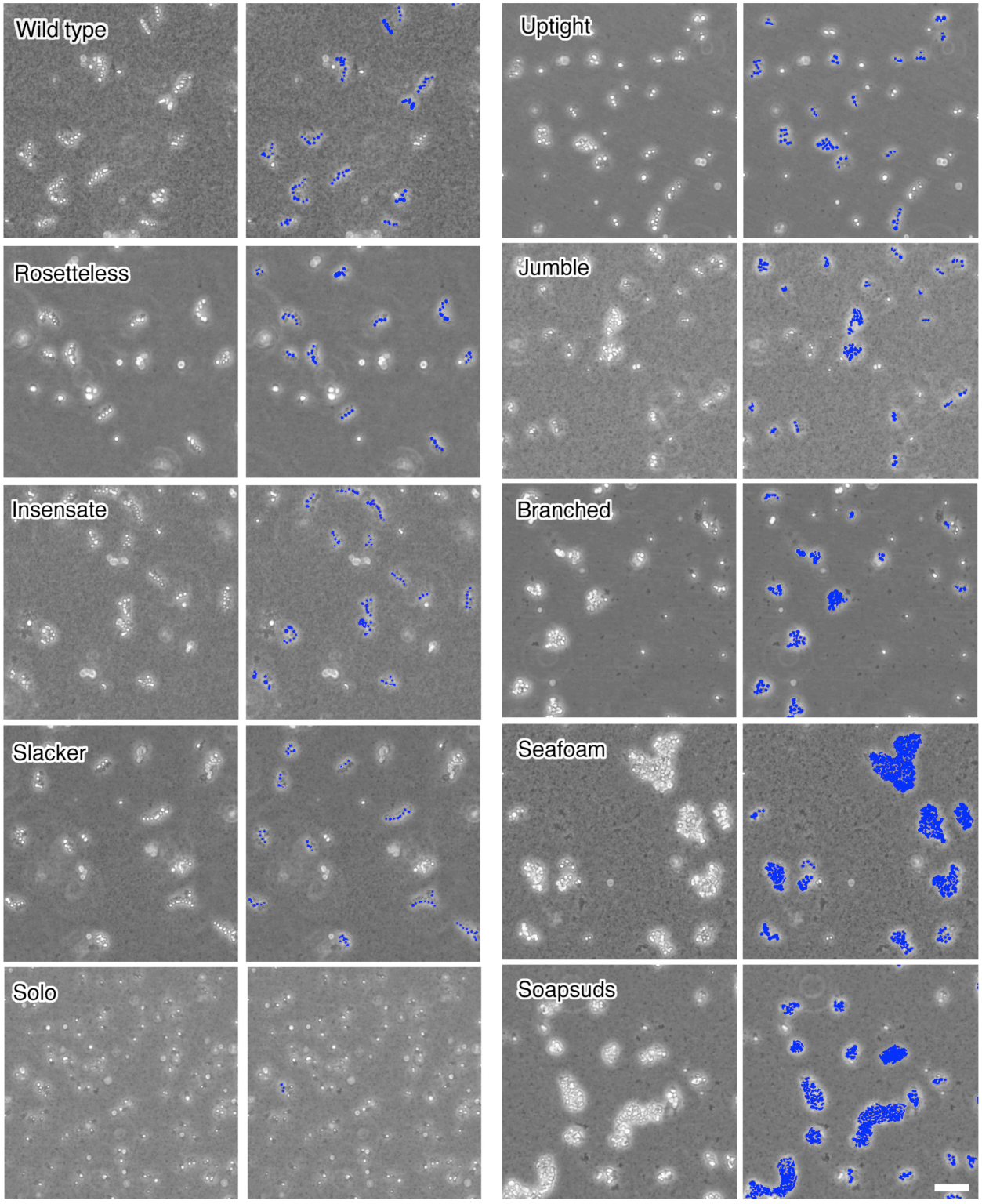
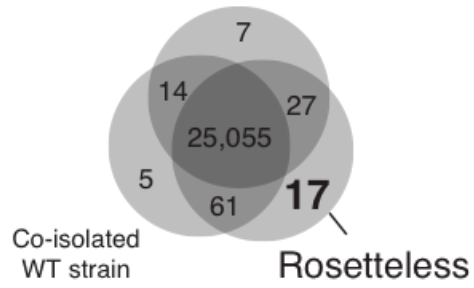

Wild type
Uptight
Rosetteless
Jumble
Insensate
Branched
Slacker
Seafoam
Solo
Soapsuds

**Figure 3.5. Identification of Rosetteless mutations. A)** Venn diagram of the high-quality single nucleotide variants (SNVs) detected in the genome sequences of Rosetteless, the parental strain, and a co-isolated wild-type strain as compared to the *S. rosetta* reference genome. The vast majority of detected SNVs (25,055) were shared among all three isolates. Only 17 unique, high-quality SNVs were detected in Rosetteless. **B)** Of the 17 unique, high-quality Rosetteless SNV predictions (Table 2), only the four SNVs shown here were experimentally verified. The SNV on supercontig 8 (bold) was the only one predicted to alter a coding region. **C)** Cloning and Sanger sequencing confirmed that the supercontig 8: 427,804 SNV was polymorphic between Rosetteless and the parental strain. Note that the Sanger sequencing shown here was from the + strand, while the gene is encoded on the – strand.

A




B

| Supercontig | Location | Position relative to genes |
|---|---|---|
| 2 | 1,595,442 | intron |
| 6 | 1,658,977 | synonymous |
| **8** | **427,804** | **splice donor** |
| 22 | 1,229,741 | intron |

C

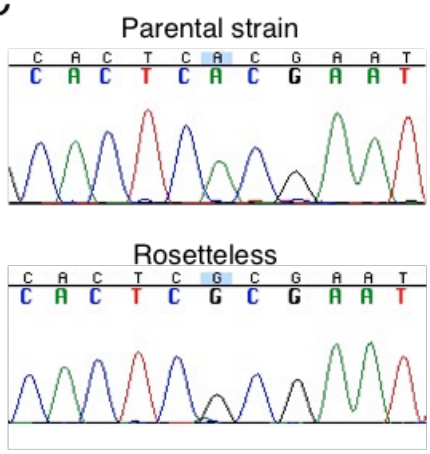
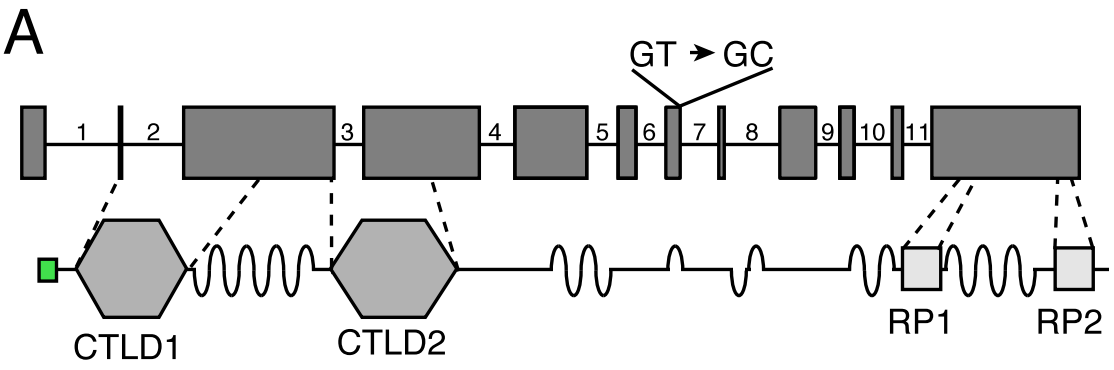


**Table 2. The 17 predicted Rosetteless-specific SNVs.** Seventeen SNV calls that were predicted to be unique to Rosetteless based upon comparisons of Illumina sequencing data among Rosetteless, the parental strain from which it was derived, and an unmutagenized wild type strain (C2E5). Genotyping of Rosetteless and wild type *S. rosetta* revealed whether the putative unique SNV was verified (V, present in Rosetteless but absent in wild type), a false positive (FP, absent from both Rosetteless and wild type), a false negative (FN, present in both Rosetteless and wild type), or a mistakenly-called SNV in a region where the *S. rosetta* reference genome was misassembled (GM).

