Glycosyltransferases promote development and prevent promiscuous cell aggregation in the choanoflagellate *S. rosetta*

Authors

Laura A. Wetzel[1], Tera C. Levin, Ryan E. Hulett, Daniel Chan, Grant A. King, Reef Aldayafleh, David S. Booth, Monika Abedin Sigg, and Nicole King[2*]

Author Affiliations

Department of Molecular and Cell Biology and Howard Hughes Medical Institute, University of California, Berkeley, CA 94720-3200, USA

[1]https://orcid.org/0000-0003-0391-2542

[2] https://orcid.org/0000-0002-6409-1111

[*]Corresponding author

Email: nking@berkeley.edu

Phone: (510) 643-9395

IMPACT STATEMENT

A genetic screen reveals that glycosyltransferases are required for proper rosette development and the prevention of cell clumping in one of the closest living relatives of animals, the choanoflagellate *S. rosetta*.

ABSTRACT

The mechanisms underlying multicellular development in the animal stem lineage may be reconstructed through the study of choanoflagellates, the closest living relatives of animals. To determine the genetic underpinnings of multicellularity in the emerging model choanoflagellate *S. rosetta,* we performed a screen for mutants with defects in multicellular rosette development. In two of the mutants, Jumble and Couscous, single cells failed to develop into orderly rosettes but instead aggregated promiscuously into amorphous clumps of cells. Both mutants mapped to lesions in genes encoding glycosyltransferases and the mutations perturbed glycosylation patterns in the extracellular matrix (ECM). In animals, glycosyltransferases transfer activated sugars to donor molecules and thereby sculpt the polysaccharide-rich ECM, regulate integrin and cadherin activity, and, when disrupted, contribute to tumorigenesis. The finding that glycosyltransferases promote proper rosette development and prevent cell aggregation in *S. rosetta* suggests a pre-metazoan role for glycosyltransferases in regulating development and preventing abnormal tumor-like multicellularity.

INTRODUCTION

The evolution of multicellular eukaryotes from their single-celled ancestors was a major transition in evolutionary history and allowed for the subsequent origin and diversification of complex macroscopic life (Leigh et al., 1995). However, despite the centrality of multicellularity to the origin and diversification of animals, little is known about the genetic and developmental mechanisms that precipitated the transition to multicellularity in the animal stem lineage. Although the first animals evolved over 600 million years ago, studying their closest living relatives, choanoflagellates, allows the reconstruction of important aspects of animal origins (Brunet and King, 2017; King et al., 2008; Ruiz-Trillo et al., 2008; Schalchian-Tabrizi et al., 2008; Sebé-Pedrós et al., 2017).

Salpingoeca rosetta is an emerging model choanoflagellate that was isolated from nature as a spherical colony of cells called a rosette. Under standard laboratory conditions, *S. rosetta* proliferates as solitary cells or as linear chain colonies that easily break apart into solitary cells (Dayel et al., 2011). When exposed to rosette inducing factors (RIFs) produced by the co-isolated prey bacterium *Algoriphagus machipongonensis, S. rosetta* instead develops into highly organized and structurally stable rosettes through a process of serial cell division (Alegado et al., 2012; Dayel et al., 2011; Fairclough et al., 2010; Woznica et al., 2016). Recent advances, including a sequenced genome (Fairclough et al., 2010), the discovery of a sexual phase to the *S. rosetta* life cycle that enables controlled mating (Levin et al., 2014; Levin and King, 2013; Woznica et al., 2017), and techniques that allow for transfection and expression of transgenes (Booth et al., 2018) have enabled increasingly detailed studies of molecular mechanisms underlying rosette development in *S. rosetta*.

In a pioneering genetic screen to identify genes required for rosette formation in *S. rosetta*, multiple rosette defect mutants were recovered that displayed a range of phenotypes (Levin et al., 2014). The first mutant to be characterized in detail was named Rosetteless; while Rosetteless cells did not develop into rosettes in the presence of RIFs, they were otherwise indistinguishable from wild type cells (Levin et al., 2014). The mutation underlying the Rosetteless phenotype was mapped to a C-type lectin, encoded by the gene *rosetteless,* the first gene shown to be required for rosette formation (Levin et al., 2014). In animals, C-type lectins function in signaling and adhesion to promote development and innate immunity (Cambi et al., 2005; Geijtenbeek and Gringhuis, 2009; Ruoslahti, 1996; Švajger et al., 2010; Zelensky and Gready, 2005). Although the molecular mechanisms by which *rosetteless* regulates rosette development remain unknown, the localization of Rosetteless protein to the rosette interior suggests that it functions as part of the extracellular matrix (ECM) (Levin et al., 2014).

Here we report on the largest class of mutants from the original rosette defect screen (Levin et al., 2014), all of which fail to develop into organized rosettes and instead form large, amorphous clumps of cells in both the absence and presence of RIFs. By mapping the mutations underlying the clumpy, rosette defect phenotypes of two mutants in this class, we identified two predicted glycosyltransferase genes that are each essential for proper rosette development. The causative glycosyltransferase mutations led to similar perturbations in the glycosylation pattern of the basal ECM. The essentiality of glycosyltransferases for rosette development combined with prior findings of the requirement of a C-type lectin highlight the importance of the ECM for regulating

multicellular rosette development and preventing spurious cell adhesion in a close relative of animals.

RESULTS

Rosette defect mutants form amorphous clumps of cells through promiscuous cell adhesion

The original rosette defect screen performed by Levin et al., 2014 yielded nine mutants that were sorted into seven provisional phenotypic classes. For this study, we screened 21,925 additional clones and identified an additional seven mutants that failed to form proper rosettes in the presence of *Algoriphagus* RIFs. (For this study, we used *Algoriphagus* outer membrane vesicles as a source of RIFs, as described in Woznica et al., 2016). Comparing the phenotypes of the 16 total rosette defect mutants in the presence and absence of RIFs allowed us to classify four broad phenotypic classes: (1) Class A mutants that have wild type morphologies in the absence of RIFs and entirely lack rosettes in the presence of RIFs, (2) Class B mutants that have wild type morphologies in the absence of RIFs and develop reduced levels of rosettes with aberrant structures in the presence of RIFs, (3) Class C mutants that produce large clumps of cells in both the presence and absence of RIFs while forming little to no rosettes in the presence of RIFs, and (4) a Class D mutant that exist primarily as solitary cells, with no linear chains of cells detected in the absence of RIFs and no rosettes detected in the presence of RIFs (Table S1).

Of the 16 rosette defect mutants isolated, seven mutants fell into Class C. For this study, we focused on four Class C mutants — Seafoam, Soapsuds, Jumble, and

Couscous (previously named Branched in Levin et al., 2014) — that form amorphous, tightly packed clumps of cells, both in the presence and absence of RIFs, but never develop into rosettes (Table 1; Figure 1A,B). We found that the clumps contain a few to hundreds of mutant cells that pack together haphazardly, unlike wild type rosettes in which all cells are oriented with their basal poles toward the rosette center and their apical flagella extending out from the rosette surface (Alegado et al., 2012; Levin et al., 2014; Woznica et al., 2016). Moreover, in contrast with the structural stability and shear resistance of wild type rosettes (Figure 1A) (Levin et al., 2014), the cell clumps formed by Class C mutants were sensitive to shear and separated into solitary cells upon pipetting or vortexing the culture (Figure 1A).

Following exposure to shear, we observed that mutant cells re-aggregated into new clumps within minutes, while wild type cells never formed clumps (Figure 1C, D; rare cell doublets were likely due to recent cell divisions). Within 30 minutes after disruption by shear force, cell clumps as large as 75, 55, 32, and 23 cells formed in Couscous, Soapsuds, Seafoam, and Jumble mutant cultures, respectively. The cell aggregation was not strain-specific, as unlabeled Jumble and Couscous mutant cells adhered to wild type cells identified by their expression of cytoplasmic mWasabi (Figure 1-figure supplement 1). Therefore, the cell clumps are not aberrant rosettes, which never form through aggregation and instead require at least 15 – 24 hours to develop clonally through serial rounds of cell division (Dayel et al., 2011; Fairclough et al., 2010). The fact that the seven Class C mutants isolated in this screen were also defective in rosette development suggests a direct link between promiscuous cell adhesion and

failed rosette development. Each of the mutants tested also displayed a mild defect in cell proliferation (Figure 1-figure supplement 2).

Improving genetic mapping in *S. rosetta* through bulk segregant analysis

We next set out to identify the causative mutation(s) underlying the clumping and rosette defect phenotypes in each of these mutants. In the Levin *et al.* 2014 study, the Rosetteless mutant was crossed to a phenotypically wild type Mapping Strain (previously called Isolate B in Levin et al., 2014) and relied on genotyping of haploid F1s at 60 PCR-verified genetic markers that differed between the Rosetteless mutant and the Mapping Strain (Levin et al., 2014). The 60 markers were distributed unevenly across the 55 Mb genome and proved to be insufficient for mapping the Class C mutants for this study. Compounding the problem, the low level of sequence polymorphism among *S. rosetta* laboratory strains and abundance of repetitive sequences in the draft genome assembly (Fairclough et al., 2013; Levin et al., 2014) made it difficult to identify and validate additional genetic markers, while genotyping at individual markers proved labor intensive and costly.

To overcome these barriers, we modified bulk segregation methods developed in other systems (Doitsidou et al., 2010; Leshchiner et al., 2012; Lister et al., 2009; Pomraning et al., 2011; Schneeberger et al., 2009; Voz et al., 2012; Wenger et al., 2010) for use in *S. rosetta.* Our strategy involved: (1) crossing mutants to the Mapping Strain (which contains previously identified sequence variants); (2) isolating heterozygous diploids identified through genotyping at a microsatellite on supercontig 1; (3) inducing meiosis; (4) growing clonal cultures of haploid F1 offspring; (5) phenotyping

the F1 offspring; (6) pooling F1 offspring based on their clumping phenotype; and (7) deeply sequencing pooled genomic DNA from the F1 mutants to find mutations that segregated perfectly with the clumping phenotype (Figure 2-figure supplement 1).

To test whether a bulk segregant approach would work in *S. rosetta*, we first analyzed a cross between the previously mapped Rosetteless mutant and the Mapping Strain (Levin et al., 2014). We isolated 38 F1s with the rosette defect phenotype from a Mapping Strain×Rosetteless cross (Levin et al., 2014), grew clonal cultures from each, pooled the resulting cultures, extracted their genomic DNA, and sequenced the pooled mutant genomes to an average coverage of 187X. Against a background of sequence variants that did not segregate perfectly with the Rosetteless phenotype, five unlinked single nucleotide variants (SNVs) and insertions/deletions (INDELs) were found to perfectly segregate with the phenotype (Table S2). Four of these detected sequence variants likely had spurious correlations with the phenotype resulting from relatively low sequencing coverage at those variants (>0.25X coverage of the entire genome) (Table S2). In contrast, the remaining SNV was detected in a well-assembled portion of the genome at a sequencing depth approaching the average coverage of the entire genome. The perfectly segregating SNV, at position 427,804 on supercontig 8, was identical to the causative mutation identified in Levin et al., 2014 (Table S2). Thus, a method based on pooling F1 haploid mutants, identifying sequence variants that perfectly segregated with the phenotype, and masking those SNVs/INDELs that were detected with >0.25X coverage of the total genome was effective for correctly pinpointing the causal mutation for Rosetteless (Figure 2-figure supplement 1). Therefore, we used this validated bulk segregant method to map the clumping mutants.

Mapping crosses were carried out for the four clumping/rosette defect mutants characterized in this study (Seafoam, Soapsuds, Jumble, and Couscous) and all four crosses yielded heterozygous diploids, demonstrating that they were competent to mate. As observed in prior studies of *S. rosetta* mating (Levin et al., 2014; Woznica et al., 2017), the diploid cells each secreted a flask-shaped attachment structure called a theca and were obligately unicellular. Therefore, the heterozygous diploids were not informative about whether the mutations were dominant or recessive as the phenotypes could only be detected in haploid cells. For Seafoam and Soapsuds, we isolated heterozygous diploids, but never recovered F1 offspring with the mutant phenotype (Table 1). The inability to recover haploids with either clumping or rosette defect phenotypes from the Seafoam×Mapping Strain and Soapsuds×Mapping Strain crosses might be explained by any of the following: (1) the clumping/rosette defect phenotypes are polygenic, (2) meiosis defects are associated with the causative mutations, and/or (3) mutant fitness defects allowed wild type progeny to outcompete the mutant progeny. In contrast, heterozygous diploids from crosses of Jumble and Couscous to the Mapping Strain produced F1 haploid progeny with both wild type and mutant phenotypes and thus allowed for the successful mapping of the causative genetic lesions, as detailed below.

Jumble maps to a putative glycosyltransferase

Following the bulk segregant approach, we identified 5 sequence variants in Jumble that segregated perfectly with both the clumping and rosette defects. Only one of these – at position 1,919,681 on supercontig 1 – had sequencing coverage of at least

0.25X of the average sequence coverage of the rest of the genome (Figure 2A; Table S3). In a backcross of mutant F1 progeny to the Mapping Strain, we confirmed the tight linkage of the SNV to the rosette defect phenotype (Figure 2B). Moreover, all F2 progeny that displayed a rosette defect also had a clumping phenotype. Given the tight linkage of both traits with the SNV and the absence of any detectable neighboring sequence variants, we infer that the single point mutation at genome position 1:1,919,681 causes both the clumping and rosette defect phenotypes in Jumble mutants.

The mutation causes a T to C transition in a gene hereafter called *jumble* (GenBank accession EGD72416/NCBI accession XM_004998928; Figure 2A). The *jumble* gene contains a single exon and is predicted to encode a 467 amino acid protein containing a single transmembrane domain. Following the convention established in Levin et al. 2014, the mutant allele, which is predicted to confer a leucine to proline substitution at amino acid position 305, is called *jumble*^{lw1}.

