Glycosyltransferase homologs prevent promiscuous cell aggregation and promote multicellular development 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

ABSTRACT

The mechanisms underlying multicellular development in the earliest animals may be revealed through the study of the closest living relatives of animals, choanoflagellates. The emerging model choanoflagellate *S. rosetta* can develop from a single cell into a multicellular "rosette" through a process of serial cell divisions. Through a screen for rosette defect mutants, we have uncovered multiple *S. rosetta* mutants in which single cells fail to develop into orderly rosettes but instead aggregate promiscuously into amorphous clumps of cells. By mapping the genetic lesions underlying two of these clumping/rosette defect mutants, Jumble and Couscous, we found that both are monogenic and caused by mutations in genes encoding glycosyltransferases, enzymes that transfer activated sugars to donor molecules. These are only the second and third genes to be implicated in the regulation of multicellularity in choanoflagellates to date. Animal glycosyltransferases contribute to the production of the polysaccharide-rich glycocalyx that covers nearly all animal cells, regulate the activity of integrins and cadherins (proteins critical for development and cell-cell adhesion), and, when disrupted, can contribute to tumorigenesis. Our finding that glycosyltransferases in *S. rosetta* are required to prevent spurious cell adhesion in single cells and to promote proper cell adhesion during rosette development 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 allowed for the diversification of complex macroscopic life (Leigh et al., 1995). Multicellularity has evolved independently in at least 16 different eukaryotic lineages (Bonner, 1998; King, 2004; Rokas, 2008), either through serial cell division or by aggregation of single cells. Both processes require cell-cell signaling and cell-cell adhesion, often mediated by interactions with a shared extracellular matrix (ECM). Perhaps because clonal multicellularity avoids the genetic conflicts associated with cell aggregation, most examples of “complex multicellularity” are found in lineages with clonal multicellularity (Knoll, 2011). In the case of animals, the complex multicellularity of the adult develops clonally as a single-celled zygote undergoes serial rounds of cell division followed by cell differentiation and tissue morphogenesis. Despite the centrality of multicellularity to the origin and diversification of animals, little is known about the genetic and developmental events that precipitated the transition to multicellularity in the animal stem lineage.

As the closest living relatives of animals (King et al., 2008; Ruiz-Trillo et al., 2008; Schalchian-Tabrizi et al., 2008), choanoflagellates can help to reveal the origins of animal multicellularity. *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). However, when exposed to rosette inducing factors (RIFs) produced by the prey bacterium *Algoriphagus machipongonensis, S. rosetta* instead develops into highly

63 organized and structurally stable rosettes through a process of serial cell division
64 (Alegado et al., 2012; Dayel et al., 2011; Fairclough et al., 2010; Woznica et al., 2016).
65 Therefore, the transition to rosette development in *S. rosetta* can be induced in culture.
66 Additionally, *S. rosetta* has a sequenced genome (Fairclough et al., 2010), a sexual
67 phase to its life cycle that enables controlled mating (Levin et al., 2014; Levin and King,
68 2013; Woznica et al., 2017), and newly developed techniques that allow for transfection
69 and expression of transgenes (Booth et al., 2018).

70 The growing experimental tractability of *S. rosetta* enabled a pioneering genetic
71 screen, in which mutagenized cells were assessed for the ability to form rosettes in the
72 presence of *Algoriphagus* RIFs (Levin et al., 2014). Multiple rosette defect mutants were
73 recovered and displayed a range of phenotypes. The first mutant characterized in detail,
74 Rosetteless, did not form rosettes in the presence of RIFs, but was otherwise
75 indistinguishable from wild type cells (Levin et al., 2014). The mutation underlying the
76 Rosetteless phenotype was mapped to a C-type lectin, encoded by the gene
77 *rosetteless,* the first gene known to be required for rosette formation (Levin et al., 2014).
78 Until the sequencing of choanoflagellate genomes and transcriptomes (Fairclough et al.,
79 2013; King et al., 2008; Richter et al., 2018), C-type lectins were thought to be restricted
80 to animals, where they act in signaling and adhesion to promote development and
81 innate immunity (Levin et al., 2014). In *S. rosetta,* Rosetteless protein localizes near the
82 membrane at the basal pole of each cell prior to rosette induction and is then secreted
83 basally as part of an extracellular matrix that coats and connects the basal pole of each
84 cell in rosettes (Levin et al., 2014). The localization of Rosetteless protein and its

requirement for rosette development led to the hypothesis that the protein may be essential for the structural integrity of rosettes (Levin et al., 2014).