| Supercontig | Genomic position | Position relative to neighboring gene(s) | Verified? |
|---|---|---|---|
| 1 | 343,191 | Intron | FN |
| 1 | 343,193 | Intron | FN |
| 1 | 401,602 | Intron | FP |
| 1 | 401,606 | Intron | FP |
| 2 | 1,595,442 | Intron | V |
| 5 | 1,669,635 | Synonymous coding | FN |
| 5 | 1,770,007 | Synonymous coding | GM |
| 6 | 1,658,977 | Synonymous coding | V |
| 6 | 1,743,400 | Synonymous coding | FN |
| 8 | 427,804 | Splice donor | V |
| 10 | 1,118,925 | Intergenic repeat | FN |
| 14 | 1,268,439 | Intron | FP |
| 22 | 1,229,741 | Intron | V |
| 29 | 385,389 | Intron | FN |
| 32 | 746,459 | Intron | GM |
| 39 | 5,326 | Intron | FN |
| 39 | 5,329 | Intron | FN |

**Figure 3.6. Defective splicing of the C-type lectin EGD82922 in Rosetteless**. **A)** Top: In the Rosetteless mutant, the only SNV predicted to disrupt a coding sequence interrupts a predicted splice donor in intron 7 (GT→GC) of the 6 kb gene *EGD82922*. *EGD82922* has 12 exons (grey boxes) and 11 introns (numbered). Bottom: *EGD82922* is predicted to encode a 1125 aa protein containing an amino-termal signal peptide (green), two C-type lectin-like domains (CTLD1 and CTLD2), several stretches of serine and threonine repeats (wavy lines), and two high-complexity internal repeats (RP1 and RP2; panel C) of unknown function. **B)** Alignment of the C-type lectin domains (CTLD1 & CTLD2) of the Rosetteless lectin with animal CTLDs from rat surfactant protein A (rat SP-A, 1R13_A) and rat mannose-binding protein (rat MBP, 2MSB_A). Conserved residues used in disulfide linkages are highlighted in blue, mannose-type sugar binding residues are highlighted in red, and asterisks mark positions of calcium ion binding. Other residues of sequence similarity or identity are highlighted in gray. **C)** Alignment of the RP1 and RP2 regions of EGD82922, with similar or identical residues highlighted in gray. **D)** To characterize the splicing of *EGD82922*, primers spanning the exon 5/6 and 10/11 splice junctions were used to amplify mature *EGD82922* from wild type or Rosetteless mutant cDNA. The wild type cDNA gave a single band consistent with its predicted size (516 bp; a), whereas the Rosetteless cDNA gave a variety of bands (including b-d). **E)** The splice isoforms from **(D)** were cloned and sequenced to determine the alterations in splicing. The wild type cDNA sample yielded only one splice isoform (a), which exactly matched the gene model prediction. The Rosetteless mutant cDNA gave several isoforms, the most common of which were: intron 7 retention (b) and the use of alternate splice donors for intron 7 (c and d). Asterisks mark variants of intron 7 that are shorter (*) or longer (**) than in the wild type splice isoform. Splice isoforms b and d are predicted to lead to frame shifts and early stop codons (arrows), while isoform c is predicted to cause a 9 amino acid in-frame insertion.




B

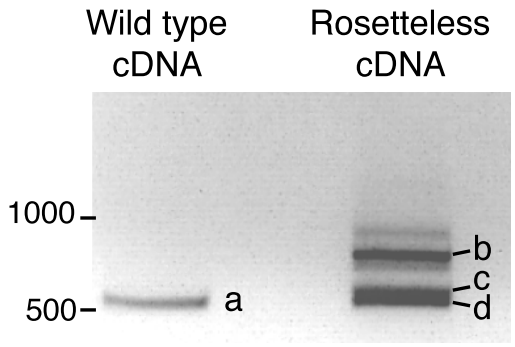
```
                                                                  *   *
   rat MBP  MPFSKVKALCSELRGTVAIPRNAEENKAIQEVAKT---SAFLGITDEVTE-GQFMY
  rat SP-A  VNFDTIKEMCTRAGGNIAVPRTPEENEAIASIAKKYNNYVYLGMIEDQTP-GDFHY
Rtls CTLD1  FSFSSAQLVCSSLSYELATIEDATQNAAIASILEAT-GTAWIGGTRQAGD-SDFQW
Rtls CTLD2  QPFDVAQETCRAYGGELASIHSANENLFAATFLDSDNSVQWLGGARSAEDPTVFEW

                          *    **
   rat MBP  VTGGRLTYSNWKKDEPNDHGSGEDCVT----IVDNGLWNDISCQASHTAVCEFPA
  rat SP-A  LDGASVSYTNWYPGEPRG-QGKEKCVE----MYTDGTWNDRGCLQYRLAVCEF--
Rtls CTLD1  VSGQPFQYTNYATGEPNSAAGTAACLA----INPDGTWFDTDCASEMGVVCT---
Rtls CTLD2  IDGTAFDFQAWYPGEPNNYRGNEACLSQGLKLVQSSQWNDAPCSQSRPFLCK---
```

C

```
RP1 CQLQTVLTCYYRDSGSCSNDGQCVIDADTAFCIPNPRVCTVVLDPTTP
RP2 CSLEESLTCYYRDA-ACEVDNACVLDATSGFCIPHPDTCTLELQAQRP
```

D




E

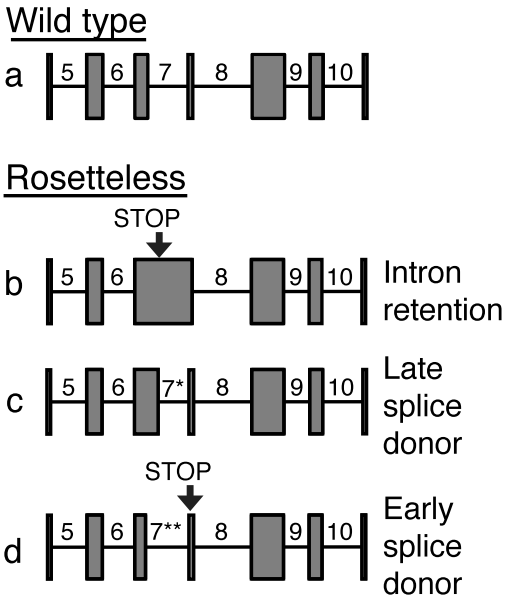


**Figure 3.7. Rosetteless maps to *EGD82922*. A)** Design of the mapping cross. Rosetteless cells were mixed with Isolate B, an *S. rosetta* culture capable of forming rosette colonies. Mating was induced by starvation. To isolate the products of outcrossed mating, cells then underwent clonal isolation, and clonal populations were genotyped to identify outcrossed, diploid heterozygotes. These heterozygotes were expanded and induced to undergo meiosis, after which clonal isolation and genotyping were repeated. Haploid progeny of the cross were homozygous at all three markers. B) 2 x 2 contingency table shows that the Rosetteless phenotype was perfectly linked to the genotype of the supercontig 8: 427,804 candidate splice donor mutation. **C)** Map of all supercontig 8 markers. Top numbers show the genetic distance between the markers and the Rosetteless phenotype in centimorgans (cM). Bottom numbers show marker genomic positions on supercontig 8. Black lines within the central bar show all sites of predicted polymorphism (i.e. possible marker positions) between Rosetteless and Isolate B.