We used recently developed methods for transgene expression in *S. rosetta* (Booth et al., 2018) to test whether expression of a *jumble* with an N- or C-terminal *monomeric teal fluorescent protein* (*mTFP*) gene fusion under the *S. rosetta elongation factor L* (*efl*) promoter could complement the mutation and rescue rosette development in the Jumble mutant (Figure 2C,D). We were able to enrich for rare transfected cells by using a construct in which the puromycin resistance gene (*pac*) was expressed under the same promoter as the *jumble* fusion gene, with the two coding sequences separated by a sequence encoding a self-cleaving peptide (Kim et al., 2011). Transfection of Jumble mutant cells with wild type *jumble-mTFP* followed by puromycin selection and

the addition of RIFs yielded cultures in which 9.33%±5.07% of cells were in rosettes (Figure 2C). Similarly, transfection of Jumble with *mTFP-jumble* followed by puromycin selection and rosette induction resulted in cultures with 7.00%±4.91% of cells in rosettes (Figure 2C). Importantly, we did not detect any rosettes when we transfected Jumble cells with *mTFP* alone, *jumble*^{lw1}*-mTFP*, or *mTFP*-jumble^{lw1}. Complementation of the Jumble mutant by the wild type *jumble* allele, albeit in a subset of the population, provided further confirmation that the *jumble*^{lw1} mutation causes the cell clumping and rosette defect phenotypes. The fact that the transfection experiment did not allow all cells to develop into rosettes may be due to any number of reasons, including incomplete selection against untransfected cells, differences in transgene expression levels in different transfected cells, and the possibility that the mTFP tag reduces or otherwise changes the activity of the Jumble protein.

We next sought to determine the function and phylogenetic distribution of the *jumble* gene. BLAST searches uncovered unannotated *jumble* homologs in nine other choanoflagellates (Figure 2-figure supplement 2A) and in fungi, but none in animals. The choanoflagellate homologs of *jumble* were detected in the transcriptomes of species representing each of the three major choanoflagellate clades (Richter et al., 2018), suggesting that *jumble* evolved before the origin and diversification of choanoflagellates. Although Interpro (Finn et al., 2017) and Pfam (Finn et al., 2016) did not reveal any known protein domains in Jumble, the NCBI Conserved Domain Search (Marchler-Bauer et al., 2017) predicted a glycosyltransferase domain with low confidence (E-value 3.87^{-03}). Moreover, two different algorithms that use predicted secondary and tertiary structures to identify potential homologs, HHphred (Zimmermann

et al., 2017) and Phyre2 (Kelly et al., 2015), predict that Jumble is related to well-annotated glycosyltransferases (HHphred: E-value 7.5^{-19} to polypeptide N-acetylgalactosaminyltransferase 4; Phyre2: Confidence 94.5% to human polypeptide n-acetylgalactosaminyltransferase 2) (Figure 2-figure supplement 2B). The Leu305Pro substitution in Jumble^{lw1} disrupts a predicted alpha helix, which we hypothesize would prevent proper folding of the Jumble protein (Figure 2A).

Glycosyltransferases play essential roles in animal development (Sawaguchi et al., 2017; Zhang et al., 2008) and cell adhesion (Müller et al., 1979; Stratford, 1992). Their biochemical functions include transferring an activated nucleotide sugar, also called a glycosyl donor, to lipid, protein, or carbohydrate acceptors (Lairson et al., 2008). Target acceptors in animals include key signaling and adhesion proteins such as integrins and cadherins, whose activities are regulated by N- and O-linked polysaccharide modifications, also referred to as N- and O-linked glycans (Larsen et al., 2017; Zhao et al., 2008). Notably, many well-characterized glycosyltransferases act in the Golgi apparatus, where they glycosylate molecules that are trafficked through the secretory system (El-Battari, 2006; Tu and Banfield, 2010). To investigate the localization of Jumble, we transfected wild type cells with a *jumble-mWasabi* gene fusion transcribed under the control of the *S. rosetta efl* promoter. Jumble-mWasabi protein localized to the apical pole of the cell body near the base of the flagellum. Based on comparisons with transmission electron micrographs of *S. rosetta* and other choanoflagellates, Jumble-mWasabi localization corresponds to the location of the Golgi apparatus, for which there is not yet a fluorescent marker in *S. rosetta* (Figure 2E,G; Figure 2-figure supplement 3A) (Leadbeater, 2015). In contrast, Jumble^{lw1}-mWasabi,

was distributed in a tubular pattern throughout the cell and co-localized with an endoplasmic reticulum (ER) marker (Figure 2F,H; Figure 2-figure supplement 3B) (Booth et al., 2018). The ER localization of Jumble^{lw1} is consistent with the hypothesis that the missense mutation disrupts proper protein folding as often misfolded proteins are retained in the ER and targeted for degradation (Kopito, 1997). The failure of the Jumble^{lw1} protein to localize properly at the Golgi apparatus strongly suggests a loss of function.

Couscous maps to a lesion in a predicted mannosyltransferase

We followed a similar strategy to map the genetic lesion(s) underlying the Couscous mutant phenotype. Using the bulk segregant approach on F1 mutant offspring from a Couscous × Mapping Strain cross, we identified eight sequence variants that segregated perfectly with the clumping and rosette defect phenotypes, of which only one – a single nucleotide deletion at position 462,534 on supercontig 22 – had sequencing coverage at least 0.25X of the average sequence coverage of the rest of the genome (Figure 3A; Table S4). The tight linkage of the deletion to both the clumping and rosette defect phenotypes was further confirmed by genotyping the sequence variant in F2 mutants resulting from backcrosses of F1 mutants to the Mapping Strain (Figure 3B). Given the tight linkage, we infer that the deletion at position 462,534 on supercontig 22 causes both clumping and the disruption of rosette development in Couscous mutant cells.

The single nucleotide deletion at position 462,534 on supercontig 22 sits in a four-exon gene, hereafter called *couscous* (GenBank accession EGD77026/ NCBI

accession XM_004990809). The mutation causes a predicted frameshift leading to an early stop codon in the mutant protein, Couscous^{lw1} (Figure 3A). As with the Jumble mutant, we were able to rescue rosette formation in a portion of the population by transfecting cells with either a *couscous-mTFP* or *mTFP-couscous* gene fusion under the *efl* promoter (Figure 3C, D), thereby increasing our confidence in the mapping results.

The predicted Couscous amino acid sequence contains a specific type of glycosyltransferase domain, an alpha-mannosyltransferase domain, that transfers activated mannose onto the outer chain of core N-linked polysaccharides and O-linked mannotriose (Strahl-Bolsinger et al., 1999). The predicted mannosyltransferase domain shares 28% and 35% amino acid sequence identity to alpha 1-2 mannosyltransferase (MNN2) proteins in *Saccharomyces cerevisiae* and *Candida albicans*, respectively, including the conserved DXD motif found in many families of glycosyltransferases (Wiggins and Munro, 1998) (Figure 3-figure supplement 1A). MNN2 proteins catalyze the addition of the first branch of mannose-containing oligosaccharides found on cell wall proteins (Rayner and Munro, 1998) and proper MNN2 activity is required for flocculation, or non-mating aggregation, in *S. cerevisiae* (Stratford, 1992). In addition to the mannosyltransferase domain, Couscous is predicted to have a PAN/Apple domain composed of a conserved core of three disulfide bridges (Ho et al., 1998; Tordai et al., 1999). PAN/Apple domains are broadly distributed among eukaryotes, including animals, where they mediate protein-protein and protein-carbohydrate interactions, often on the extracellular surface of the cell (Ho et al., 1998; Tordai et al., 1999).

315 In wild type cells transfected with a *couscous-mWasabi* transgene under the *efl*
316 promoter, Couscous was found in puncta scattered about the cytosol, collar and cell
317 membrane (Figure 3-figure supplement 1B, C). While Couscous-mWasabi was clearly
318 not localized to the Golgi, the puncta may co-localize with the ER, where
319 glycosyltransferases are also known to function (El-Battari, 2006; Tu and Banfield,
320 2010). However, despite attempting to co-transfect cells with *couscous-mWasabi* and a
321 marker of the ER, we were unable to detect any cells expressing both constructs. In
322 addition, it is possible that the fusion of Couscous to a fluorescent protein or its
323 overexpression interfered with its proper localization in *S. rosetta.* Therefore, we are
324 currently uncertain about the subcellular localization of Couscous protein.
325
326 **Jumble and Couscous mutants lack proper sugar modifications at the basal pole**
327 Because both Jumble and Couscous have mutations in putative
328 glycosyltransferases, we hypothesized that the abundance or distribution of cell surface
329 sugars, called glycans, on Jumble and Couscous mutant cells might be altered. To
330 investigate the distribution of cell surface glycans, we stained live *S. rosetta* with diverse
331 fluorescently labelled sugar-binding lectins. Of the 22 lectins tested, 21 either did not
332 recognize *S. rosetta* or had the same staining pattern in wild type, Jumble and
333 Couscous cells (Table S5).
334 The remaining lectin, jacalin, bound to the apical and basal poles of wild type
335 cells (Figure 4A, B, B'). Jacalin also brightly stained the ECM filling the center of
336 rosettes in a pattern reminiscent of the Rosetteless C-type lectin (Levin et al., 2014)
337 (Figure 4A, B'), although the two were not imaged simultaneously because jacalin does

not bind after cell fixation and labelled Rosetteless antibodies accumulate strongly in the food vacuoles of live cells. In contrast with wild type cells, the basal patch of jacalin staining was absent or significantly diminished in Couscous and Jumble mutants, both in the presence and absence of RIFs (Figure 4 C-F). Interestingly, the apical patch of jacalin staining in mutant cells appeared similar to wild type cells. This may explain the lack of a clear difference in bands detected with jacalin by western blot between wild type and mutants whole cell lysates (Figure 4-figure supplement 1).

The loss of basal jacalin staining indicated that the Jumble and Couscous mutations either disrupt proper trafficking of sugar-modified molecules to the basal pole of cells or alter the glycosylation events themselves. Thus, we examined whether the basal secretion of Rosetteless protein was disrupted in the mutant strains. In both Jumble and Couscous, Rosetteless properly localized to the basal pole, but its expression did not increase nor was it secreted upon treatment with RIFs, as normally occurs in wild type cells (Figure 4-figure supplement 2). Because Rosetteless is required for rosette development, this failure to properly upregulate and secrete Rosetteless might contribute to the rosette defect phenotype in Jumble and Couscous cells.

DISCUSSION

Of the 16 rosette defect mutants isolated in Levin *et al.* 2014 and in this study, almost half (7) also display a mild to severe clumping phenotype. This suggests that mechanisms for preventing promiscuous adhesion among wild type cells can be easily disrupted. We found that the clumping phenotype results from promiscuous adhesion of

mutant cells to other mutant or wild type cells rather than from incomplete cytokinesis. A recent study revealed that the bacterium *Vibrio fischeri* induces *S. rosetta* to form swarms of cells, visually similar to the mutant clumps, as part of their mating behavior (Woznica et al., 2017). However, it seems unlikely that the clumping Class C mutants is related to swarming; the cell fusion and subsequent settling of diploid cells characteristic of *V. fischeri*-induced mating have not been observed in the class C mutants cultured without *V. fischeri*.

For both Jumble and Couscous, the causative mutations mapped to predicted glycosyltransferase genes. Consistent with its role as a glycosyltransferase, Jumble localized to the Golgi apparatus, but Couscous appeared to localize in cytoplasmic puncta and to the cell membrane. We predict that the glycosyltransferase mutations are loss of function alleles, given that transfection of mutant *S. rosetta* with the wild type alleles was sufficient to complement each of the mutations. While we have not uncovered the target(s) of the glycotransferases or the exact nature of the interplay between the two phenotypes, disruption of the glycocalyx at the basal pole of both Jumble and Couscous mutant cells (Figure 4) hints that the regulation of ECM could play a role in preventing clumping and in promoting proper rosette development.

One possible explanation for the clumping phenotype is that *jumble* and *couscous* are required to regulate the activity of cell surface adhesion molecules and receptors. Glycosylation regulates the activities of two key adhesion proteins in animals: integrins that regulate ECM adhesion, and cadherins that, among their various roles in cell signaling and animal development, bind other cadherins to form cell-cell adhesions called adherens junctions (Larsen et al., 2017; Zhao et al., 2008). Cadherin activity can

384 be either positively or negatively regulated by glycosyltransferases. For example,
385 epithelial cadherin (E-cadherin) is modified by N-acetylglucosaminyltransferase III (GnT-
386 III) whose activity leads to increased cell adhesion and N-acetylglucosaminyltransferase
387 V (GnT-V) whose activity leads to decreased cell adhesion (Carvalho et al., 2016;
388 Granovsky et al., 2000). GnT-V knockdown enhances cell-cell adhesion mediated by E-
389 cadherin and the related N-cadherin (Carvalho et al., 2016; Guo et al., 2009). The
390 inactivation of E-cadherin, including through over- or under- expression of GnT-V or
391 GnT-III, is considered to be a hallmark of epithelial cancers (Hirohashi and Kanai,
392 2003). *S. rosetta* expresses 29 different cadherins (Nichols et al., 2012) and it is
393 possible that mutations to *jumble* and *couscous* disrupt regulatory glycosylation of a cell
394 adhesion molecules like cadherins.