Here we report on a class of mutants from the original rosette-defect screen that display a distinct phenotype; mutant cells in this class fail to form organized rosettes, but instead form large, amorphous clumps of cells in both the absence and presence of RIFs. We hereafter refer to this as a clumpy/rosette defect phenotype, as the two defects appear linked. By mapping the mutations underlying the clumpy/rosette defect phenotype using a newly developed bulk sequencing approach, we identified two predicted glycosyltransferase genes that are each essential for proper rosette development. 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

Four rosette defect mutants form amorphous clumps of cells through promiscuous cell adhesion

The original rosette defect screen performed by Levin *et al.,* 2014 (Levin et al., 2014) yielded nine mutants that were sorted into seven provisional phenotypic classes. For this study, we continued the screen and identified an additional eight mutants that failed to form proper rosettes in the presence of *Algoriphagus* RIFs. Comparing the phenotypes of the 17 total rosette defect mutants in the presence and absence of RIFs allowed us to identify four broad phenotypic classes: (1) Class A mutants that produce

normal chains in the absence of RIFs and entirely lack rosettes in the presence of RIFs, (2) Class B mutants that produce normal chains 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 unusual clumps in the absence of RIFs and little to no rosettes in the presence of RIFs, and (4) Class D mutants that exist primarily as solitary swimmers, with no chains detected in the absence of RIFs and no rosettes detected in the presence of RIFs (Table S1).

Of the 17 rosette defect mutants isolated, eight mutants fell into Class C. For this study, we focused on four Class C mutants — Seafoam, Soapsuds, Jumble, and Couscous [previously named Branched (Levin et al., 2014)] — that form amorphous, tightly packed clumps of cells, both in the presence and absence of RIFs (Table 1; Figure 1A). 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 agitation of the culture flask (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 adhesion was not strain-specific, as unlabeled Jumble and Couscous mutant cells adhered to wild type cells identified by their expression of cytoplasmic mWasabi (Figure S1). Therefore, the cell clumps are not aberrant rosettes, which never form through aggregation and instead form clonally through serial rounds of cell division (Dayel et al., 2011; Fairclough et al., 2010). The fact that eight 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 S2).

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. The single prior mapping cross performed in *S. rosetta* revealed a lesion in a C-type lectin gene, *rosetteless*, that prevented proper rosette development (Levin et al., 2014). In the Levin *et al.* 2014 study, the Rosetteless mutant was crossed to a phenotypically wild type Mapping Strain [previously called Isolate B (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 up clonal cultures of haploid F1 offspring; (5) phenotyping the F1 offspring; (6) pooling F1 offspring based on their clumping phenotype; and (7) sequencing pooled genomic DNA from the F1 mutants to find mutations that segregated perfectly with the clumping phenotype (Figure S3).

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 (Table 2). 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. Four of these detected sequence variants were likely artefacts resulting from relatively low sequencing coverage that caused assembly errors, which are common in repetitive sections of the genome (Levin