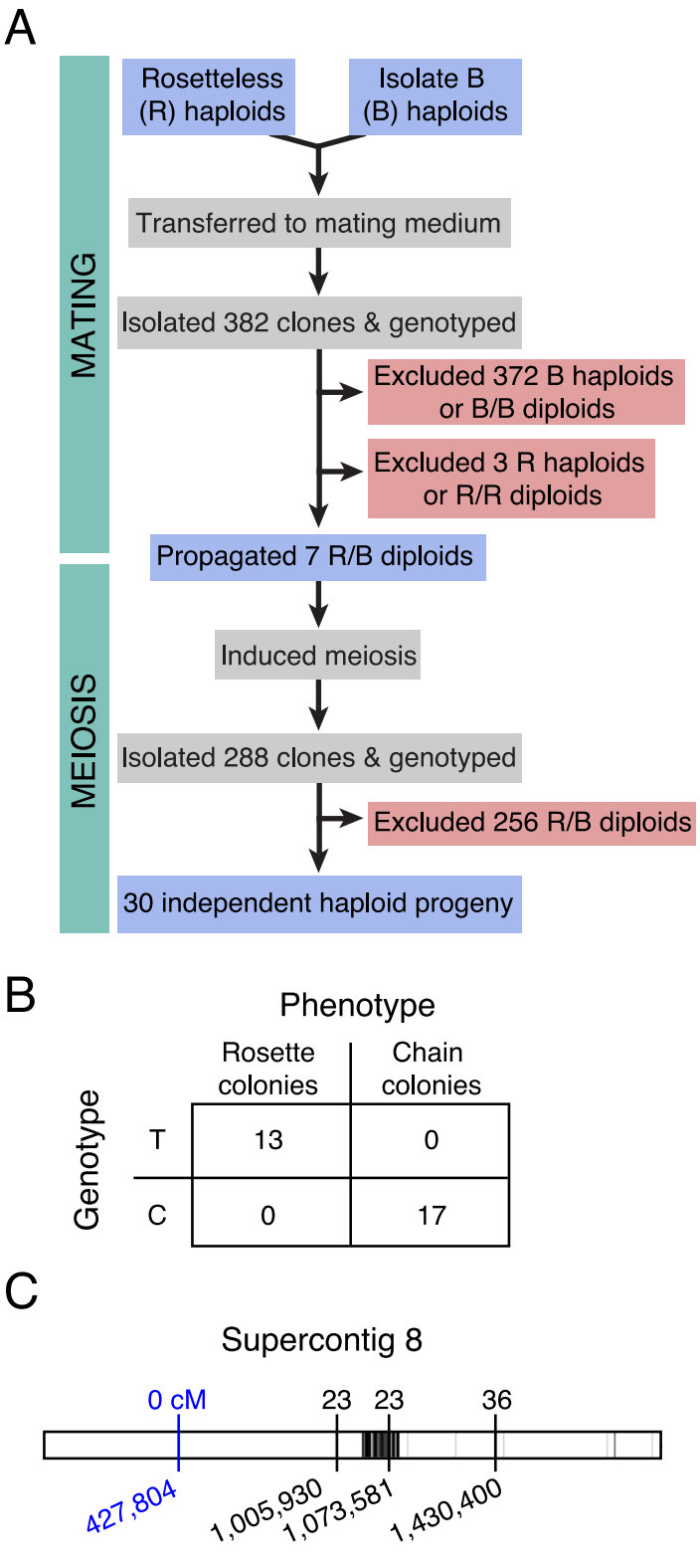

A
MATING
MEIOSIS
Rosetteless (R) haploids
Isolate B (B) haploids
Transferred to mating medium
Isolated 382 clones & genotyped
Excluded 372 B haploids or B/B diploids
Excluded 3 R haploids or R/R diploids
Propagated 7 R/B diploids
Induced meiosis
Isolated 288 clones & genotyped
Excluded 256 R/B diploids
30 independent haploid progeny
B
Phenotype
Rosette colonies
Chain colonies
Genotype
T
13
0
C
0
17
C
Supercontig 8
0 cM
23
23
36
427,804
1,005,930
1,073,581
1,430,400

**Figure 3.8. Map of polymorphisms and markers used in the cross.** Horizontal bars represent the 25 largest supercontigs in the *S. rosetta* genome. The positions of each predicted polymorphism, based on comparisons of the genome sequences of Rosetteless and Isolate B, are mapped semi-transparently in black, such that regions with high variant density are darker than those with lower variant density. Red symbols show positions genotyped in all cross isolates, including the perfectly linked supercontig 8: 427,804 mutation (*). Note that the supercontig 8: 427,804 mutation lies within a haplotype block with little to no polymorphism between Rosetteless and Isolate B.

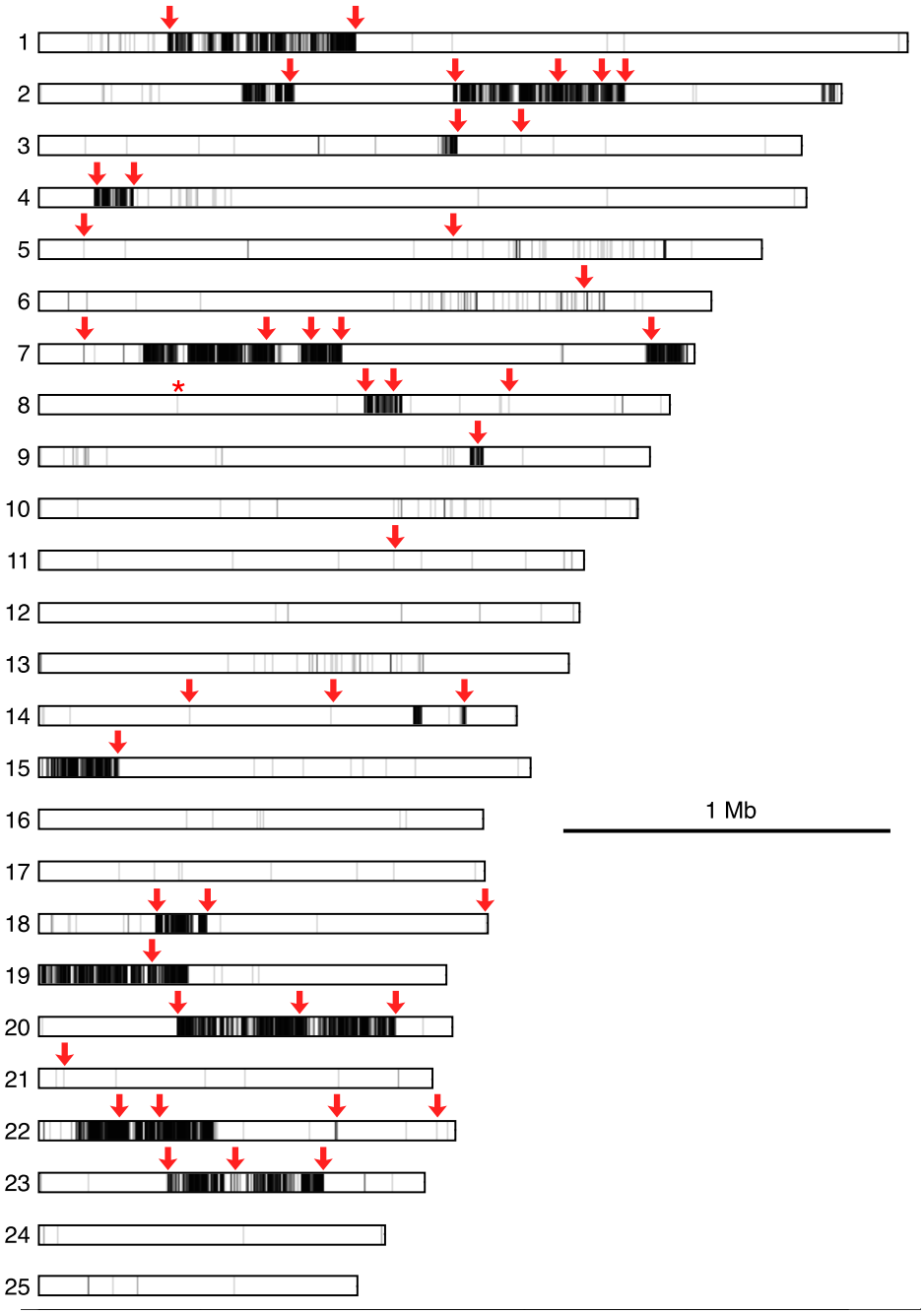


**Figure 3.9. A linkage map for *S. rosetta*.** Linkage map of *S. rosetta*, based on the genotypes of the 30 Rosetteless-Isolate B cross progeny. The color of each box represents the genetic distance between two markers (shades of gray; see legend) or the genetic distance between a marker and the Rosetteless phenotype (shades of blue; see legend), thresholded such that non-significant distances are colored white (one-tailed Fisher's exact test, p>0.05). The boxes representing the four SNVs that were unique to Rosetteless are outlined in red. Gray lines show the boundaries between assembled linkage groups, labeled at the top of the figure. The supercontig locations of each marker are listed, in order, on the left and bottom sides of the figure. Note that the Rosetteless phenotype (blue) is significantly linked only to linkage group 5/supercontig 8 and is perfectly linked to the splice donor mutation.