395 Another possibility is that *jumble* and *couscous* add a protective sugar layer to
396 the cell surface and loss of glycosyltransferase activity reveals underlying sticky
397 surfaces. If *jumble* and *couscous* add branches to existed sugar modifications, their loss
398 of function could expose new sugar moieties at the cell surface that act as ligands for
399 lectins that aggregate cells. Lectins mediate cell aggregation in diverse organisms
400 (Colin Hughes, 1992). For example, sponges such as *Geodia cydonium* can be
401 disaggregated into single cells and then reaggregated through lectin binding of a post-
402 translational sugar modification (Müller et al., 1979). In *S. cerevisiae*, the
403 mannosyltransferase MNN2 adds mannose structures to the cell wall that are
404 recognized by aggregating lectins and MNN2 is required for proper flocculation (Rayner
405 and Munro, 1998; Stratford, 1992). Exposing new sugars on the cell surface in Jumble

406 and Couscous could lead to spurious aggregation, potentially by lectins or other sugar
407 binding proteins.

408 It is somewhat more difficult to infer how increased clumping in single cells might
409 interfere with rosette development. One possibility is that the disruption of ECM
410 glycosylation that we hypothesize might promote clumping may also prevent the proper
411 maturation of the ECM needed for rosette development (Figure 5). A prior study showed
412 that only *S. rosetta* cells recognized with the lectin wheat germ agglutinin (WGA) are
413 competent to form rosettes, which suggests that glycosylation might be necessary for
414 rosette development (Dayel et al., 2011). While WGA staining does not appear to be
415 perturbed in Jumble and Couscous (Table S2), jacalin staining at the basal pole
416 appears severely reduced or abolished compared to wild type. Jacalin staining was
417 enriched in the center of wild type rosettes in a pattern reminiscent of Rosetteless,
418 which is required for rosette development (Levin et al., 2014). Intriguingly, in Jumble
419 and Couscous, Rosetteless localized to the correct pole, but did not become enriched
420 upon rosette induction, indicating that the ECM did not properly mature. Rosetteless has
421 mucin-like Ser/Thr repeats that are predicted sites of heavy glycosylation and two C-
422 type lectin domains that would be expected to bind to sugar moieties (Levin et al.,
423 2014). Therefore, it is possible that Rosetteless might be regulated either through direct
424 glycosylation or through the glycosylation of potential binding partners by Jumble and
425 Couscous.

426 The clumping, rosetteless mutants underscore the differences between cell
427 aggregation and a regulated clonal developmental program, such as embryogenesis or
428 rosette development. Aggregation is hypothesized to only be evolutionarily stable if

limited to close relatives (Gilbert et al., 2007; Kuzdzal-Fick et al., 2011). Recent efforts to experimentally evolve multicellularity in yeast and the green alga *Chlamydomonas reinhardtii* have all resulted in isolates that formed clonally and not through aggregation (Herron et al., 2018; Ratcliff et al., 2013, 2012). Although aggregation and clonal multicellularity both rely on cell-cell adhesion and cell signaling, the fact that they are mutually exclusive in *S. rosetta* and in experimental evolution studies argues against their co-occurrence in the animal stem lineage.

MATERIALS AND METHODS

Media preparation, strains, and cell culture

Unenriched artificial seawater (ASW), AK artificial seawater (AK), cereal grass media (CG), and high nutrient (HN) media were prepared as described previously (Booth et al., 2018; Levin et al., 2014; Levin and King, 2013). The wild type strain, from which each mutant was generated, was the described strain SrEpac (ATCC PRA-390; accession number SRX365844) in which *S. rosetta* is co-cultured monoxenically with the prey bacterium *Echinicola pacifica* (Levin et al., 2014; Levin and King, 2013; Nedashkovskaya et al., 2006). Seafoam, Soapsuds, and Couscous (previously named Branched) were generated through X-ray mutagenesis and Jumble was generated by EMS mutagenesis as documented in Levin et al., 2014. For routine culturing, wild type and mutant cultures were diluted 1:10 every 2-3 days in HN media. The Mapping Strain, (previously called Isolate B in Levin et al., 2014) used for mapping crosses (accession number SRX363839) was grown in the presence of rosette-inducing *A. machipongonensis* bacteria (ATCC BAA-2233). The Mapping Strain was maintained in

25% CG media diluted in ASW and passaged 1:10 every 2-3 days. For transfection of *S. rosetta*, cells were maintained in 5% (vol/vol) HN media in AK seawater (Booth et al., 2018). Rosette induction was performed with *A. machipongonensis* outer membrane vesicles (OMVs) prepared as in Woznica et al., 2016 and referred to here as rosette inducing factors (RIFs).

Imaging and quantifying rosette phenotypes

To image rosette phenotypes (Figure 1A), cells were plated at a density of $1x10^4$ cells/ml in 3 ml HN media either with or without *Algoriphagus* RIFs. Cultures were imaged after 48 hr of rosette induction in 8-well glass bottom dishes (Ibidi 15 µ-Slide 8 well Cat. No. 80826) that were coated with 0.1 mg/mL poly-D-lysine (Sigma) for 15 min and washed 3 times with water to remove excess poly-D-lysine. For imaging wild type and mutant cultures in the presence and absence of RIFs (Figure 1A top two panels), 200 µl of cells were plated with a wide bore pipette tip for minimal disruption and allowed to settle for 5 min. For images of vortexed cells (Figure 1A bottom panel), 200 µl of cells were vortexed for 15 s before plating and imaged within 10 min of plating to prevent re-clumping. Cells were imaged live by differential interference contrast microscopy using a Zeiss Axio Observer.Z1/7 Widefield microscope with a Hammatsu Orca-Flash 4.0 LT CMOS Digital Camera and a 63x/NA1.40 Plan-Apochromatic oil immersion lens with 1.6X optivar setting.

To quantify rosette induction (Figure 1B), cells were plated at a density of $1x10^4$ cells/ml in 3 ml HN media with RIFs. After 48 hr, an aliquot of cells was vortexed vigorously for 15 secs and fixed with formaldehyde. To determine the percentage of cells in rosettes, the relative number of single cells and cells within rosettes were scored

475 using a hemocytometer. Rosettes were counted as a group of 3 or more cells with
476 organized polarity relative to a central focus after exposure to vortexing.

477 **Imaging and quantification of cell clumping**

478 Clumps were quantified using a modified protocol from Woznica et al., 2017
479 (Figure 1C). To prevent cells from sticking to the bottom of the glass dishes, 8 well
480 glass bottom dishes (Ibidi 15 µ-Slide 8 well Cat. No. 80826) were coated with 1% BSA
481 for 1 hr and washed 3 times with water to remove any residual BSA. Cells were diluted
482 to $5x10^5$ cells/mL, vortexed for 15 s to break apart any pre-formed clumps and plated in
483 the BSA pre-treated dishes. For quantification, DIC images were taken using a Zeiss
484 Axio Observer.Z1/7 Widefield microscope with a Hammatsu Orca-Flash 4.0 LT CMOS
485 Digital Camera and a 20x objective. Images were collected for each strain from 10
486 distinct locations throughout the well.

487 Images were batch processed in ImageJ for consistency. To accurately segment
488 the phase bright cells and limit signal from the phase dark bacteria the following
489 commands were applied with default settings: 'Smooth' (to reduce background bacterial
490 signal), 'Find Edges' (to highlight the phase-bright choanoflagellate cells), 'Despeckle'
491 (to remove noise), 'Make Binary' (to convert to black and white), 'Dilate' (to expand to
492 smooth jagged edges from segmentation), 'Erode' (to return to the same size as before
493 dilate), and 'Fill Holes' (to fill any remaining small holes). Finally, images were analyzed
494 with the 'Analyze Particles' command to calculate the area of the clump and only
495 particles larger than 20 μm^2 were kept to filter out any remaining bacterial signal. Cell
496 equivalents/clump (Figure 1C, right y axis) were calculated by dividing the area of the
497 clump by the area of a representative individual cell (as approximated by averaging the

498 area of the wild type cells). Data are presented as violin boxplots, showing the median
499 cell number (middle line), interquartile range (white box), and range excluding outliers
500 (thin line). A minimum of 630 clumps from two biological replicates were measured for
501 each condition.

502 **Performing mapping crosses**

503 Mapping crosses for each mutant strain (Seafoam, Soapsuds, Jumble, and
504 Couscous) with Mapping Strain (previously described as Isolate B) were attempted
505 using both methods previously shown to induce mating in *S. rosetta*: nutrient limitation
506 for 11 days and addition of 2.5-5% *V. fischeri* conditioned media (Levin and King, 2013;
507 Woznica et al., 2017). Both methods were effective at inducing mating for all attempted
508 crosses; here, we report which method was used to generate data for each individual
509 cross. Cells induced to mate were plated by limiting dilution to isolate diploid clones.
510 Clonal isolates were allowed to grow for 5-7 days and screened for populations of
511 thecate cells, as these are the only documented diploid cell type (Levin et al., 2014;
512 Woznica et al., 2017). From each population of thecate cells, we extracted DNA from 75
513 µl of cells by scraping cells from the plate, harvesting and pelleting the cells,
514 resuspending in 10 µl of base solution (25 mM NaOH, 2 mM EDTA), transferring
515 samples into PCR plates, boiling at 100°C for 20 min, followed by cooling at 4°C for 5
516 min, and then adding 10 µl Tris solution (40 mM Tris-HCl, pH 7.5). We used 2 µl of this
517 sample as the DNA template for each genotyping reaction. We identified heterozygous
518 strains through genotyping by PCR at a single microsatellite genotyping marker at
519 position 577,135 on supercontig 1 (Forward primer: GACAGGGCAAAACAGACAGA
520 and Reverse primer: CCATCCACGTTCATTCTCCT) that distinguishes a 25 bp deletion

521 in the Mapping Strain (199 bp) from the strain used to generate the mutants (217 bp).
522 Isolates containing PCR products of both sizes were inferred to be diploid. Meiosis was
523 induced by rapid passaging every day in CG medium. For both Seafoam and Soapsuds,
524 we were able to generate putative outcrossed diploids by crossing to the Mapping Strain
525 based on the genotyping marker on supercontig 1, but we only could only clonally
526 isolate populations of F1 haploids with rosettes and never isolated any F1 haploids with
527 the clumpy, rosetteless phenotype.

528 For the successful cross of Jumble to the Mapping Strain, we induced mating by
529 starvation using the approach of Levin and King 2013. First, we started with rapidly
530 growing, regularly passaged strains, pelleted $2x10^6$ cells/mL of each strain together and
531 resuspended in 10mL of ASW lacking any added nutrients. After 11 days of starvation in
532 ASW, we pelleted all cells (presumably including diploid cells resulting from mating) and
533 resuspended in 100% CG media to recover any diploids. After 3 days of recovery, we
534 isolated clones by limiting dilution in 10% CG media in ASW (vol/vol). The probability of
535 clonal isolation in this step was 0.91-0.93 (calculated using the Poisson distribution and
536 the number of choanoflagellate-free wells per plate; Levin and King, 2013). Three
537 clonally isolated heterozygous populations, each containing almost exclusively thecate
538 cells, were identified through genotyping by PCR at a supercontig 1 microsatellite as
539 described above. To induce meiosis, heterozygotes were diluted 1:2 in 25% CG media
540 in ASW (vol/vol) every 1-2 days for 8 days. As soon as rosettes and swimming cells
541 were observed, we repeated the serial dilution to isolate clones (probability of clonal
542 isolation 0.85-0.98). We collected any clonally isolated populations that formed rosettes
543 or clumps and ignored any wells containing thecate cells assuming that these

represented diploid cells that had not undergone meiosis. 56% of all non-thecate isolates displayed the cell clumping phenotype and 44% of all non-thecate isolates were capable of forming rosettes, consistent with Mendelian segregation of a single locus (X^2=1.162, df=1, p=0.28). Isolates were genotyped with the marker on supercontig 1 to ensure that independent assortment of the genotype and the phenotype indeed occurred. In total, 30 clumpy F1s were collected for bulk segregation analysis.

For the successful Couscous cross, we induced mating using *V. fisheri* conditioned media using the approach of Woznica et al., 2017. A mixture of $1x10^6$ Couscous and Mapping Strain cells at stationary growth were pelleted and resuspended in 5% *V. fischeri* conditioned media in ASW (vol/vol). After 24 hr, the cells were pelleted, resuspended in 5% HN media in ASW (vol/vol), and allowed to recover for 24 hr. We then isolated clones by limiting dilution in 10% CG media in ASW (vol/vol). The probability of clonal isolation in this step was between 0.97-0.98. We extracted DNA as described above and identified heterozygous clones through genotyping by PCR at a single microsatellite genotyping marker on supercontig 1. Four clonally isolated heterozygous populations, containing almost exclusively thecate cells, were identified. To induce meiosis, heterozygotes were passaged 1:2 in 25% CG media in ASW (vol/vol) every 1-2 days for 8 days. As soon as rosettes and swimming cells were observed, we repeated clonal isolation (probability of clonal isolation 0.78-0.97). We collected any clonally isolated populations that formed rosettes or clumps and ignored any wells containing thecate cells assuming that these represented diploid cells that had not undergone meiosis. Only 14.6% of non-thecate isolates were clumps; this deviation from a Mendelian ratio (X^2=225.63, df=1, p<0. 5.34^{-51}) may indicate a potential fitness

567 defect of the mutant phenotype. Isolates were genotyped with the marker on
568 supercontig 1 to ensure that independent assortment indeed occurred. In total, 22
569 clumpy F1s were collected for bulk segregant analysis.

570 **Whole genome sequencing**

571 Jumble and Couscous were whole genome sequenced individually to identify the
572 mutation(s) carried in each strain. To do this, Jumble and Couscous cells were grown to
573 stationary phase in 500 mL of 5% HN media in ASW (vol/vol). To generate pooled
574 genomic DNA for bulk segregant analysis, we grew up $5x10^6$ cells of each of the 38 F1s
575 with the rosetteless phenotype from the Rosetteless×Mapping Strain cross (Levin et al.,
576 2014), $5x10^6$ cells of each of the 30 F1s with the clumpy phenotype from the
577 Jumble×Mapping Strain cross, and $5x10^6$ cells of each of the 22 F1s with the clumpy
578 phenotype from the Couscous×Mapping Strain cross. For each cross, the F1 cells were
579 pelleted, frozen, and combined during lysis for DNA extraction. For all samples, we
580 performed a phenol-chloroform DNA extraction and used a CsCl gradient to separate *S.*
581 *rosetta* DNA from contaminating *E. pacifica* DNA by GC content (King et al., 2008).