et al., 2014) (Table 2). 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 (Levin et al., 2014) (Table 2). Thus, a method based on pooling of F1 haploid mutants, identification of sequence variants that perfectly segregated with the phenotype, and masking of those SNVs/INDELs that were detected with coverage far below that of the total genome was effective for correctly pinpointing the causal mutation for Rosetteless (Figure S3). 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 could not be assessed for either rosette defect or clumping phenotypes. 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 defect. Only one of these – at position 1,919,681 on supercontig 1 – had sequencing coverage of at least 0.5X of the average sequence coverage of the rest of the genome (Figure 2A; Table 3). In a backcross of mutant F1 progeny to the Mapping Strain, we confirmed the tight linkage of the SNV to the rosetteless 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 transversion 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). Although stable integration of transgenes has not yet been established in *S. rosetta*, 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 sequence encoding a self-cleaving peptide (Wang et al., 2015). After 48 hours of puromycin selection and the addition of RIFs, we detected 9.33%±5.07% and 7.00%±4.91% of Jumble cells in rosettes with expression of either *jumble-mTFP* or *mTFP-jumble*, respectively (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 only the wild type *jumble* allele, albeit in a subset of the population, provides further confirmation that the mapping to GenBank accession EGD72416 was correct. The fact that the transfection experiment did not to 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. While BLAST searches did not uncover any clear *jumble* homologs in non-choanoflagellates, *jumble* homologs with clear sequence identity were detected in the transcriptomes of at least 9 other species of choanoflagellates representing each of the three major choanoflagellate clades (Richter et al., 2018) (Figure S4). Its broad

distribution among choanoflagellates suggests that *jumble* evolved before the origin and diversification of choanoflagellates. Interpro (Finn et al., 2017) and Pfam (Finn et al., 2016) did not reveal any known protein domains in Jumble. However, the NCBI Conserved Domain Search (Marchler-Bauer et al., 2017) predicted a glycosyltransferase domain with low confidence (E-value 3.87^{-03}). Moreover, two structural homology predictors, HHphred (Zimmermann et al., 2017) and Phyre (Kelly et al., 2015), revealed a constellation of structural domains (e.g. alpha helices and beta sheets; Figure S4) in Jumble that resemble those in well-annotated glycosyltransferases. The Leu305Pro substitution in Jumble^{lw1} disrupts one of those domains, 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). 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 currently a fluorescent marker in *S. rosetta* (Figure 2D,F; Figure S5) (Leadbeater, 2015). Notably, many well-characterized glycosyltransferases act in the Golgi apparatus, where they glycosylate molecules as they are trafficked through the secretory system (El-Battari, 2006; Tu and Banfield, 2010). In contrast, Jumble^{lw1}-mWasabi, was distributed in a tubular pattern throughout the cell and co-localized with an endoplasmic reticulum (ER) marker (Figure 2E,G; Figure S5B) (Booth et al., 2018). The failure of the Jumble^{lw1} protein to localize properly at the Golgi apparatus may cause a loss of function and help to explain the clumping and rosette defect phenotypes.

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 0.5X of the average sequence coverage of the rest of the genome (Figure 3A; Table 4). 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 S6A). In these fungi, MNN2 catalyzes 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, 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).

In wild type cells transfected with a *couscous-mWasabi* transgene under the *efl* promoter, Couscous was found in puncta scattered about the cytosol, collar and cell membrane (Figure S6B, C). While Couscous-mWasabi was clearly not localized to the Golgi, the puncta may co-localize with the ER, where glycosyltransferases are also known to function (El-Battari, 2006; Tu and Banfield, 2010). However, despite attempting to co-transfect cells with *couscous-mWasabi* and a marker of the ER, we were unable to detect any cells expressing both constructs. In addition, it is possible that the fusion of Couscous to a fluorescent protein or its overexpression interfered with its proper localization in *S. rosetta.* Therefore, we are currently uncertain about the subcellular localization of Couscous protein.

Jumble and Couscous mutants lack proper sugar modifications at the basal pole

Because both Jumble and Couscous have mutations in putative glycosyltransferases, we hypothesized that the abundance or distribution of cell surface sugars, called glycans, on Jumble and Couscous mutant cells might be altered. To investigate the distribution of cell surface glycans, we stained live *S. rosetta* with diverse fluorescently labelled sugar-binding lectins. Of the 22 lectins tested, 21 either did not recognize *S. rosetta* or had the same staining pattern in wild type, Jumble and Couscous cells (Table S2).