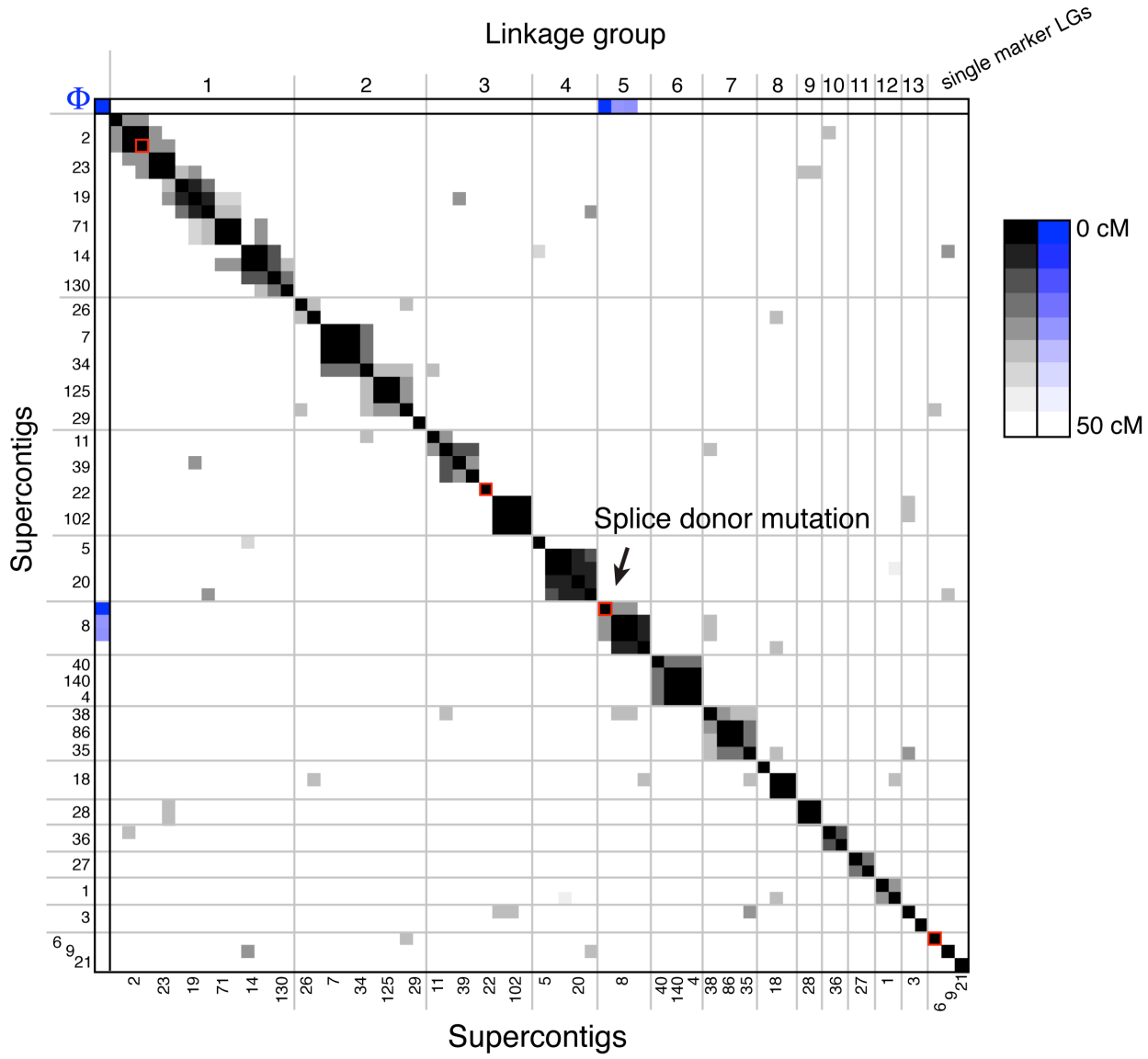


**Table 3. Primers used for genotyping and assessing splicing.** Primers used to genotype microsatellites by fragment analysis (MF), genotype microsatellites by 2% agarose gel (MG), verify and genotype SNPs using Sanger sequencing (SNP), assess the splicing of *EGD82922* from cDNA (SPL), or sequence the coding region of *EGD82922* from gDNA (C). One primer was directly fluorescently labeled with 6-carboxyfluorescein (6FAM).

| Primer name | Genomic position | Primer use | Primer sequence |
|---|---|---|---|
| gt_indel_7 | sc2: 1,730,188 | MF | TGTAAAACGACGGCCAGTCACAACAGAGAAGCCGGATG<br>GCACATGAGCGCTTAGAATTC |
| gt_indel_2 | sc7: 855,120 | MF | TGTAAAACGACGGCCAGTGCCTTATCAACTCCCTGTGC<br>GCGTGATTGCCTTCGTTTCA |
| gt_indel_9 | sc8: 1,073,581 | MF | 6FAM-ATGGGCGCCTTGAACAAAC<br>TGCTCCAGATGTCAGCTTCA |
| gt2 | sc8: 427,804 | SNP | GGACACGCGAGACATAGACA<br>TTTCATTTGGGATACGCACA |
| gt3 | sc22: 1,229,741 | SNP | GAAATTCACGGTGACGGACT<br>ACGCAATCGGTAAAGGTCAC |
| gt15 | sc1: 401,602 | SNP | CAATTGCACCAAGCAAGAGA<br>TGCATGTCTCTCACGAGTCC |
| gt18 | sc5: 1,770,007 | SNP | GCTATGGTGGCGAGTACGTG<br>CTGATAGCACTTGCGGCATA |
| gt23 | sc29: 385,389 | SNP | GGGCATCCAACACATACACA<br>TCGTCCGTTCAAACTCTTCC |
| indel1 | sc22: 391,927 | MG | GGAACAGGGAGCCACAACTA<br>ACGCTCGCCTACTACGACTC |
| Rtls_2L | Spans exons 5 and 6 of *EGD82922* | SPL | CCGCATTTGAGGCCGAAT |
| Rtls_1R | Spans exons 10 and 11 of *EGD82922* | SPL & C | CAGCATCAATCACGCACTGA |
| Rtls_4L | sc8: 430573 | C | GCTATGTGGAGTTCTCCGATG |
| Rtls_2R | sc8: 426601 | C | TGAACGATGCACTGGGTTG |
| Rtls_L1 | sc8: 431369 | C | CTGCGCCCGTTGATGTTG |
| Rtls_L3 | sc8: 428490 | C | GTGGCCCTCGCTTGTGGT |
| Rtls_L4 | sc8: 431240 | C | CACACACAGGTCGTTTGCTT |
| Rtls_L5 | sc8: 429836 | C | CGACAACGACGAGAGAGACA |
| Rtls_L6 | sc8: 429763 | C | CAATCCAGAAGTGCCAGGTT |
| Rtls_R1 | sc8: 428389 | C | AAAACGCGTTCCTCCACTGC |
| Rtls_R2 | sc8: 425565 | C | GTGGTGGAAGGTGCAACAGG |
| Rtls_R3 | sc8: 429763 | C | AACCTGGCACTTCTGGATTG |
| Rtls_R4 | sc8: 428351 | C | GATTGCAGTGAGGGGTTCAC |
| Rtls_R6 | sc8: 427237 | C | CGTGGACGATGTTGTCCTAA |

# Chapter 4

# Future directions and closing remarks

## Future directions

### Characterization of sexual reproduction in *S. rosetta*

Through my observations of ploidy switching, haplotype blocks, and cell fusion in *S. rosetta*, I demonstrated that this species undergoes both sex and meiosis and that the sexual life cycle can be controlled in the lab (Chapter 2). These inferences were further solidified by the Mendelian inheritance I observed in the first experimental cross of *S. rosetta* (Fig. 3.7). The existence of a sexual cycle and anisogamous mating in choanoflagellates is exciting, in part because it may be relevant to the ancestry of sperm/egg reproduction in animals. However, before we can address such hypotheses, there is still much to learn about how *S. rosetta* undergoes its sexual life cycle.

Perhaps the most important piece of missing information is the relationship between the *S. rosetta* sexual cycle, ploidy, and the previously characterized morphological cell types (Fig. 1.2) [44]. Understanding this relationship is a prerequisite to evaluating whether there is a relationship between *S. rosetta* multicelluarity and sexual reproduction. During the Rosetteless-Isolate B cross (Fig. 3.7), the haploid cells were chains and rosettes at the beginning of the cross, then turned thecate as they became starved and diploid, and turned into chains and rosettes again when the cells became haploid following meiosis. However, although this tells us that chains and rosettes can be haploid and thecate cells can be diploid, it does not tell us if the reverse is possible or if morphological transitions require a shift in ploidy. Currently, the main challenge in relating ploidy and morphological cell types is that ploidy can only be measured on a population-wide scale, and it is unknown which cell morphologies correspond to which level of DNA content. At the same time, it is challenging to obtain *S. rosetta* cultures that are exclusively one morphological cell type, and so it is not currently feasible to measure the ploidy of pure cultures. Future studies may be able to link ploidy and morphological cell type through the establishment of methods to measure ploidy on a single-cell level (e.g. fluorescent *in situ* hybridization) and/or the discovery of markers for the morphological cell types that can be sorted and analyzed by flow cytometry.