582 Multiplexed, 150 bp paired-end libraries were prepared and sequenced on an
583 Illumina HiSeq 4000. Raw reads were trimmed with TrimmomaticPE (Bolger et al.,
584 2014) to remove low quality base calls. Trimmed reads were mapped to the *S. rosetta*
585 reference genome (Fairclough et al., 2013) using Burrows-Wheeler Aligner (Li and
586 Durbin, 2009), and we removed PCR duplicates with Picard
587 (http://broadinstitute.github.io/picard/). We realigned reads surrounding indel calls using
588 GATK (Depristo et al., 2011) and called variants using SAMtools and bcftools (Li et al.,
589 2009).

590 **Bulk segregant sequencing analysis**

591 No large region of the genome (i.e. haplotype block) was found to co-segregate
592 with the mutant phenotype in any of the crosses, likely because of the sparse, uneven
593 distribution of genetic markers and/or high recombination rates. Sequence variants from
594 the pooled samples were culled using vcftools vcf-isec (Danecek et al., 2011): (1) to
595 keep only any sequence variants in the pooled samples that were shared with the
596 parental mutant strain since any causative mutations should be present in both the
597 pooled sample and the parental mutant strain, and (2) to remove any sequence variants
598 in the pooled samples that were shared with the Mapping Strain (Isolate B), wild type
599 (previously Isolate C), or the unmutagenized control from the Rosetteless mutagenesis
600 (C2E5) since any of these sequence variants should not be causative for rosette defects
601 (Levin et al., 2014; Levin and King, 2013). The remaining variants were filtered by
602 quality: depth >2, quality score >10, and reference allele not N. The remaining list
603 represents high quality variants in the pooled population that are shared with the mutant
604 to the exclusion 3 different strains competent to form rosettes. Segregating variants
605 were determined by dividing the number of reads that map to the alternative allele by
606 the total number of high quality reads determined by SAMtools and bcftools (Li et al.,
607 2009); any variants with >99% of reads that map to the alternative allele were
608 considered variants that segregated perfectly with the mutant phenotype.

609 **Backcrosses**

610 To test the linkage of clumpy phenotype and the predicted causative mutation
611 from the bulk segregant analysis, F1s with the clumpy phenotype from the
612 Jumble×Mapping Strain and Couscous×Mapping Strain were backcrossed to the

Mapping Strain. For the Jumble F1 backcross, $1x10^6$ cells grown up from a clonally isolated F1 with the clumpy phenotype from Jumble×Mapping Strain and $1x10^6$ Mapping Strain cells were mixed, pelleted, and resuspended in 10 mL of 5% *V. fischeri* conditioned media in ASW (vol/vol). After 24 hr, the *V. fischeri* conditioned media was replaced with 25% CG media in ASW (vol/vol) and cells were plated to limiting dilution. Clonally isolated thecate populations were genotyped by PCR of the microsatellite on supercontig 1 as described above and 4 heterozygous diploids populations were identified (probability of clonal isolation 0.79-0.95). The heterozygotes were rapidly passaged for 2 weeks to induce meiosis before being plated for clonal isolation (probability of clonal isolation 0.95-0.98). 12 F2s with the clumpy phenotype and 9 F2s with the rosette phenotype were identified (Figure 2B). Their DNA was extracted using Base-Tris method described above and the region around the causal mutation was amplified. The resultant PCR product was digested for 4 hr with BfaI, which cleaves the mutant allele but not the wild type allele, and products of the digest were distinguished by agarose gel electrophoresis.

For the two Couscous F1 backcrosses, $2.5x10^5$ cells from either one of two F1s with the clumpy phenotype from Couscous×Mapping Strain cross and $2.5x10^5$ Mapping Strain cells were mixed, pelleted, resuspending in 0.5 mL of 2.5% *V. fischeri* conditioned media in ASW (vol/vol). After 24 hr, *V. fischeri* conditioned media was replaced with 25% CG media in ASW (vol/vol) and cells were plated to limiting dilution (probability of clonal isolation 0.85-0.97). Clonally isolated thecate populations were genotyped by PCR of the microsatellite on supercontig 1 as described above and 3 heterozygous diploids (6 total) were identified in each cross. Isolates were rapidly

636 passaged for 2 weeks to induce meiosis before being plated for clonal isolation
637 (probability of clonal isolation 0.88-0.97). 51 F2s with the clumpy phenotype and 38 F2s
638 with the rosette phenotype were identified (Figure 3B); their DNA was extracted using
639 Base-Tris method described above, the region around the causal mutation was
640 amplified, and the resultant PCR product was Sanger sequenced.

641 **Jumble and Couscous domain and structure prediction and alignment**

642 Protein domains encoded by *jumble* (Figure 2A) and *couscous* (Figure 3A) were
643 predicted using Interpro (Finn et al., 2017), PFAM (Finn et al., 2016), and the NCBI
644 Conserved Domain Search (Marchler-Bauer et al., 2017). Structural homology analysis
645 of Jumble was performed with Phyre2 (Kelly et al., 2015) and HHphred (Zimmermann et
646 al., 2017). The structure of the human N-acetylgalactosaminyltransferase 4 (GlcNAc T4)
647 catalytic domain (HHphred: E-value 7.5^{-19}) was aligned to the predicted Jumble
648 structure generated by HHphred using the PyMOL Molecular Graphics System, Version
649 2.0 Schrödinger, LLC (Figure 2-figure supplement 2B). Other choanoflagellate
650 homologs of *jumble* were determined by reciprocal BLAST of the 20 sequenced
651 choanoflagellate transcriptomes (Richter et al., 2018) and alignment was performed with
652 ClustalX (Larkin et al., 2007) (Figure 2-figure supplement 2A). The alignment of
653 Couscous to yeast MNN2 glycosyltransferase domains were performed with ClustalX
654 (Larkin et al., 2007) (Figure 3-figure supplement 1).

655 **Generating transgenic constructs**

656 Jumble (GenBank accession EGD72416/NCBI accession XM_004998928) and
657 Couscous (GenBank accession EGD77026/ NCBI accession XM_004990809) were
658 cloned from wild type cDNA prepared as described in Booth et al., 2018. Jumble^{lw1} was

659 cloned from cDNA prepared from the Jumble mutant. Couscous^{lw1} could not be cloned
660 from cDNA directly (possibly because of low mRNA levels due to nonsense mediate
661 decay or simply because of high GC content of the gene). However, the 1 bp deletion
662 in *Couscous^{lw1}* was confirmed by Sanger sequencing of genomic Couscous DNA. Site
663 directed mutagenesis of the wild type gene was used to generate the mutant allele.
664 For complementation (Figure 2C,D and 3C,D), constructs were generated from a
665 plasmid with a pUC19 backbone with a 5' *S. rosetta* elongation factor L (*efl)* promoter,
666 monomeric teal fluorescent protein (*mTFP)*, and the 3' UTR from actin (Addgene ID
667 NK633) (Booth et al., 2018). A puromycin resistance gene was synthesized as a gene
668 block and codon optimized for *S. rosetta*. The puromycin resistance gene (puro) was
669 inserted after the *efl* promoter and separated from fluorescent reporters by self-
670 cleaving 2A peptide from the porcine virus (P2A) (Kim et al., 2011). Copies of *jumble,*
671 *jumble^{lw1}, couscous,* and *couscous^{lw1}*were inserted either 5' or 3' of the mTFP and
672 separated from mTFP by a flexible linker sequence (SGGSGGS) through Gibson
673 cloning.
674 For fluorescent localization (Figure 2E-H, Figure 2-figure supplement 3B, Figure
675 3-figure supplement B,C), constructs were generated from a pUC19 backbone with a
676 5' *S. rosetta* elongation factor L (*efl)* promoter, mWasabi, and 3' UTR from actin.
677 Copies of *jumble, jumble^{lw1},* and *couscous* were inserted either 5' of the mWasabi
678 separated by a flexible linker sequence (SGGSGGS) through Gibson cloning. Plasma
679 membrane and ER markers from Booth et al., 2018 were used as previously described
680 (Addgene ID NK624 and NK644).
681 **_S. rosetta_ transfection and transgene expression**

Transfection protocol was followed as described in Booth et al., 2018 (http://www.protocols.io/groups/king-lab). Two days prior to transfection, a culture flask (Corning, Cat. No. 353144) was seeded with Jumble, Couscous, or wild type cells at a density of 5,000 cells/ml in 200 ml of 1x HN Media. After 36-48 hr of growth, bacteria were washed away from the cells in three consecutive rounds of centrifugation and resuspension in sterile AK seawater. After the final wash, the cells were resuspended in a total volume of 100 µl AK and counted on a Luna-FL automated cell counter (Logos Biosystems). The remaining cells were diluted to a final concentration of $5x10^7$ cells/ml and divided into 100 µl aliquots. Each aliquot of cells pelleted at 2750 x g, resuspend in priming buffer (40 mM HEPES-KOH, pH 7.5; 34 mM Lithium Citrate; 50 mM L-Cysteine; 15% (w/v) PEG 8000; and 1 µM papain), and incubated at room temperature for 30 mins to remove extracellular material coating the cells. Priming buffer was quenched with 50 mg/ml bovine serum albumin-fraction V (Sigma). Cells were pelleted at 1250 x g and resuspend in 25 µl of SF buffer (Lonza). Each transfection reaction was prepared by adding 2 µl of "primed" cells to a mixture of 16 µl of SF buffer, 2 µl of 20 µg/ µl pUC19; 1 µl of 250 mM ATP, pH 7.5; 1 µl of 100 mg/ml Sodium Heparin; and 1 µl of each reporter DNA construct at 5 µg/µl. Transfections were carried out in 96-well nucleofection plate (Lonza) in a Nucleofector 4d 96-well Nucleofection unit (Lonza) with the CM-156 pulse. Immediately after nucleofection, 100 µl of ice-cold recovery buffer (10 mM HEPES-KOH, pH 7.5; 0.9 M Sorbitol; 8% (w/v) PEG 8000) was added to the cells and incubated for 5 min. The whole volume of the transfection reaction plus the recovery buffer was transferred to 1 ml of 1x HN media in a 12-well plate. After cells

recovered for 1 hr, 5 µl of a 10 mg frozen *E. pacifica* pellet resuspend in 1 ml of AK seawater was added to each well and RIFs were added if looking at rosette induction.

Transgenic Complementation

For complementation, Jumble mutants were transfected with the following constructs: (1) *pefl-puro-P2A-Jumble-mTFP*, (2) *pefl-puro-P2A-Jumble^{lw1}-mTFP*, (3) *pefl-puro-P2A-mTFP-Jumble*, (4) *pefl-puro-P2A-mTFP-Jumble^{lw1}*, and (5) *pefl-puro-P2A-mTFP*; and Couscous with the following constructs: (1) *pefl-puro-P2A-Couscous-mTFP*, (2) *pefl-puro-P2A-Couscous^{lw1}-mTFP*, (3) *pefl-puro-P2A-mTFP-Couscous*, (4) *pefl-puro-P2A-mTFP-Couscous^{lw1}*, and (5) *pefl-puro-P2A-mTFP*. Transfected cells were grown an additional 24 hr after transfection to allow for transgene expression, and then 40 µg/ml puromycin was added for selection. Selection occurred for 48 hr before rosette induction was counted by hemocytometer. After vortexing for 15 sec and fixing with formaldehyde, 200 cells of each transfection well were counted on a hemocytometer to determine percentage of cells in rosettes (Figure 2C, Figure 3C). Complementation was repeated on 2 biological replicates with 3 technical transfection replicates each. Representative rosette images (Figure 2D, Figure 3D) were taken on by confocal microscopy using Zeiss Axio Observer LSM 880 a C-Apochromat 40x/NA1.20 W Korr UV-Vis-IR water immersion objective.

Live cell imaging

Glass-bottom dishes for live cell imaging were prepared by corona-treating and poly-D-lysine coating as described in Booth et al., 2018. Transfected cells were prepared for microscopy by pelleting 1-2 ml of cells and resuspend in 200 µl of 4/5 ASW with 100 mM LiCl to slow flagellar beating. Cells were plated on glass-bottom dishes

and covered by 200 µl of 20% (w/v) Ficoll 400 dissolved in 4/5 ASW with 100 mM LiCl. Confocal microscopy was performed on a Zeiss Axio Observer LSM 880 with an Airyscan detector and a 63x/NA1.40 Plan-Apochromatic oil immersion objective. Confocal stacks were acquired in super-resolution mode using ILEX Line scanning and two-fold averaging and the following settings: 35 nm x 35 nm pixel size, 100 nm z-step, 0.9-1.0 µsec/pixel dwell time, 850 gain, 458 nm laser operating at 1-6% laser power, 561 nm laser operating at 1-2% laser power, 458/561 nm multiple beam splitter, and 495-550 nm band-pass/570 nm long-pass filter. Images were processed using the automated Airyscan algorithm (Zeiss).