Another area in need of future research concerns *S. rosetta* gametogenesis. Although I found that *S. rosetta* can self-fertilize via the generation of both male and female gamete types from clonal lines (Fig. 2.1 and 2.7), we do not know how *S. rosetta* generates the male and female gamete cells. It will be interesting to learn whether gametes are generated through a specialized cell division or whether asexually reproducing haploid cells can simply differentiate into gametes. If it occurs through a specialized cell division, can all *S. rosetta* cells generate both male and female gamete types, or are there subpopulations in the culture that are specialized for the production of a single gamete type? Such questions

could probably be answered through careful time-lapse microscopy, especially if microfluidic devices can be used to prevent cells from swimming out of the field of view.

The process of *S. rosetta* gamete recognition and fusion also remains quite mysterious. I was able to observe a few, rare cell fusion events (Fig. 2.7), but the vast majority of gametes adhered to one another for a substantial time and then parted without fusing. Is cell fusion merely a slow and inefficient process? Or is there some kind of signaling that occurs between male and female cells before cell fusion is permitted? The observation of a bridge of DNA between tightly adherent male and female gametes (Fig 2.8), although anecdotal, suggests that information at some level is exchanged prior to fusion. I also observed that many populations in the process of sexual reproduction have a significant number of 3n nuclei (Fig. 2.1B and 2.3E), perhaps indicating that more than one fusion event occurred per cell. Thus, assuming there is some regulation of sexual cell fusion, this may nevertheless be a somewhat sloppy process. Finally, although I observed seven cell fusion events that were all between a male and female gamete, it does not rule out the possibility that isogamous fusion could also occur. Here again, further time-lapse microscopy could be informative through the observation of additional cell fusion events, especially in combination with approaches to make choanoflagellate mating more synchronous within the culture.

**Suggestions for optimizing crosses in choanoflagellates**

Through the use of different culture media, I was able to induce sex and meiosis to perform the first experimental cross in a choanoflagellate. Although we were only able to induce a relatively low rate of outcrossed mating (1.8%) and meiosis (11%) in this cross, I suggest the following modifications could increase cross efficiency further. First, in this cross mating was induced through an 11-day incubation in nutrient poor media, during which time the choanoflagellates grazed down the bacteria present and themselves became starved. This long incubation provided a large window for the Rosetteless/Isolate B ratio to diverge from the initial 50:50 mix, which likely decreased the rate of outcrossed mating as compared to self-mating. A more rapid starvation procedure, perhaps by actively eliminating the bacteria with antibiotics, would likely allow for more synchronous and more random mating. Second, the *S. rosetta* cultures crossed here were cultured with different bacterial species and grown in different culture media prior to the cross; crosses involving more similar parental strains may also allow for more synchronous mating. Third, in Fig. 2.1, I observed a nearly complete diploid-to-haploid switch using *S. rosetta* from the SrEpac strain grown in HN media, suggesting the crosses between haploid cells in this background could exhibit much more efficient meiosis. Finally, identification of outcrossed clones by genotyping, while effective, is also time-intensive and relatively low-throughput. The use of parental strains carrying either selectable markers or fluorescent markers could greatly simplify choanoflagellate crosses through drug selection of or flow sorting.

**A future for choanoflagellate genetics**

At this point, many of the tools are in place for using forward genetics to understand rosette formation or any other phenotype of interest in *S. rosetta*. I have investigated multiple methods for mutagenesis, each of which seems to cause a mild decrease on cell number on average (Fig. 3.3). I have developed a simple, albeit somewhat labor intensive, method for isolating mutant clones by visual screening (Fig. 3.1). I have discovered the *S. rosetta* sexual life cycle and a means for controlling it in experimental crosses through the use of different culture media (Fig. 2.1 and 3.7). And finally, through extensive genotyping of the first cross, I have developed an initial set of validated markers genome-wide and generated a preliminary linkage map for *S. rosetta* (Fig. 3.9).

So what can we do with these tools and how can they be improved further? One take home message from the Rosetteless sequencing was that this mutant contained very few mutations (Fig. 3.5)—so few that it seemed likely that Rosetteless was actually a spontaneous mutant that was largely unaffected by the EMS. The enhanced cell death and elevated rate of mutant recovery following X-ray as opposed to EMS mutagenesis (Fig. 3.1B and 3.3A) suggests that our X-ray protocol was more effective, but further optimization of both protocols will likely prove useful. At the same time, the small number of Rosetteless mutations greatly facilitated mapping, and so it may be desirable for future studies to pursue spontaneous mutants.

Although we pursued a visual screening method as the simplest option for a proof-of-principle screen, future studies that can isolate mutants through selection are likely to be higher throughput and less labor intensive. Such selections could be based on size (e.g. separation of rosettes from single cells through a physical filter), differential staining with a fluorescent marker (e.g. using flow sorting), or viability (e.g. using a cytotoxic drug). The establishment of *S. rosetta* lines carrying different selectable markers could also prove useful for facilitating the downstream steps of a cross and for evaluating mutant phenotypes.

Another current limitation in mapping mutations in *S. rosetta* is the lack of highly polymorphic strains to use in crosses. As large regions of the genome are identical in all of the haploid strains available to date (Fig. 3.8), we are unable to genotype cross isolates in these regions and thus cannot detect linkage between the 'marker deserts' and other parts of the genome. Until and unless more polymorphic strains can be used, we will still be limited in the markers available to us and the linkage map will remain incomplete. There are several possible ways to obtain more polymorphic strains. First, if *S. rosetta* were re-isolated from the wild, this would be a natural choice for use in crosses. Second, from the genome sequencing of isolate A (Fig. 2.4), we know that this strain has additional regions of polymorphism that are not found in isolate B, and I infer that the original *S. rosetta* isolate was likely an outcrossed heterozygote, carrying even more polymorphism relative to SrEpac. Thus, the generation of another haploid, monoxenic line of *S. rosetta* from the original ATCC-50818 culture may thus yield an appropriate, more polymorphic strain. However, future researchers considering approaches 1 or 2 should be forewarned that the generation of an *S. rosetta* monoxenic line from a diverse microbial culture is both time and labor intensive. A third approach would involve subjecting an existing monoxenic line to

extensive mutagenesis, in the hopes that mutations could be generated in the 'marker deserts'. Such an approach would probably require the use of EMS or another mutagen that generates point mutations, as large chromosomal deletions or inversions would be undesirable for recombination mapping. We still have much to learn about how mutagenesis works in *S. rosetta*, so it remains unclear which mutagen(s) would be appropriate or what doses would be effective. For now, it is probably acceptable to use the existing Isolate B line for mapping, but the isolation of a new polymorphic line will be useful in the long term.

With these tools, I hope that we will eventually be able to fully elucidate the rosette formation pathway in *S. rosetta*. Previously, candidate gene approaches and transcriptional profiling proved relatively ineffective for identifying rosette formation genes [39], but with the forward genetics approach we can reasonably infer that we have now found the first gene in this pathway: the *rosetteless* C-type lectin. As none of the other eight mutants we isolated have mutations in the *rosetteless* locus, there seems to be much more out there to discover! The continuation of rosette formation screens in addition to the further study of these existing mutants will be important first steps towards identifying new genes.

## Closing remarks

Although choanoflagellates can yield evolutionary insights unobtainable in other organisms, it can be challenging to work with a new model organism that has few molecular tools, a modest bibliography of previous work, and a small research community. The lack of transgenic tools in particular has limited our ability to understand either the genetic basis of choanoflagellate multicellularity or the functions of interesting genes in choanoflagellates, both of which are necessary for understanding the role of co-option and *de novo* evolution in the origin of animal multicellularity.

And yet, all is not lost! It was precisely these limitations that inspired me to pursue forward genetics as an alternative (and arguably better) approach for linking genotype and phenotype in choanoflagellates. Although performing choanoflagellate genetics is not yet easy, the fact that this genetic screen was possible at all is really quite remarkable. When I began my thesis, the *S. rosetta* genome had just been sequenced, we had single monoxenic line of the cells in lab (Px1), and we knew that as-yet-uncharacterized bacterial lipids induced rosette formation. We didn't know the chromosome number or the ploidy of *S. rosetta*, much less that there was a sexual cycle occurring, unbeknownst to us, in our own cultures. No one had ever generated or isolated a mutant of any choanoflagellate species. And obviously, no phenotypic defects had ever been mapped to a gene in an experimental cross. So, I think we have come a long way. The project has been much more successful than I ever anticipated, due, I think, to a combination of personal stubbornness, an amazing and hard working undergraduate researcher, lots of good ideas and advice from those around me, and frankly, good luck.