Lectin staining and jacalin quantification

All fluorescein lectins from kits I, II, and III from Vector Lab (FLK-2100, FLK-4100, and FLK-4100) were tested for recognition in wild type, Jumbled, and Couscous (Table S5). Cells were plated on poly-D-Lysine coated wells of a 96-well glass bottom plate, lectins were added at a concentration of 1:200 and imaged immediately using Zeiss Axio Observer.Z1/7 Widefield microscope with a Hammatsu Orca-Flash 4.0 LT CMOS Digital Camera and a 20x objective. For further jacalin image analysis (Figure 4), cells were plated on a poly-D-Lysine coated glass bottom dish, 1:400 fluorescein labelled-jacalin and 1:200 lysotracker Red DN-99 (overloaded to visualize the cell body) and were imaged immediately by confocal microscopy using Zeiss Axio Observer LSM 880 a 63x/NA1.40 Plan-Apochromatic oil immersion objective. Images were taken with the following settings: 66 nm x 66 nm pixel size, 64 nm z-step, 0.34 µsec/pixel dwell time, 488 nm laser operating at 0.2% laser power with 700 master gain, and 561 nm laser operating at 0.0175% laser power with 750 master gain. Fifteen unique fields of

view chosen based on lysotracker staining. Induced cells were treated with OMVs 24 hr before imaging.

To process images, Z-stack images were max projected using ImageJ. Individual cells were chosen based on the ability to clearly see a horizontally oriented collar by lysotracker and cropped to only include a single cell. The maximum fluorescence intensity pixel of the jacalin channel was determined for the cropped image and was used to normalize the fluorescence intensity. To measure jacalin staining around the cell body, a line was drawn using only the lysotracker staining from the point where the collar and the cell body meet on one side of the cell around the cell to the other and the fluorescence intensity was measured along the line. To compare between cells, the lines drawn around the cell body were one-dimensional interpolated in R to include 150 points and normalized to the length of the line. The average fluorescence intensity was plotted over the length of the line drawn around the cell body for Jumble, Couscous, and wild type induced and uninduced with a 95% confidence interval (Figure 4F). Measurements were taken from two biological replicates with at least 59 cells in total from each condition.

Wild type and mutant clumping assays

Wild type cells transfected with the puromycin resistance gene and mWasabi separated by the P2A self-cleaving peptide under the *efl* promoter and maintained in 40 µg/mL puromycin to enrich for positive transformants. For clumping assays, equal numbers of mWasabi-wt cells either uninduced or induced to form rosettes were mixed with either Jumble or Couscous, vortexed, and plated on BSA treated 8-well glass bottom dishes. DIC and fluorescent images were obtained after 30 mins using Zeiss

773 Axio Observer.Z1/7 Widefield microscope with a Hammatsu Orca-Flash 4.0 LT CMOS
774 Digital Camera and a 40x/NA1.40 Plan-Apochromatic lens (Figure 1-figure supplement
775 1).

776 **Wild type and mutant growth curves**

777 All cells strains were plated at a density of $1x10^4$ cells/ml in 3 ml HN media.
778 Every 12 hr an aliquot of cells was vortexed vigorously for 15 sec, fixed with
779 formaldehyde, and counted by hemacytometer. Curves were generated from the
780 average ± SD from 2 biological replicates with 3 technical replicates each (Figure 1-
781 figure supplement 2).

782 **Jacalin Western blot**

783 Whole cell lysates were made from pelleting $1x10^7$ cells at 4C at 3,000 x g and
784 resuspending in lysis buffer (20 mM Tris-HCl, pH 8.0; 150 mM KCl; 5 mM MgCl2; 250
785 mM Sucrose; 1 mM DTT; 10 mM Digitonin; 1 mg/ml Sodium Heparin; 1 mM Pefabloc
786 SC; 0.5 U/µl DNaseI; 1 U/µl SUPERaseIN). Cells were incubated in lysis buffer for 10
787 min on ice and passed through 30G needle 5x. Insoluble material was pelleted at 6,000
788 x g for 10 min at 4C. Lysate ($1x10^6$ cells/sample) was run on 4-20% TGX mini-gel (Bio-
789 Rad) for 45 min at 200 V and transferred onto 0.2 µm nitrocellulose membrane using
790 Trans-Blot Turbo Transfer System (Bio-Rad) with semi-dry settings 25V for min. The
791 blot was blocked for 30 min with Odyssey PBS Block (Li-cor). The blot was probed with
792 biotinylated jacalin (1:4,000; Vector Labs) and E7 anti-tubulin antibody (1:10,000;
793 Developmental Studies Hybridoma Bank) diluted in block for 1 hr, and then with IRDye
794 800 streptavidin (1:1,000; Li-cor) and IRDye 700 mouse (1:1,000; Li-cor) in PBST [PBS

795 with %1 Tween 20 (v/v)]. Blot was imaged on Licor Odyssey (Figure 4-figure
796 supplement 1).

797 **Rosetteless immunofluorescence staining and imaging**

798 Immunofluorescence (Figure 4-figure supplement 2) was performed previously
799 described in Levin et al., 2014 with the modifications for better cytoskeleton
800 preservation described in Booth et al., 2018. Two mL of dense wild type, Jumble, and
801 Couscous cells, that were either uninduced or induced with RIFs for 24 hr, were allowed
802 to settle on poly-L-lysine coated coverslips (BD Biosciences) for 30 min. Cells were
803 fixed in two steps: 6% acetone in cytoskeleton buffer (10 mM MES, pH 6.1; 138 KCl, 3
804 mM $MgCl_2$; 2 mM EGTA; 675 mM Sucrose) for 5 and 4% formaldehyde with diluted in
805 cytoskeleton buffer for 20 min. The coverslips were gently washed three times with
806 cytoskeleton buffer. Cells were permeabilized with permeabilization buffer [100 mM
807 PIPES, pH 6.95; 2 mM EGTA; 1 mM $MgCl_2$; 1% (w/v) bovine serum albumin-fraction V;
808 0.3% (v/v Triton X-100)] for 30 min. Cells were stained with the anti-Rosetteless
809 genomic antibody at 3.125 ng/µl (1:400), E7 anti-tubulin antibody (1:1000;
810 Developmental Studies Hybridoma Bank), Alexa fluor 488 anti-mouse and Alexa fluor
811 647 anti-rabbit secondary antibodies (1:1000 each; Molecular Probes), and 6 U/ml
812 rhodamine phalloidin (Molecular Probes) before mounting in Prolong Gold antifade
813 reagent with DAPI (Molecular Probes).

814 Images were acquired on a Zeiss LSM 880 Airyscan confocal microscope with a
815 63x objective (as described for live cell imaging) by frame scanning in the super-
816 resolution mode with the following settings: 30 nm x 30 nm pixel size; 100 nm z-step;
817 561 nm laser operating at 1.5% power with 700 master gain, and 488 nm laser

operating at 2.0% power with 800 master gain. Wild type rosettes were imaged with 633 nm laser operating at 0.3% laser power and 650 master gain to prevent overexposure of Rosetteless, but all other conditions were operating at 2% laser power and 650 master gain in the 633 nm channel.

ACKNOWLEDGEMENTS

Hannah Elzinga, Lily Helfrich, and Max Coyle helped with experiments and reagent preparation. We thank Iswar Hariharan and members of the King lab for helpful discussions, research support, mutant naming suggestions, and comments on the manuscript, especially Kayley Hake, Ben Larson, Tess Linden, and Thibaut Brunet. This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 OD018174 Instrumentation Grant.

831 **REFERENCES**

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

835 Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina
836 sequence data. *Bioinformatics* **30**:2114–2120. doi:10.1093/bioinformatics/btu170

837 Booth DS, Szmidt-Middleton H, King N. 2018. Choanoflagellate transfection illuminates
838 their cell biology and the ancestry of animal septins. *bioRxiv*.
839 doi:https://doi.org/10.1101/343111

840 Brunet T, King N. 2017. The Origin of Animal Multicellularity and Cell Differentiation.
841 *Developmental Cell* **43**:124–140. doi:10.1016/j.devcel.2017.09.016

842 Cambi A, Koopman M, Figdor CG. 2005. How C-type lectins detect pathogens. *Cellular*
843 *Microbiology* **7**:481–488. doi:10.1111/j.1462-5822.2005.00506.x

844 Carvalho S, Catarino TA, Dias AM, Kato M, Almeida A, Hessling B, Figueiredo J,
845 Gartner F, Sanches JM, Ruppert T, Miyoshi E, Pierce M, Carneiro F, Kolarich D,
846 Seruca R, Yamaguchi Y, Taniguchi N, Reis CA, Pinho SS. 2016. Preventing E-
847 cadherin aberrant N-glycosylation at Asn-554 improves its critical function in gastric
848 cancer. *Oncogene* **35**:1619–1631. doi:10.1038/onc.2015.225

849 Colin Hughes R. 1992. Lectins as cell adhesion molecules. *Current Opinion in Structural*
850 *Biology* **2**:687–692. doi:10.1016/0959-440X(92)90202-I

851 Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE,
852 Lunter G, Marth GT, Sherry ST, McVean G, Durbin R. 2011. The variant call format
853 and VCFtools. *Bioinformatics* **27**:2156–2158. doi:10.1093/bioinformatics/btr330

854 Dayel MJ, Alegado RA, Fairclough SR, Levin TC, Nichols SA, McDonald K, King N.
855 2011. Cell differentiation and morphogenesis in the colony-forming choanoflagellate
856 *Salpingoeca rosetta*. *Developmental Biology* **357**:73–82.
857 doi:10.1016/j.ydbio.2011.06.003

858 Depristo MA, Banks E, Poplin R, Garimella K V., Maguire JR, Hartl C, Philippakis AA,
859 Del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM,
860 Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. 2011. A framework
861 for variation discovery and genotyping using next-generation DNA sequencing
862 data. *Nature Genetics* **43**:491–501. doi:10.1038/ng.806

863 Doitsidou M, Poole RJ, Sarin S, Bigelow H, Hobert O. 2010. *C. elegans* mutant
864 identification with a one-step whole-genome-sequencing and SNP mapping
865 strategy. *PLoS ONE* **5**:1–7. doi:10.1371/journal.pone.0015435

El-Battari A. 2006. Autofluorescent Proteins for Monitoring the Intracellular Distribution of Glycosyltransferases. *Methods in Enzymology* **416**:102–120. doi:10.1016/S0076-6879(06)16007-3

Fairclough SR, Chen Z, Kramer E, Zeng Q, Young S, Robertson HM, Begovic E, Richter DJ, Russ C, Westbrook MJ, Manning G, Lang BF, Haas B, Nusbaum C, King N. 2013. Premetazoan genome evolution and the regulation of cell differentiation in the choanoflagellate *Salpingoeca rosetta*. *Genome Biology* **14**:1–15. doi:10.1186/gb-2013-14-2-r15

Fairclough SR, Dayel MJ, King N. 2010. Multicellular development in a choanoflagellate. *Current Biology* **20**:875–876. doi:10.1016/j.cub.2010.09.014

Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, Chang HY, Dosztanyi Z, El-Gebali S, Fraser M, Gough J, Haft D, Holliday GL, Huang H, Huang X, Letunic I, Lopez R, Lu S, Marchler-Bauer A, Mi H, Mistry J, Natale DA, Necci M, Nuka G, Orengo CA, Park Y, Pesseat S, Piovesan D, Potter SC, Rawlings ND, Redaschi N, Richardson L, Rivoire C, Sangrador-Vegas A, Sigrist C, Sillitoe I, Smithers B, Squizzato S, Sutton G, Thanki N, Thomas PD, Tosatto SCE, Wu CH, Xenarios I, Yeh LS, Young SY, Mitchell AL. 2017. InterPro in 2017-beyond protein family and domain annotations. *Nucleic Acids Research* **45**:D190–D199. doi:10.1093/nar/gkw1107

Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A. 2016. The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Research* **44**:D279–D285. doi:10.1093/nar/gkv1344

Geijtenbeek TBH, Gringhuis SI. 2009. Signalling through C-type lectin receptors: shaping immune responses. *Nature Reviews Immunology* **9**:465–479. doi:10.1038/nri2569

Gilbert OM, Foster KR, Mehdiabadi NJ, Strassmann JE, Queller DC. 2007. High relatedness maintains multicellular cooperation in a social amoeba by controlling cheater mutants. *Proceedings of the National Academy of Sciences* **104**:8913–8917. doi:10.1073/pnas.0702723104

Granovsky M, Fata J, Pawling J, Muller WJ, Khokha R, Dennis JW. 2000. Suppression of tumor growth and metastasis in Mgat5-deficient mice. *Nature Medicine* **6**:306–312. doi:10.1038/73163

Guo HB, Johnson H, Randolph M, Pierce M. 2009. Regulation of homotypic cell-cell adhesion by branched N-glycosylation of N-cadherin extracellular EC2 and EC3 domains. *Journal of Biological Chemistry* **284**:34986–34997. doi:10.1074/jbc.M109.060806

Herron MD, Borin JM, Boswell JC, Walker J, Chen I-CK, Knox CA, Boyd M, Rosenzweig F, Ratcliff WC. 2018. De novo origin of multicellularity in response to

905 predation. *bioRxiv* 247361. doi:10.1101/247361

906 Hirohashi S, Kanai Y. 2003. Cell adhesion system and human cancer morphogenesis.
907 *Cancer Science* **94**:575–581. doi:10.1111/j.1349-7006.2003.tb01485.x

908 Ho DH, Badellino K, Baglia FA, Walsh PN. 1998. A Binding Site for Heparin in the Apple
909 3 Domain of Factor XI. *Journal of Biological Chemistry* **273**:16382–16390.

910 Kelly LA, Mezulis S, Yates C, Wass M, Sternberg M. 2015. The Phyre2 web portal for
911 protein modelling, prediction, and analysis. *Nature Protocols* **10**:845–858.
912 doi:10.1038/nprot.2015-053

913 Kim JH, Lee SR, Li LH, Park HJ, Park JH, Lee KY, Kim MK, Shin BA, Choi SY. 2011.
914 High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in
915 human cell lines, zebrafish and mice. *PLoS ONE* **6**:1–8.
916 doi:10.1371/journal.pone.0018556

917 King N, Westbrook MJ, Young SL, Kuo A, Abedin M, Chapman J, Fairclough S, Hellsten
918 U, Isogai Y, Letunic I, Marr M, Pincus D, Putnam N, Rokas A, Wright KJ, Zuzow R,
919 Dirks W, Good M, Goodstein D, Lemons D, Li W, Lyons JB, Morris A, Nichols S,
920 Richter DJ, Salamov A, Bork P, Lim WA, Manning G, Miller WT, McGinnis W,
921 Shapiro H, Tjian R, Grigoriev I V., Rokhsar D, Sequencing JGI, Bork P, Lim WA,
922 Manning G, Miller WT, McGinnis W, Shapiro H, Tjian R, Grigoriev I V., Rokhsar D.
923 2008. The genome of the choanoflagellate *Monosiga brevicollis* and the origin of
924 metazoans. *Nature* **451**:783–8. doi:10.1038/nature06617

925 Kopito RR. 1997. ER quality control: the cytoplasmic connection. *Cell* **88**:427–30.
926 doi:10.1016/S0092-8674(00)81881-4

927 Kuzdzal-Fick JJ, Fox SA, Strassman JE, Queller DC. 2011. High Relatedness Is
928 Necessary and Sufficient to Maintain Multicellularity in Dictyostelium. *Science*
929 **334**:1548–1551. doi:10.3334/ORNLDAAC/797

930 Lairson LL, Henrissat B, Davies GJ, Withers SG. 2008. Glycosyltransferases:
931 Structures, Functions, and Mechanisms. *Annual Review of Biochemistry* **77**:521–
932 555. doi:10.1146/annurev.biochem.76.061005.092322

933 Larkin MA, Blackshields G, Brown NP, Chenna R, Mcgettigan PA, McWilliam H,
934 Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG.
935 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**:2947–2948.
936 doi:10.1093/bioinformatics/btm404

937 Larsen ISB, Narimatsu Y, Joshi HJ, Siukstaite L, Harrison OJ, Brasch J, Goodman KM,
938 Hansen L, Shapiro L, Honig B, Vakhrushev SY, Clausen H, Halim A. 2017.
939 Discovery of an O-mannosylation pathway selectively serving cadherins and
940 protocadherins. *Proceedings of the National Academy of Sciences* **114**:201708319.
941 doi:10.1073/pnas.1708319114

Leadbeater BS. 2015. The choanoflagellates: evolution, biology, and ecology, 1st ed. Cambridge: Cambridge University Press.

Leigh EG, Smith JM, Szathmary E. 1995. The Major Transitions of Evolution. *Evolution* **49**:1302. doi:10.2307/2410462

Leshchiner I, Alexa K, Kelsey P, Adzhubei I, Austin-Tse CA, Cooney JD, Anderson H, King MJ, Stottmann RW, Garnaas MK, Ha S, Drummond IA, Paw BH, North TE, Beier DR, Goessling W, Sunyaev SR. 2012. Mutation mapping and identification by whole-genome sequencing. *Genome Research* **22**:1541–1548. doi:10.1101/gr.135541.111

Levin TC, Greaney AJ, Wetzel L, King N. 2014. The Rosetteless gene controls development in the choanoflagellate *S. rosetta. eLife* **3**:e04070. doi:10.7554/eLife.04070

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

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**:1754–1760. doi:10.1093/bioinformatics/btp324

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**:2078–2079. doi:10.1093/bioinformatics/btp352

Lister R, Gregory BD, Ecker JR. 2009. Next is now: new technologies for sequencing of genomes, transcriptomes, and beyond. *Current Opinion in Plant Biology* **12**:107–118. doi:10.1016/j.pbi.2008.11.004

Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Geer LY, Bryant SH. 2017. CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures. *Nucleic Acids Research* **45**:D200–D203. doi:10.1093/nar/gkw1129

Müller WE, Zahn RK, Kurelec B, Müller I, Uhlenbruck G, Vaith P. 1979. Aggregation of sponge cells. *Journal of Biological Chemistry* **254**:1280–1287.

Nedashkovskaya OI, Kim SB, Vancanneyt M, Lysenko AM, Shin DS, Park MS, Lee KH, Jung WJ, Kalinovskaya NI, Mikhailov V V., Bae KS, Swings J. 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. doi:10.1099/ijs.0.64156-0

Nichols SA, Roberts BW, Richter DJ, Fairclough SR, King N. 2012. Origin of metazoan cadherin diversity and the antiquity of the classical cadherin/β-catenin complex. *Proceedings of the National Academy of Sciences* **109**:13046–51.

doi:10.1073/pnas.1120685109

Pomraning KR, Smith KM, Freitag M. 2011. Bulk segregant analysis followed by high-throughput sequencing reveals the *Neurospora* cell cycle gene, ndc-1, to be allelic with the gene for ornithine decarboxylase, spe-1. *Eukaryotic Cell* **10**:724–733. doi:10.1128/EC.00016-11

Ratcliff WC, Denison RF, Borrello M, Travisano M. 2012. Experimental evolution of multicellularity. *Proceedings of the National Academy of Sciences* **109**:1595–1600. doi:10.1073/pnas.1115323109

Ratcliff WC, Herron MD, Howell K, Pentz JT, Rosenzweig F, Travisano M. 2013. Experimental evolution of an alternating uni- and multicellular life cycle in *Chlamydomonas reinhardtii*. *Nature Communications* **4**:1–7. doi:10.1038/ncomms3742

Rayner JC, Munro S. 1998. Identification of the MNN2 and MNN5 mannosyltransferases required for forming and extending the mannose branches of the outer chain mannans of *Saccharomyces cerevisiae*. *Journal of Biological Chemistry* **273**:26836–26843. doi:10.1074/jbc.273.41.26836

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

Ruiz-Trillo I, Roger AJ, Burger G, Gray MW, Lang BF. 2008. A phylogenomic investigation into the origin of Metazoa. *Molecular Biology and Evolution* **25**:664–672. doi:10.1093/molbev/msn006

Ruoslahti E. 1996. Brain extracellular matrix. *Glycobiology* **6**:489–492. doi:https://doi.org/10.1093/glycob/6.5.489

Sawaguchi S, Varshney S, Ogawa M, Sakaidani Y, Yagi H, Takeshita K, Murohara T, Kato K, Sundaram S, Stanley P, Okajima T. 2017. O-GlcNAc on NOTCH1 EGF repeats regulates ligand-induced Notch signaling and vascular development in mammals. *eLife* **6**. doi:10.7554/eLife.24419

Schalchian-Tabrizi K, Minge MA, Espelund M, Orr R, Ruden T, Jakobsen KS, Cavalier-Smith T. 2008. Multigene phylogeny of Choanozoa and the origin of animals. *PLoS ONE* **3**. doi:10.1371/journal.pone.0002098

Schneeberger K, Ossowski S, Lanz C, Juul T, Petersen AH, Nielsen KL, Jørgensen JE, Weigel D, Andersen SU. 2009. SHOREmap: Simultaneous mapping and mutation identification by deep sequencing. *Nature Methods* **6**:550–551. doi:10.1038/nmeth0809-550

Sebé-Pedrós A, Degnan BM, Ruiz-Trillo I. 2017. The origin of Metazoa: a unicellular perspective. *Nature Reviews Genetics* **18**:498–512. doi:10.1038/nrg.2017.21

1015 Strahl-Bolsinger S, Gentzsch M, Tanner W. 1999. Protein O-mannosylation. *Biochimica*
1016 *et Biophysica Acta - General Subjects* **1426**:297–307. doi:10.1016/S0304-
1017 4165(98)00131-7

1018 Stratford M. 1992. Yeast Flocculation : Receptor Definition by mnn Mutants and
1019 Concanavalin A. *Yeast* **8**:635–645.

1020 Švajger U, Anderluh M, Jeras M, Obermajer N. 2010. C-type lectin DC-SIGN: An
1021 adhesion, signalling and antigen-uptake molecule that guides dendritic cells in
1022 immunity. *Cellular Signalling* **22**:1397–1405. doi:10.1016/J.CELLSIG.2010.03.018

1023 Tordai H, Bányai L, Patthy L. 1999. The PAN module: The N-terminal domains of
1024 plasminogen and hepatocyte growth factor are homologous with the apple domains
1025 of the prekallikrein family and with a novel domain found in numerous nematode
1026 proteins. *FEBS Letters* **461**:63–67. doi:10.1016/S0014-5793(99)01416-7

1027 Tu L, Banfield DK. 2010. Localization of Golgi-resident glycosyltransferases. *Cellular*
1028 *and Molecular Life Sciences* **67**:29–41. doi:10.1007/s00018-009-0126-z

1029 Voz ML, Coppieters W, Manfroid I, Baudhuin A, von Berg V, Charlier C, Meyer D,
1030 Driever W, Martial JA, Peers B. 2012. Fast homozygosity mapping and
1031 identification of a zebrafish enu-induced mutation by whole-genome sequencing.
1032 *PLoS ONE* **7**:1–10. doi:10.1371/journal.pone.0034671

1033 Wenger JW, Schwartz K, Sherlock G. 2010. Bulk segregant analysis by high-throughput
1034 sequencing reveals a novel xylose utilization gene from *Saccharomyces cerevisiae*.
1035 *PLoS Genetics* **6**:18. doi:10.1371/journal.pgen.1000942

1036 Wiggins C a, Munro S. 1998. Activity of the yeast MNN1 alpha-1,3-mannosyltransferase
1037 requires a motif conserved in many other families of glycosyltransferases.
1038 *Proceedings of the National Academy of Sciences of the United States of America*
1039 **95**:7945–7950. doi:10.1073/pnas.95.14.7945

1040 Woznica A, Cantley AM, Beemelmanns C, Freinkman E, Clardy J, King N. 2016.
1041 Bacterial lipids activate, synergize, and inhibit a developmental switch in
1042 choanoflagellates. *Proceedings of the National Academy of Sciences* **113**:7894–
1043 7899. doi:10.1073/pnas.1605015113

1044 Woznica A, Gerdt JP, Hulett RE, Clardy J, King N. 2017. Mating in the Closest Living
1045 Relatives of Animals Is Induced by a Bacterial Chondroitinase. *Cell* **170**:1175–
1046 1183.e11. doi:10.1016/j.cell.2017.08.005

1047 Zelensky AN, Gready JE. 2005. The C-type lectin-like domain superfamily. *FEBS*
1048 *Journal* **272**:6179–6217. doi:10.1111/j.1742-4658.2005.05031.x

1049 Zhang L, Zhang Y, Ten Hagen KG. 2008. A mucin-type O-glycosyltransferase
1050 modulates cell adhesion during *Drosophila* development. *Journal of Biological*
1051 *Chemistry* **283**:34076–34086. doi:10.1074/jbc.M804267200

1052 Zhao Y, Sato Y, Isaji T, Fukuda T, Matsumoto A, Miyoshi E, Gu J, Taniguchi N. 2008.
1053 Branched N-glycans regulate the biological functions of integrins and cadherins.
1054 *FEBS Journal* **275**:1939–1948. doi:10.1111/j.1742-4658.2008.06346.x

1055 Zimmermann L, Stephens A, Nam SZ, Rau D, Kübler J, Lozajic M, Gabler F, Söding J,
1056 Lupas AN, Alva V. 2017. A Completely Reimplemented MPI Bioinformatics Toolkit
1057 with a New HHpred Server at its Core. *Journal of Molecular Biology* **430**:2237–
1058 2243. doi:10.1016/j.jmb.2017.12.007

1059
1060

FIGURE LEGENDS

Figure 1. Mutant cells aggregate and fail to form rosettes. (A) Wild type cells are unicellular or form linear chains in the absence of rosette inducing factors (RIFs) and develop into organized spherical rosettes. Rosettes are resistant to shear force and survive vortexing. Four class C mutants — Seafoam, Soapsuds, Couscous, and Jumble — form disordered clumps of cells in the presence and absence of RIFs. The clumps are not resistant to vortexing and fall apart into single cells. **(B)** Class C mutants do not form any detectable rosettes. Rosette development was measured as the % of cells in rosettes after 48 hr in the presence of RIFs and is shown as mean ± SEM. n.d. = no detected rosettes. **(C)** Class C mutants quickly aggregated into large clumps after disruption by vortexing. After vortexing, wild type and mutant cells were incubated for 30 minutes in the absence of RIFs and clump sizes were quantified by automated image analysis. Data are presented as violin boxplots, showing the median cell number (middle line), interquartile range (white box), and range excluding outliers (thin line). All mutants had significantly larger masses of cells (two-tailed t-test ****p<0.0001) than found in cultures of wild type cells. **(D)** Clumping occurred within minutes after vortexing in the Class C mutants without RIFs, revealing that the clumps form by aggregation and not through cell division. DIC images obtained at 0, 15, and 30 minutes post-vortexing. Scale bar = 20 μm.