Thanks to everyone for helping me to do awesome science. I've had a great time.

## References


1. King, N. (2004). The Unicellular Ancestry of Animal Development. Developmental Cell *7*, 313–325.

2. Knoll, A. H. (2011). The Multiple Origins of Complex Multicellularity. Annual Review of Earth and Planetary Sciences *39*, 217–239.

3. Steenkamp, E. T., Wright, J., and Baldauf, S. L. (2006). The protistan origins of animals and fungi. Molecular Biology and Evolution *23*, 93–106.

4. Morris, S. C. (1993). The Fossil Record and the Early Evolution of the Metazoa. Nature *361*, 219–225.

5. Martindale, M. Q. (2005). The evolution of metazoan axial properties. Nat Rev Genet *6*, 917–927.

6. Nielsen, C. (2008). Six major steps in animal evolution: are we derived sponge larvae? Evol Dev *10*, 241–257.

7. 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., et al. (2009). The origin of Metazoa: a transition from temporal to spatial cell differentiation. Bioessays *31*, 758–768.

8. Kalinka, A. T., and Tomancak, P. (2012). The evolution of early animal embryos:conservation or divergence? Trends Ecol Evol (Amst) *27*, 385–393.

9. Richter, D. J., and King, N. (2013). The genomic and cellular foundations of animal origins. Annual Review of Genetics *47*, 509–537.

10. Nichols, S. A., Dirks, W., Pearse, J. S., and King, N. (2006). Early evolution of animal cell signaling and adhesion genes. Proc Natl Acad Sci USA *103*, 12451–12456.

11. Adamska, M., Degnan, B. M., Green, K., and Zwafink, C. (2011). What sponges can tell us about the evolution of developmental processes. Zoology *114*, 1–10.

12. Srivastava, M., Simakov, O., Chapman, J., Fahey, B., Gauthier, M. E. A., Mitros, T., Richards, G. S., Conaco, C., Dacre, M., Hellsten, U., et al. (2010). The Amphimedon queenslandica genome and the evolution of animal complexity. Nature *466*, 720–726.

13. Erwin, D. H. (2009). Early origin of the bilaterian developmental toolkit. Philosophical Transactions of the Royal Society B: Biological Sciences *364*, 2253–2261.

14. Pires-daSilva, A., and Sommer, R. J. (2003). The evolution of signalling pathways in animal development. Nat Rev Genet *4*, 39–49.

15. King, N. (2005). Choanoflagellates. Curr Biol *15*, R113–4.

16. Lang, B. F., O'Kelly, C., Nerad, T., Gray, M. W., and Burger, G. (2002). The closest unicellular relatives of animals. Current Biology *12*, 1773–1778.

17. Carr, M., Leadbeater, B. S. C., Hassan, R., Nelson, M., and Baldauf, S. L. (2008). Molecular phylogeny of choanoflagellates, the sister group to Metazoa. Proceedings of the National Academy of Sciences *105*, 16641–16646.

18. Nitsche, F., Carr, M., Arndt, H., and Leadbeater, B. S. C. (2011). Higher level taxonomy and molecular phylogenetics of the Choanoflagellatea. Journal of Eukaryotic Microbiology *58*, 452–462.

19. Paps, J., Medina-Chacón, L. A., Marshall, W., Suga, H., and Ruiz-Trillo, I. (2012). Molecular Phylogeny of Unikonts: New Insights into the Position of Apusomonads and Ancyromonads and the Internal Relationships of Opisthokonts. Protist *164*, 1–11.

20. Philippe, H., Derelle, R., Lopez, P., Pick, K., Borchiellini, C., Boury-Esnault, N., Vacelet, J., Renard, E., Houliston, E., Queinnec, E., et al. (2009). Phylogenomics Revives Traditional Views on Deep Animal Relationships. Current Biology *19*, 706–712.

21. Ruiz-Trillo, I., Roger, A. J., Burger, G., Gray, M. W., and Lang, B. F. (2008). A phylogenomic investigation into the origin of metazoa. Molecular Biology and Evolution *25*, 664–672.

22. Torruella, G., Derelle, R., Paps, J., Lang, B. F., Roger, A. J., Shalchian-Tabrizi, K., and Ruiz-Trillo, I. (2012). Phylogenetic relationships within the Opisthokonta based on phylogenomic analyses of conserved single-copy protein domains. Molecular Biology and Evolution *29*, 531–544.

23. James-Clark, H. (1867). Conclusive proofs of the animality of the ciliate sponges, and of their affinities with the Infusoria flagellata. The Annals and Magazine of Natural History *19*, 13–18.

24. Kent, W. S. (1880). A manual of the Infusoria: including a description of all known Flagellate, Ciliate, and Tentaculiferous Protozoa, British and foreign, and an account of the organization and the affinities of the sponges (London: David Bogue).

25. Goldberg, W. M., and Taylor, G. T. (1989). Cellular structure and ultrastructure of the black coral *Antipathes aperta*: 1. Organization of the tentacular epidermis and nervous system. Journal of Morphology *202*, 239–253.

26. Lyons, K. M. (1973). Collar cells in planula and adult tentacle ectoderm of the solitary coral *Balanophyllia regia* (Anthozoa Eupsammiidae). Z Zellforsch Mikrosk Anat *145*, 57–74.

27. Martinez, A., Lopez, J., Villaro, A. C., and Sesma, P. (1991). Choanocyte-like cells in the digestive system of the starfish *Marthasterias glacialis* (Echinodermata). Journal of Morphology *208*, 215–225.

28. Milanesi, C. (1971). Choanocytes in endostyle of *Botryllus schlosseri* (Pallas) stolidobranchiata. Journal of submicroscopic cytology *{3}*, 359–363.

29. Norrevang, A. (1964). Choanocytes in the skin of *Harrimania kupfferi* (Enteropneusta). Nature *204*, 398–399.

30. Walker, C. W. (1979). Ultrastructure of the somatic portion of the gonads in asteroids, with emphasis on flagellated-collar cells and nutrient transport. Journal of Morphology *162*, 127–161.

31. Alegado, R. A., and King, N. (2014). Bacterial influences on animal origins. CSHL Perspectives *In Press.*

32. Mah, J. L., Christensen-Dalsgaard, K. K., and Leys, S. P. (2014). Choanoflagellate and choanocyte collar-flagellar systems and the assumption of homology. Evol Dev *16*, 25–37.

33. King, N., and Carroll, S. B. (2001). A receptor tyrosine kinase from choanoflagellates: molecular insights into early animal evolution. Proceedings of the National Academy of Sciences *98*, 15032–15037.

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

35. Abedin, M., and King, N. (2008). The Premetazoan Ancestry of Cadherins. Science *319*, 946–948.

36. Manning, G., Young, S. L., Miller, W. T., and Zhai, Y. (2008). The protist, *Monosiga brevicollis*, has a tyrosine kinase signaling network more elaborate and diverse than found in any known metazoan. Proceedings of the National Academy of Sciences *105*, 9674–9679.

37. Young, S. L., Diolaiti, D., Conacci-Sorrell, M., Ruiz-Trillo, I., Eisenman, R. N., and King, N. (2011). Premetazoan ancestry of the Myc-Max network. Molecular Biology and Evolution *28*, 2961–2971.

38. King, N., Westbrook, M. J., Young, S. L., Kuo, A., Abedin, M., Chapman, J., Fairclough, S., Hellsten, U., Isogai, Y., Letunic, I., et al. (2008). The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. Nature *451*, 783–788.

39. Fairclough, S. R., Chen, Z., Kramer, E., Zeng, Q., Young, S., Robertson, H. M., Begovic, E., Richter, D. J., Russ, C., Westbrook, M. J., et al. (2013). Premetazoan genome

evolution and the regulation of cell differentiation in the choanoflagellate Salpingoeca rosetta. Genome Biology *14*, R15.

40. Li, W., Young, S. L., King, N., and Miller, W. T. (2008). Signaling properties of a non-metazoan Src kinase and the evolutionary history of Src negative regulation. J Biol Chem *283*, 15491–15501.

41. Li, W., Scarlata, S., and Miller, W. T. (2009). Evidence for Convergent Evolution in the Signaling Properties of a Choanoflagellate Tyrosine Kinase. Biochemistry *48*, 5180–5186.

42. Schultheiss, K. P., Craddock, B. P., Tong, M., Seeliger, M., and Miller, W. T. (2013). Metazoan-like signaling in a unicellular receptor tyrosine kinase. BMC Biochem. *14*.

43. Leadbeater, B. S. (1983). Life-history and ultrastructure of a new marine species of *Proterospongia* (Choanoflagellida). Journal of the Marine Biological Association of the United Kingdom *63*, 135–160.

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

45. Buss, L. W. (1987). Evolution of Individuality (Princeton, NJ: Princeton University Press).

46. Bonner, J. T. (2009). First signals: the evolution of multicellular development (Princeton, NJ: Princeton University Press).

47. Fairclough, S. R., Dayel, M. J., and King, N. (2010). Multicellular development in a choanoflagellate. Current Biology *20*, R875–6.

48. Alegado, R. A., Brown, L. W., Cao, S., Dermenjian, R. K., Zuzow, R., Fairclough, S. R., Clardy, J., and King, N. (2012). A bacterial sulfonolipid triggers multicellular development in the closest living relatives of animals. eLife *1*, e00013.

49. Haeckel, E. (1874). Die Gastraea-Theorie, die phylogenetische Classification des Thierreichs und die Homologie der Keimblatter. Jenaische Zeitschrift für Naturwissenschaft *8*, 1–55.

50. Ruvinsky, A. (1997). Sex, meiosis and multicellularity. Acta Biotheor *45*, 127–141.

51. Grosberg, R. K., and Strathmann, R. R. (2007). The evolution of multicellularity: A minor major transition? Annu Rev Ecol Evol Syst *38*, 621–654.

52. Kuzdzal-Fick, J. J., Fox, S. A., Strassmann, J. E., and Queller, D. C. (2011). High relatedness is necessary and sufficient to maintain multicellularity in Dictyostelium. Science *334*, 1548–1551.

53. Libby, E., and B Rainey, P. (2013). A conceptual framework for the evolutionary origins of multicellularity. Phys Biol *10*, 035001.

54. Peacock, L., Bailey, M., Carrington, M., and Gibson, W. (2014). Meiosis and Haploid Gametesin the Pathogen Trypanosoma brucei. Curr Biol *24*, 181–186.

55. Lee, S. C., Ni, M., Li, W., Shertz, C., and Heitman, J. (2010). The Evolution of Sex: a Perspective from the Fungal Kingdom. Microbiology and Molecular Biology Reviews *74*, 298–340.

56. Cooper, M. A., Adam, R. D., Worobey, M., and Sterling, C. R. (2007). Population Genetics Provides Evidence for Recombination in Giardia. Current Biology *17*, 1984–1988.

57. O'Gorman, C. M., Fuller, H. T., and Dyer, P. S. (2009). Discovery of a sexual cycle in the opportunistic fungal pathogen Aspergillus fumigatus. Nature *457*, 471–474.

58. Bennett, R. J., and Johnson, A. D. (2005). Mating in Candida albicans and the search for a sexual cycle. Annu Rev Microbiol *59*, 233–255.

59. Welch, D. M., and Meselson, M. (2000). Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange. Science *288*, 1211–1215.

60. Carr, M., Leadbeater, B. S. C., and Baldauf, S. L. (2010). Conserved Meiotic Genes Point to Sex in the Choanoflagellates. Journal of Eukaryotic Microbiology *57*, 56–62.

61. Sager, R., and Granick, S. (1954). Nutritional control of sexuality in Chlamydomonas reinhardi. J. Gen. Physiol. *37*, 729–742.

62. Miller, J. J. (1957). The metabolism of yeast sporulation. II. Stimulation and inhibition by monosaccharides. Can. J. Microbiol. *3*, 81–90.

63. Blackburn, S. I., Hallegraeff, G. M., and Bolch, C. J. (1989). Vegetative reproduction and sexual life cycle of the toxic dinoflagellate *Gymnodinium catenatum* from Tasmania, Australia. Journal of Phycology *25*, 577–590.

64. King, N., Young, S. L., Abedin, M., Carr, M., and Leadbeater, B. S. C. (2009). Starting and maintaining Monosiga brevicollis cultures. Cold Spring Harbor Protocols *2009*, pdb.prot5148.

65. Atlas, R. M. (2004). Handbook of Microbiological Media 3rd ed. (CRC Press).

66. Alegado, R. A., Grabenstatter, J. D., Zuzow, R., Morris, A., Huang, S. Y., Summons, R. E., and King, N. (2013). *Algoriphagus machipongonensis* sp. nov., co-isolated with a colonial choanoflagellate. International Journal of Systematic and Evolutionary Microbiology *63*, 163–168.

67. King, N., Young, S. L., Abedin, M., Carr, M., and Leadbeater, B. S. C. (2009). Isolation

of Single Choanoflagellate Cells from Field Samples and Establishment of Clonal Cultures. Cold Spring Harbor Protocols *2009*, pdb.prot5147–pdb.prot5147.

68. King, N., Young, S. L., Abedin, M., Carr, M., and Leadbeater, B. S. C. (2009). Separation of Choanoflagellate and Bacterial Genomic DNA. Cold Spring Harbor Protocols *2009*, pdb.prot5154–pdb.prot5154.

69. Nedashkovskaya, O. I., Kim, S. B., Vancanneyt, M., Lysenko, A. M., Shin, D. S., Park, M. S., Lee, K. H., Jung, W. J., Kalinovskaya, N. I., Mikhailov, V. V., et al. (2006). *Echinicola pacifica* gen. nov., sp. nov., a novel flexibacterium isolated from the sea urchin *Strongylocentrotus intermedius*. International Journal of Systematic and Evolutionary Microbiology *56*, 953–958.

70. Shoemaker, N. B., Barber, R. D., and Salyers, A. A. (1989). Cloning and characterization of a Bacteroides conjugal tetracycline-erythromycin resistance element by using a shuttle cosmid vector. J Bacteriol *171*, 1294–1302.

71. Price, C., Reardon, E., and Guillard, R. (1978). Collection of Dinoflagellates and Other Marine Microalgae by Centrifugation in Density Gradients of a Modified Silica Sol. Limnology and Oceanography *23*, 548–553.

72. Woo, J. L., and Berk, A. J. (2007). Adenovirus ubiquitin-protein ligase stimulates viral late mRNA nuclear export. J. Virol. *81*, 575–587.

73. Wersto, R. P., Chrest, F. J., Leary, J. F., Morris, C., Stetler-Stevenson, M., and Gabrielson, E. (2001). Doublet discrimination in DNA cell-cycle analysis. Cytometry *46*, 296–306.

74. Uchida, M., Sun, Y., McDermott, G., Knoechel, C., Le Gros, M. A., Parkinson, D., Drubin, D. G., and Larabell, C. A. (2011). Quantitative analysis of yeast internal architecture using soft X-ray tomography. Yeast *28*, 227–236.

75. Hutter, K. J., and Eipel, H. E. (1978). Flow cytometric determinations of cellular substances in algae, bacteria, moulds and yeasts. Antonie Van Leeuwenhoek *44*, 269–282.

76. Lohse, M., Bolger, A. M., Nagel, A., Fernie, A. R., Lunn, J. E., Stitt, M., and Usadel, B. (2012). RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. Nucleic Acids Research *40*, W622–W627.

77. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

78. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

79. Bradley, R. K., Roberts, A., Smoot, M., Juvekar, S., Do, J., Dewey, C., Holmes, I., and Pachter, L. (2009). Fast statistical alignment. PLoS Comput Biol *5*, e1000392.

80. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189–1191.

81. Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat Meth *9*, 671–675.

82. King, N., Young, S. L., Abedin, M., Carr, M., and Leadbeater, B. S. C. (2009). Visualizing the Subcellular Localization of Actin, Beta-tubulin, and DNA in Monosiga brevicollis. Cold Spring Harbor Protocols *2009*, pdb.prot5150–pdb.prot5150.

83. Cousin, A., Heel, K., Cowling, W. A., and Nelson, M. N. (2009). An efficient high-throughput flow cytometric method for estimating DNA ploidy level in plants. Cytometry *75A*, 1015–1019.

84. Lu, S., Zong, C., Fan, W., Yang, M., Li, J., Chapman, A. R., Zhu, P., Hu, X., Xu, L., Yan, L., et al. (2012). Probing Meiotic Recombination and Aneuploidy of Single Sperm Cells by Whole-Genome Sequencing. Science *338*, 1627–1630.