Figure 2. Jumble maps to a predicted glycosyltransferase that localizes to the Golgi apparatus. (A) Jumble has a predicted transmembrane domain (marked TM) and secondary structure (alpha helices marked by black rectangles). Structural homology

algorithms predict that Jumble is related to well-characterized glycosyltransferases (Figure 2-figure supplement 2). The mutant gene has a T to C mutation at nucleotide 1109 that causes an amino acid substitution of proline to leucine at amino acid position 305. **(B)** A backcross of a mutant F1 progeny to the Mapping Strain yielded nine rosette-forming F2 isolates with the wild type T allele and twelve clumpy F2 isolates with the *jumble*^{lw1} C allele. The inheritance significantly deviated from expected Mendelian inheritance of unlinked traits and confirmed the tight linkage between the *jumble*^{lw1} allele to the clumpy, rosetteless phenotype. X^2 = Chi-squared value, d.f. = degrees of freedom. **(C,D)** Transgenic expression of *jumble-mTFP* and *mTFP-jumble* rescued rosette development in the Jumble mutant, but *jumble*^{lw1}*-mTFP*, *mTFP-jumble*^{lw1}, or *mTFP* did not. RIFs were added immediately after transfection and 40 µg/mL puromycin was added 24 hours post-transfection to select for transformants. **(C)** Rosette development was measured as the % of cells in rosettes 72 hr post-transfection and shown as mean ± SD. n.d. = no detected rosettes. (n=200 cells counted from each of 3 technical replicates; 2 biological replicates). **(D)** Rosettes transgenically complemented with *jumble-mTFP* in the Jumble mutant appeared phenotypically wild type and most cells in rosettes had detectable fluorescent expression at the apical base of the cell. Representative rosette shown. **(E-H)** To examine localization, Jumble-mWasabi or Jumble^{lw1}-mWasabi (cyan) under the *efl* promoter were co-expressed with membrane marker-mCherry (magenta) in wild type *S. rosetta.* Jumble-mWasabi localizes to the apical pole of cells grown **(E)** without RIFs or **(G)** with RIFs, consistent with the localization of the Golgi apparatus. When expressed in otherwise wild type cells grown **(F)** without RIFs or **(H)** with RIFs, the mutant Jumble^{lw1}-mWasabi incorrectly localizes to

the ER and food vacuole. Boxes indicate the inferred location of the Golgi apparatus at the apical pole of the cell. The food vacuole (asterisk) was often visualized due to autofluoresence from ingested bacteria or through accumulation of the fluorescent markers in the food vacuole, perhaps through autophagy. For reference, arrows indicate the base of the flagellum although the flagellum may not be visible in the plane of focus shown. Scale bar = 5 μm.

Figure 3. Couscous maps to a predicted mannosyltransferase with a PAN/Apple domain. (A) Couscous has a predicted signal sequence (S), a PAN/Apple domain (PAN), and a mannosyltransferase domain. The causative lesion is a 1-base pair deletion at nucleotide position 2447 that causes a frameshift at amino acid 728, resulting in 75 amino acids that do not align between the wild type and mutant (Cous) sequences, and an early stop codon (*) at amino acid 803. **(B)** Independent backcrosses of two individual mutant F1 progeny to the Mapping Strain yielded 38 rosette-forming F2 isolates with the wild type GCCC allele and 51 clumpy F2 isolates with the *couscous*^{lw1} GCC allele. The inheritance significantly deviated from expected Mendelian inheritance of unlinked traits and confirmed the tight linkage between the *couscous*^{lw1} allele to the clumpy, rosetteless phenotype. X^2 = Chi-squared value, d.f. = degrees of freedom. **(C, D)** Rosette formation in Couscous mutant cells can be rescued by transgenic expression of *couscous-mTFP* or *mTFP-couscous*, but not *couscous*^{lw1}-*mTFP*, *mTFP-couscous*^{lw1}, or *mTFP* alone. RIFs were added immediately after transfection and 40 μg/mL puromycin was added 24 hours post-transfection to select for positive transformants. **(C)** Rosette development (mean ± SD) was measured as the %

of cells in rosettes 72 hr after transfection and treatment with RIFs . n.d. = no detected rosettes. (n=200 cells counted from each of 3 technical replicates; 2 biological replicates). **(D)** Rosettes transgenically complemented with *couscous-mTFP* in the Couscous mutant appeared phenotypically wild type. Representative rosette shown. Scale bar = 5 μm.

Figure 4. Disruption of basal glycosylation patterns in Jumble and Couscous mutants. FITC-labelled jacalin binds the apical and basal poles of wild type single cells **(B)** and becomes enriched in the ECM in the center of rosettes **(A, B'** boxed region from **A)**. Although FITC-jacalin staining appeared normal at the apical poles of Jumble (**C)** and Couscous **(D)** mutant cells, FITC-jacalin staining at the basal poles of cells was undetectable in cells grown either in the absence (-RIFs; **C, D**) or presence (+RIFs; **C', D'**) RIFs. Arrows mark the apical pole and arrowheads mark the basal pole. **(E)** Cartoon depicts how jacalin fluorescence was measured. Starting with micrographs of FITC-jacalin stained cells, a line was drawn tracing from one edge of the collar around the cell body to the other edge of the collar, and the underlying fluorescent signal was normalized for cell size and background intensity. **(F)** The average normalized fluorescence intensity of jacalin measured in at least 59 cells for each condition was graphed against the normalized length of the cell body (n=2 biological replicates). Jumble and Couscous -/+RIFs have reduced jacalin binding at the basal pole compared to wild type -/+RIFs. Gray shadows indicate 95% confidence intervals. Scale bar = 5 μm.

Figure 5. Model for promiscuous clumping in rosette defective Class C mutants. Wild type *S. rosetta* has a glycosylated basal patch of ECM (red) as marked by the lectin jacalin that becomes enriched during the course of rosette formation. The Rosetteless protein, required for rosette formation and speculated to play a structural role in holding rosettes together, localizes to the same location on the basal pole of cells and becomes similarly enriched as rosette form. Mutants lack the glycosylated basal patch of jacalin staining. The altered cell surface could lead to clumping, either through mis-regulation of cell adhesion molecules or exposure of a normally masked adhesive cell surface. The same alteration that allows clumping of Class C mutants also prevents rosette develoment, perhaps by disrupting glycan modification on the Rosetteless protein or one of its interaction partners.

Figure 1

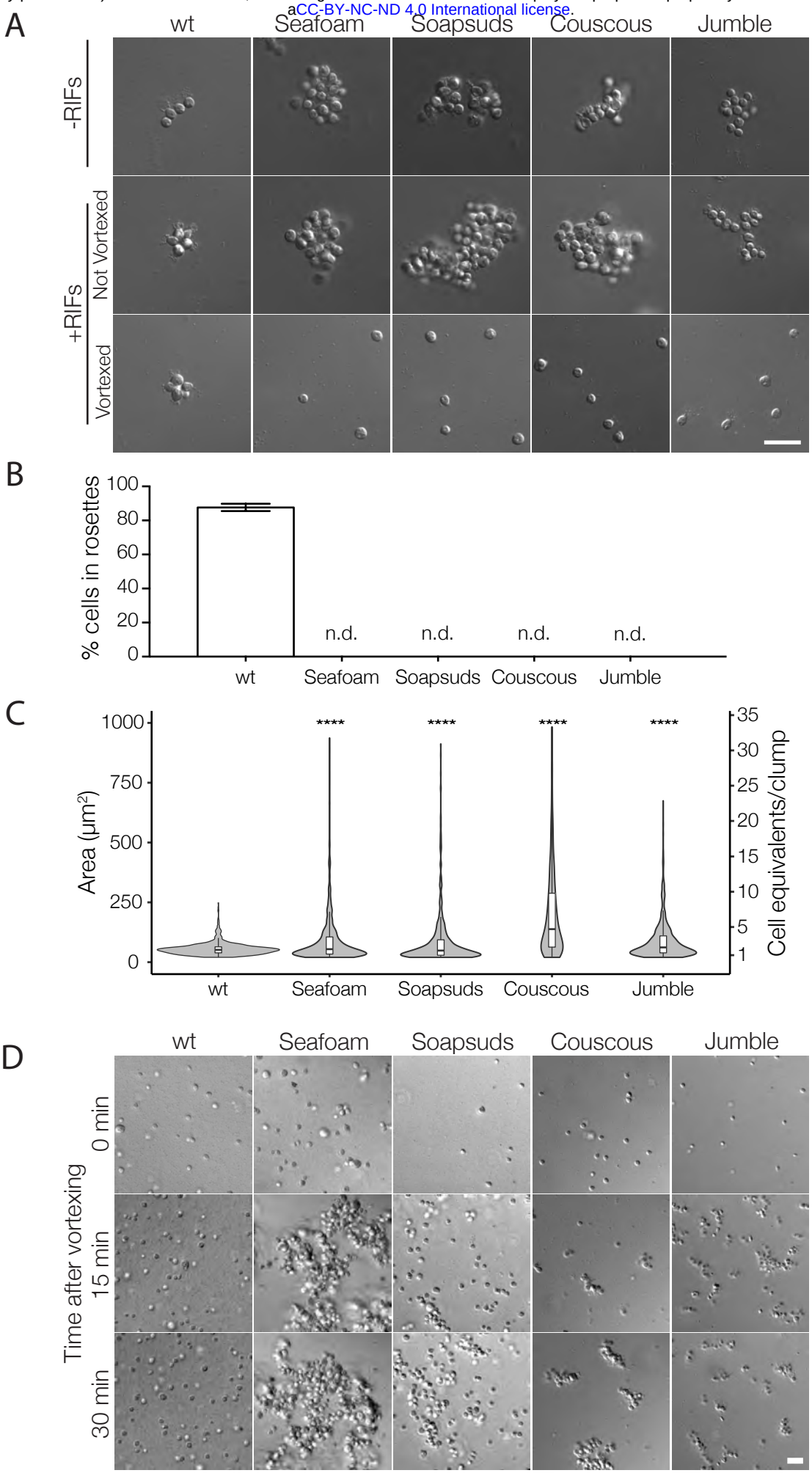
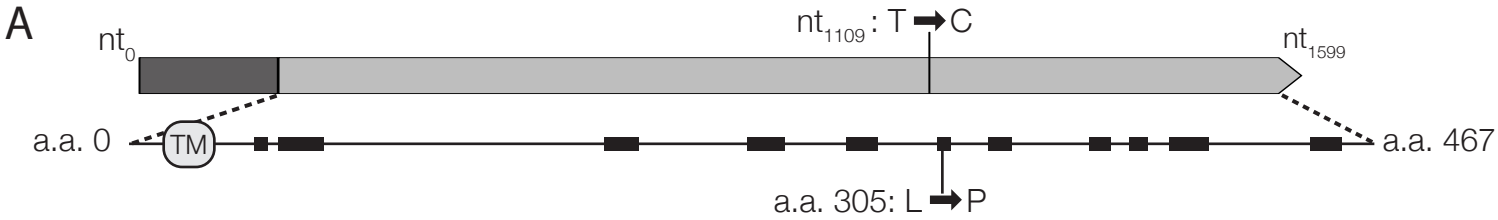