85. Marshall, W. L., and Berbee, M. L. (2010). Population-Level Analyses Indirectly Reveal Cryptic Sex and Life History Traits of Pseudoperkinsus tapetis (Ichthyosporea, Opisthokonta): A Unicellular Relative of the Animals. Molecular Biology and Evolution *27*, 2014–2026.

86. Hickman, M. A., Zeng, G., Forche, A., Hirakawa, M. P., Abbey, D., Harrison, B. D., Wang, Y.-M., Su, C.-H., Bennett, R. J., Wang, Y., et al. (2013). The “obligate diploid” Candida albicans forms mating-competent haploids. Nature *494*, 55–59.

87. Schatten, H., and A, C. (2000). Fertilization in Invertebrates. In Fertilization in Protozoa and Metazoan Animals: Cellular and Molecular Aspects, J. Tarin and A. Cano, eds. (Berlin: Springer), pp. 27–73.

88. Pommerville, J., and Fuller, M. S. (1976). The cytology of the gametes and fertilization of *Allomyces macrogynus*. Archives of microbiology *109*, 21–30.

89. Pommerville, J. (1982). Morphology and physiology of gamete mating and gamete fusion in the fungus Allomyces. Journal of Cell Science *53*, 193–209.

90. Koschwanez, J. H., Foster, K. R., and Murray, A. W. (2013). Improved use of a public good selects for the evolution of undifferentiated multicellularity. eLife *2*, e00367–e00367.

91. Szathmáry, E., and Smith, J. M. (1995). The major evolutionary transitions. Nature *374*, 227–232.

92. Levin, T. C., and King, N. (2013). Evidence for sex and recombination in the choanoflagellate *Salpingoeca rosetta*. Current Biology *23*, 2176–2180.

93. Sebé-Pedrós, A., Roger, A. J., Lang, F. B., King, N., and Ruiz-Trillo, I. (2010). Ancient origin of the integrin-mediated adhesion and signaling machinery. Proceedings of the National Academy of Sciences *107*, 10142–10147.

94. Sebé-Pedrós, A., Ariza-Cosano, A., Weirauch, M. T., Leininger, S., Yang, A., Torruella, G., Adamski, M., Adamska, M., Hughes, T. R., and Gómez-Skarmeta, J. L. (2013). Early evolution of the T-box transcription factor family. Proceedings of the National Academy of Sciences *110*, 16050–16055.

95. DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet *43*, 491–498.

96. Schuelke, M. (2000). An economic method for the fluorescent labeling of PCR fragments. Nature Biotechnology *18*, 233–234.

97. Broman, K. W., Wu, H., Sen, S., and Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. Bioinformatics *19*, 889–890.

98. Brenner, S. (1974). The genetics of *Caenorhabditis elegans*. Genetics *77*, 71–94.

99. Spencer, J. F., and Spencer, D. M. (1996). Mutagenesis in yeast. In Methods in Molecular Biology, I. Evans, ed. (Totowa, NJ: Humana Press Inc.), pp. 17–38.

100. Kim, Y., Schumaker, K. S., and Zhu, J.-K. (2006). EMS mutagenesis of Arabidopsis. Methods Mol Biol *323*, 101–103.

101. Sarin, S., Prabhu, S., O'Meara, M. M., Pe'er, I., and Hobert, O. (2008). *Caenorhabditis elegans* mutant allele identification by whole-genome sequencing. Nat Meth *5*, 865–867.

102. Blumenstiel, J. P., Noll, A. C., Griffiths, J. A., Perera, A. G., Walton, K. N., Gilliland, W. D., Hawley, R. S., and Staehling-Hampton, K. (2009). Identification of EMS-Induced Mutations in *Drosophila melanogaster* by Whole-Genome Sequencing. Genetics *182*, 25–32.

103. Doitsidou, M., Poole, R. J., Sarin, S., Bigelow, H., and Hobert, O. (2010). C. elegans Mutant Identification with a One-Step Whole-Genome-Sequencing and SNP Mapping Strategy. PLoS ONE *5*, e15435.

104. Voz, M. L., Coppieters, W., Manfroid, I., Baudhuin, A., Berg, Von, V., Charlier, C., Meyer, D., Driever, W., Martial, J. A., and Peers, B. (2012). Fast Homozygosity Mapping and Identification of a Zebrafish ENU-Induced Mutation by Whole-

Genome Sequencing. PLoS ONE *7*, e34671.

105. Wurtzel, O., Dori-Bachash, M., Pietrokovski, S., Jurkevitch, E., and Sorek, R. (2010). Mutation Detection with Next-Generation Resequencing through a Mediator Genome. PLoS ONE *5*, e15628.

106. Zuryn, S., Le Gras, S., Jamet, K., and Jarriault, S. (2010). A Strategy for Direct Mapping and Identification of Mutations by Whole-Genome Sequencing. Genetics *186*, 427–430.

107. Harper, M. A., Chen, Z., Toy, T., Machado, I. M. P., Nelson, S. F., Liao, J. C., and Lee, C. J. (2011). Phenotype sequencing: Identifying the genes that cause a phenotype directly from pooled sequencing of independent mutants. PLoS ONE *6*, e16517.

108. Uchida, N., Sakamoto, T., Kurata, T., and Tasaka, M. (2011). Identification of EMS-Induced Causal Mutations in a Non-Reference *Arabidopsis thaliana* Accession by Whole Genome Sequencing. Plant and Cell Physiology *52*, 716–722.

109. Carraway, K. L., and Hull, S. R. (1991). Cell surface mucin-type glycoproteins and mucin-like domains. Glycobiology *1*, 131–138.

110. Drickamer, K., and Dodd, R. B. (1999). C-Type lectin-like domains in *Caenorhabditis elegans*: predictions from the complete genome sequence. Glycobiology *9*, 1357–1369.

111. Lareau, L. F., Brooks, A. N., Soergel, D. A. W., Meng, Q., and Brenner, S. E. (2007). The coupling of alternative splicing and nonsense-mediated mRNA decay. In Alternative Splicing in the Postgenomic Era, B. J. Blencowe and B. R. Graveley, eds. (New York, NY: Springer), pp. 190–211.

112. Mendel, G. (1866). Versuche über Pflanzenhybriden. Verhandlungen des naturforschenden Vereines in Brunn *4*, 3–47.

113. Nüsslein-Volhard, C., and Wieschaus, E. (1980). Mutations affecting segment number and polarity in *Drosophila*. Nature *287*, 795–801.

114. Mayer, U., Ruiz, R. A. T., Berleth, T., Miseéra, S., and Juürgens, G. (1991). Mutations affecting body organization in the *Arabidopsis* embryo. Nature *353*, 402–407.

115. Haffter, P., Granato, M., Brand, M., Mullins, M. C., Hammerschmidt, M., Kane, D. A., Odenthal, J., van Eeden, F. J., Jiang, Y. J., Heisenberg, C. P., et al. (1996). The identification of genes with unique and essential functions in the development of the zebrafish, *Danio rerio*. Development *123*, 1–36.

116. Yamaguchi, Y. (2000). Lecticans: organizers of the brain extracellular matrix. Cellular and Molecular life sciences : CMLS *57*, 276–289.

117. Zelensky, A. N., and Gready, J. E. (2005). The C-type lectin-like domain superfamily. FEBS Journal *272*, 6179–6217.

118. Cambi, A., Koopman, M., and Figdor, C. G. (2005). How C-type lectins detect pathogens. Cellular Microbiology *7*, 481–488.

119. Geijtenbeek, T. B. H., and Gringhuis, S. I. (2009). Signalling through C-type lectin receptors: shaping immune responses. Nature Reviews Immunology *9*, 465–479.

120. Švajger, U., Anderluh, M., Jeras, M., and Obermajer, N. (2010). C-type lectin DC-SIGN: an adhesion, signalling and antigen-uptake molecule that guides dendritic cells in immunity. Cellular Signalling *22*, 1397–1405.

121. Ishikawa, E., Ishikawa, T., Morita, Y. S., Toyonaga, K., Yamada, H., Takeuchi, O., Kinoshita, T., Akira, S., Yoshikai, Y., and Yamasaki, S. (2009). Direct recognition of the mycobacterial glycolipid, trehalose dimycolate, by C-type lectin Mincle. J Exp Med *206*, 2879–2888.

122. Biller, S. J., Schubotz, F., Roggensack, S. E., Thompson, A. W., Summons, R. E., and Chisholm, S. W. (2014). Bacterial Vesicles in Marine Ecosystems. Science *343*, 183–186.