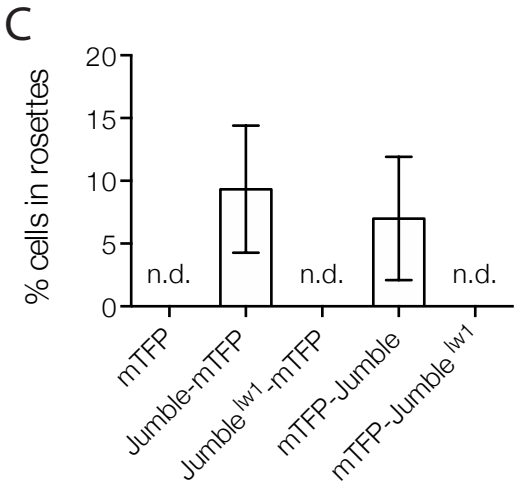
Figure 2

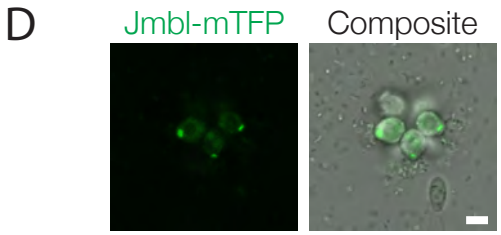


B

Jumble		
	Rosettes	Clumps
T	9	0
C	0	12

X^2=21.9 , df=3, $p < 0.001$





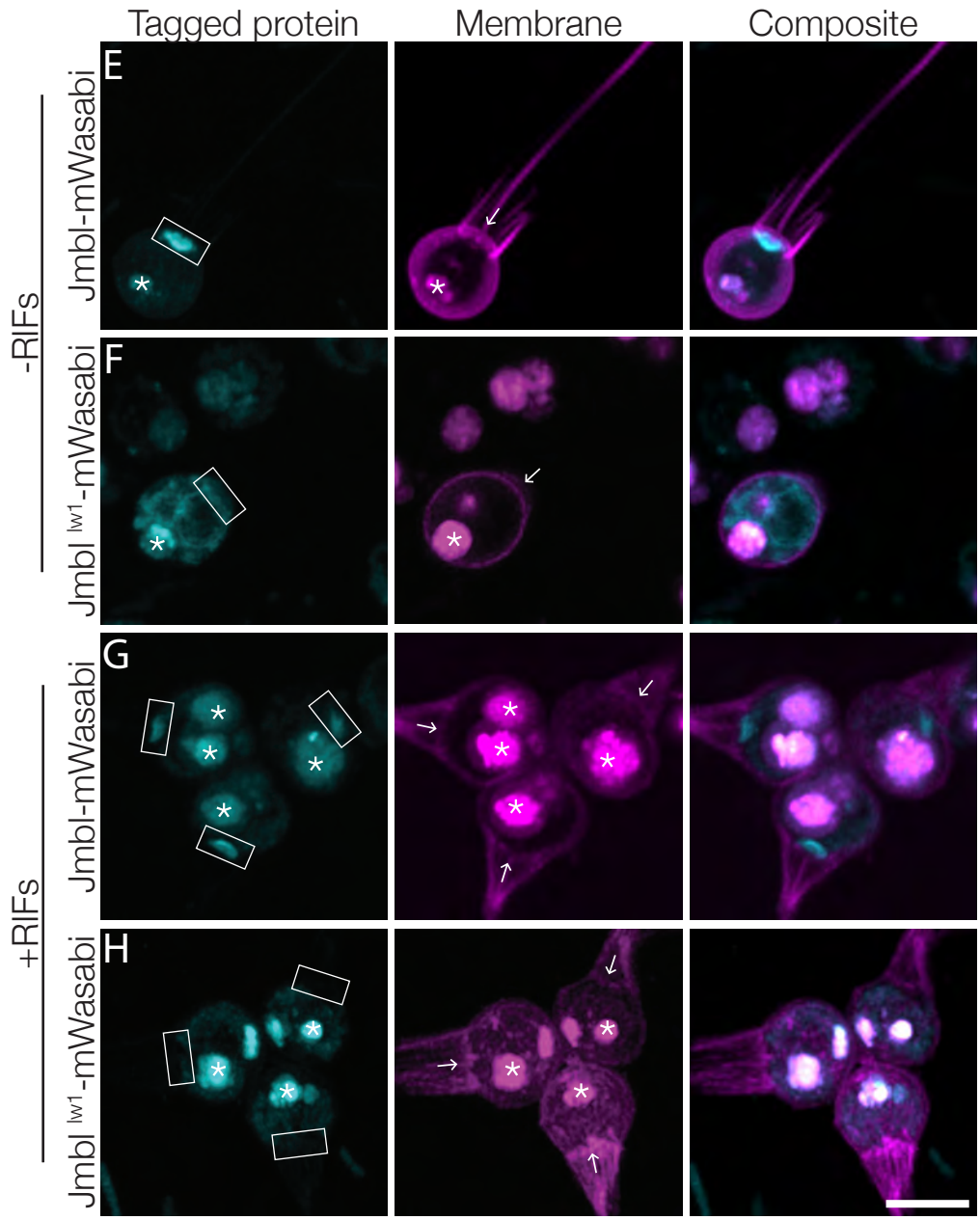
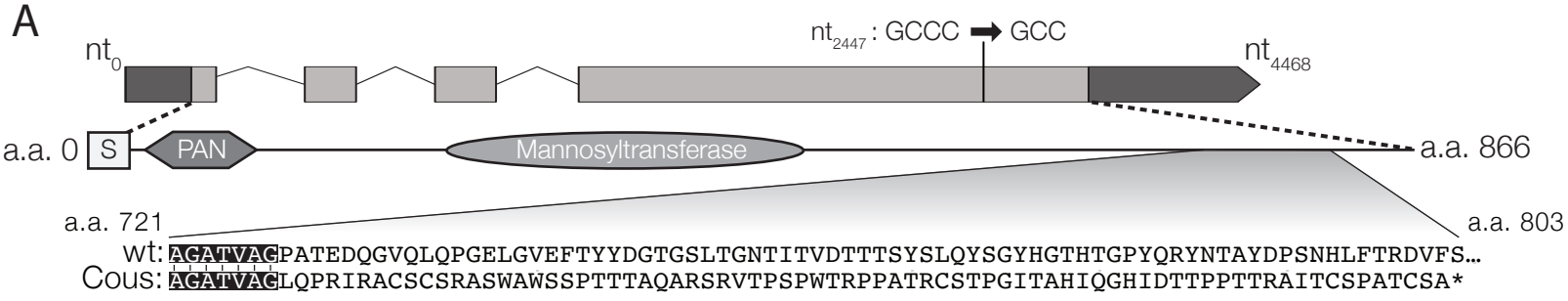


Figure 3

A

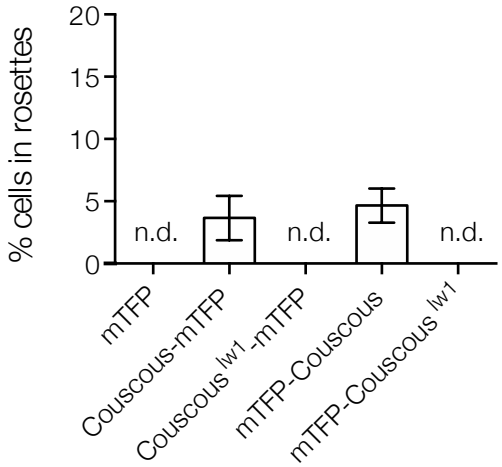


B

Couscous		
	Rosettes	Clumps
GCCC	38	0
GCC	0	51

X^2=92.8 , df=3, $p < 0.001$

C



D

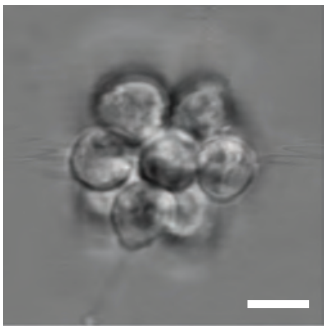
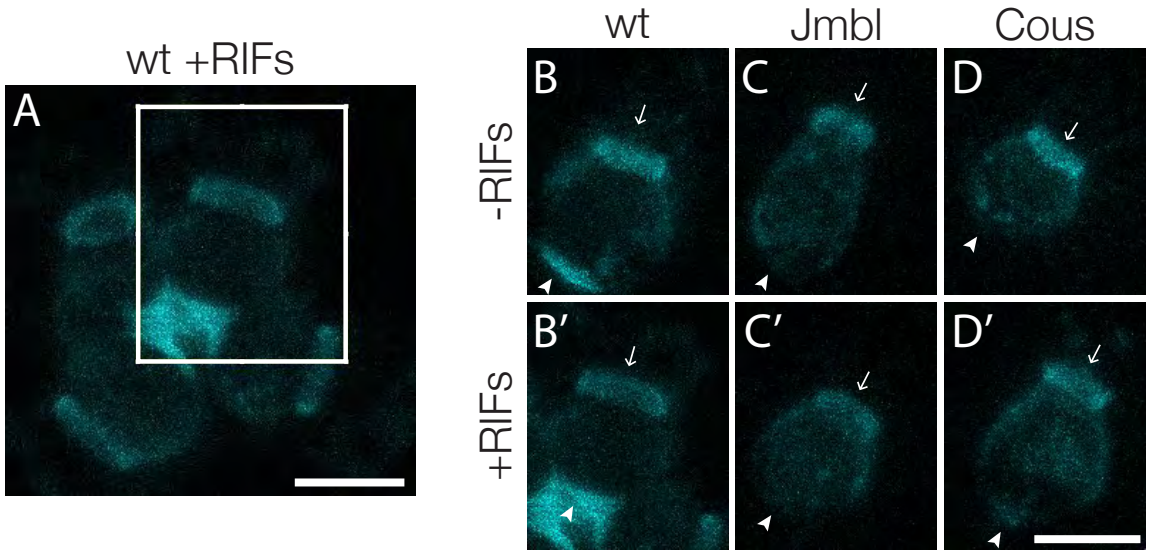


Figure 4



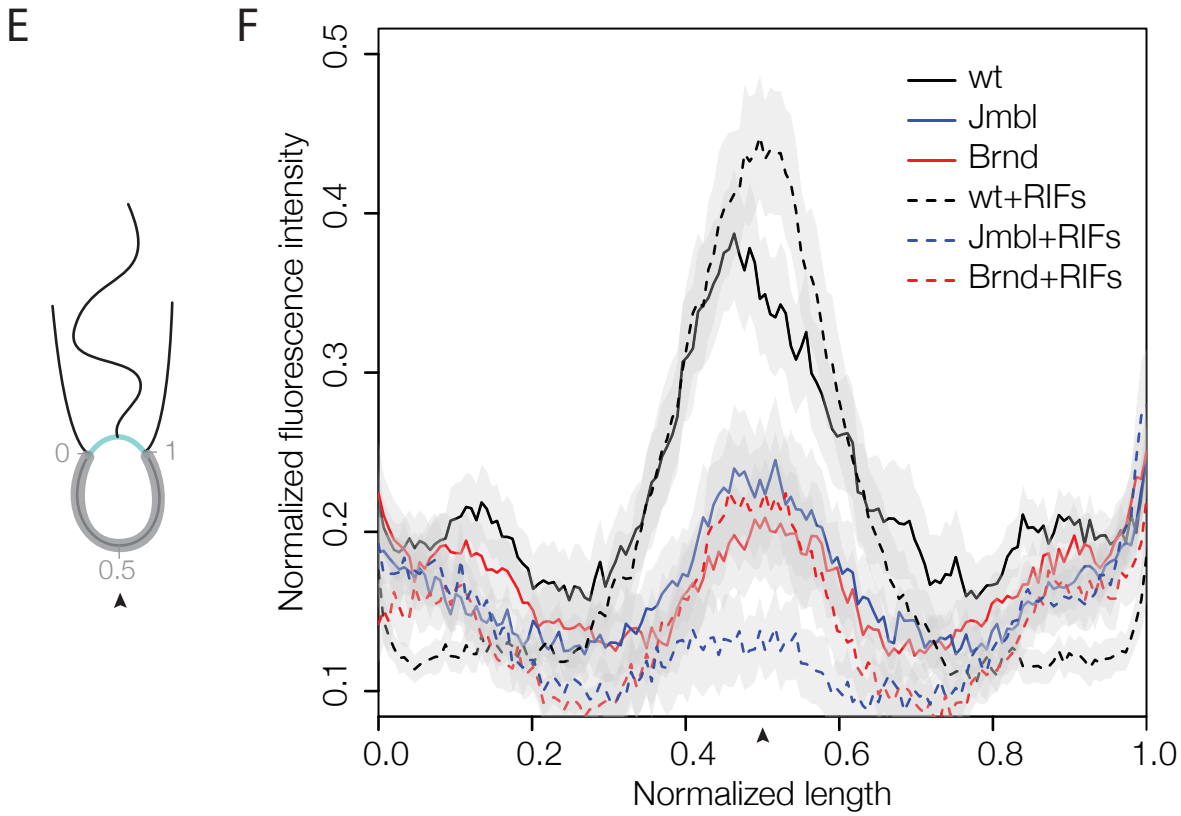


Figure 5

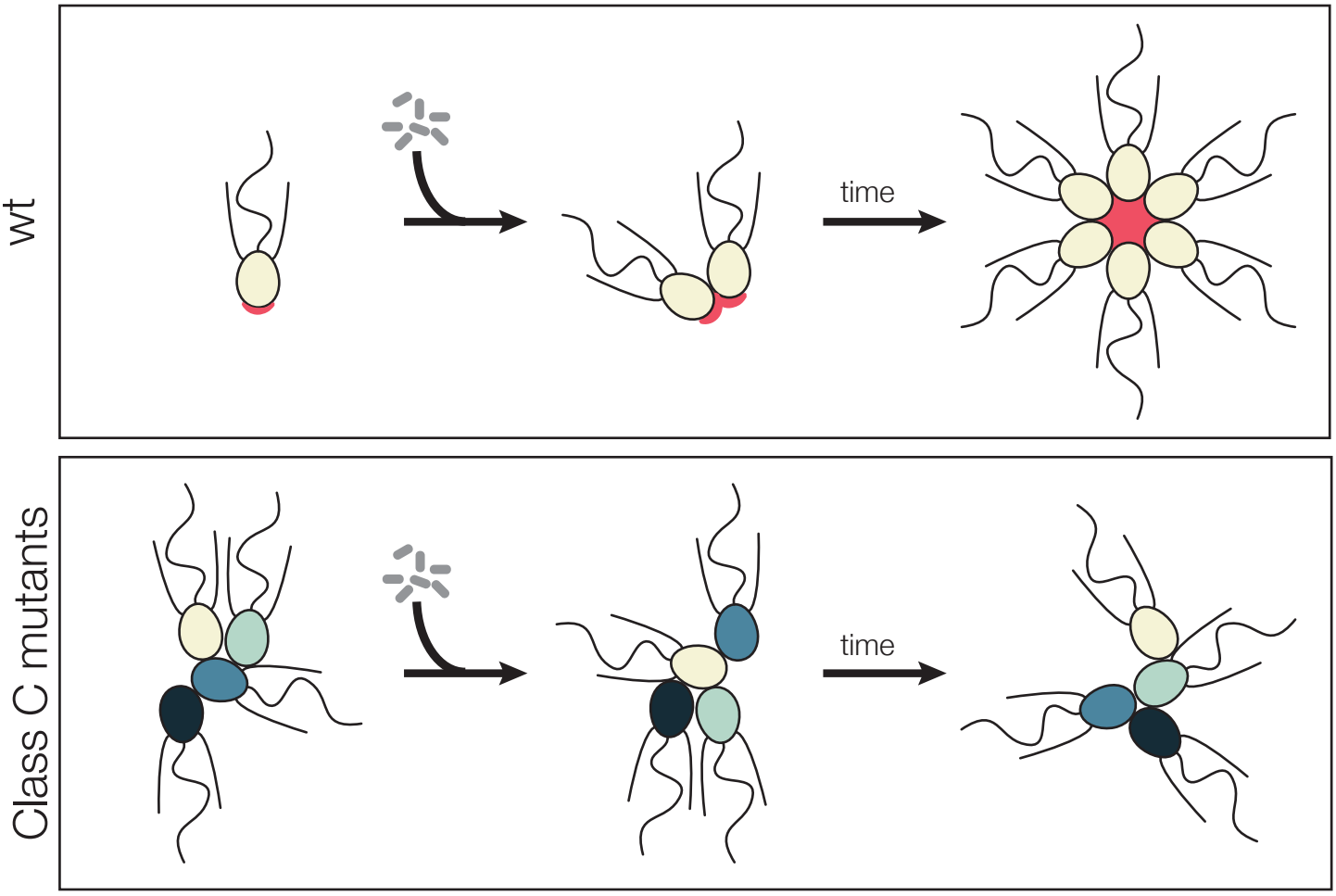


Table 1. Phenotypes of wild type and Class C mutants

Strain	% cells in rosettes	Cell interactions	Successful outcross?
wild type	87.7	Non-clumping	Yes
Seafoam	0	Clumping	No
Soapsuds	0	Clumping	No
Couscous	0	Clumping	Yes
Jumble	0	Clumping	Yes