The remaining lectin, jacalin, bound to the apical and basal poles of wild type cells (Figure 4B, B'). In rosettes, jacalin also brightly stained the interior ECM in a

pattern reminiscent of Rosetteless (Levin et al., 2014) (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, in Couscous and Jumble mutants the basal patch of jacalin staining was absent or significantly diminished, both in the presence and absence of RIFs (Figure 4 C, D). Interestingly, the apical patch of jacalin binding 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 mutants (Figure S7).

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 S8). 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 17 rosette defect mutants isolated in Levin *et al.* 2014 and in this study, almost half (8) also display a mild to severe clumping phenotype. This suggests that mechanisms for preventing promiscuous adhesion among wild type cells are diverse and 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. Importantly, the tight coupling of the rosette defect and clumping phenotypes survived recombination in both the Jumble×Mapping Strain and Couscous×Mapping Strain crosses, indicating that the mutation that causes promiscuous adhesion in each of these mutants also explains the loss of rosette development.

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. In apicomplexan and oomycete parasites, PAN/Apple domains are found in secreted adhesive proteins that mediate parasite attachment and invasion of host cells (Brown et al., 2001). Therefore, Couscous might play an additional role in cell-cell adhesion or signal transduction due to the presence of the PAN/Apple domain.

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

383 receptors. Glycosylation regulates the activities of two key adhesion proteins in animals:
384 integrins that regulate ECM adhesion, and cadherins that, among their various roles in
385 cell signaling and animal development, bind other cadherins to form cell-cell adhesions
386 called adherens junctions [22,26]. Cadherin activity can be either positively or negatively
387 regulated by glycosyltransferases. For example, epithelial cadherin (E-cadherin) is
388 modified by N-acetylglucosaminyltransferase III (GnT-III) whose activity leads to
389 increased cell adhesion and N-acetylglucosaminyltransferase V (GnT-V) whose activity
390 leads to decreased cell adhesion (Carvalho et al., 2016; Granovsky et al., 2000). GnT-V
391 knockdown enhances cell-cell adhesion mediated by E-cadherin and the related N-
392 cadherin (Carvalho et al., 2016; Guo et al., 2009). The inactivation of E-cadherin,
393 including through over- or under- expression of GnT-V or GnT-III, is considered to be a
394 hallmark of epithelial cancers (Hirohashi and Kanai, 2003). While we do not yet know
395 the targets of Jumble and Couscous activity, *S. rosetta* expresses 29 different cadherins
396 (Nichols et al., 2012). Therefore, mutations to *jumble* and *couscous* might disrupt
397 regulatory glycosylation of a cell adhesion molecule, like cadherin, and thereby promote
398 adhesion.

399 Another possibility is that *jumble* and *couscous* add a protective sugar layer to
400 the cell surface and loss of glycosyltransferase activity reveals underlying sticky
401 surfaces. If *jumble* and *couscous* add branches to existed sugar modifications, their loss
402 of function could expose new sugar moieties at the cell surface that might act as ligands
403 for lectins that aggregate cells. Lectins mediate cell aggregation in diverse organisms.
404 For example, sponges such as *Geodia cydonium* can be disaggregated into single cells
405 and then reaggregated through lectin binding of a post-translational sugar modification

406 (Müller et al., 1979). In *S. cerevisiae*, the mannosyltransferase MNN2 adds mannose
407 structures to the cell wall that are recognized by aggregating lectins and MNN2 is
408 required for proper flocculation (Rayner and Munro, 1998; Stratford, 1992). Exposing
409 new sugars on the cell surface in Jumble and Couscous could lead to spurious
410 aggregation, potentially by lectins or other sugar binding proteins.

411 It is somewhat more difficult to infer how increased clumping in single cells might
412 interfere with rosette development. One possibility is that the disruption of ECM
413 glycosylation that we hypothesize might promote clumping may also prevent the proper
414 maturation of the ECM needed for rosette development (Figure 5). A prior study showed
415 that *S. rosetta* cells lacking a target of the lectin wheat germ agglutinin (WGA) are not
416 competent to form rosettes, highlighting the potential importance of glycosylated
417 molecules for rosette development (Dayel et al., 2011). While WGA staining does not
418 appear to be perturbed in Jumble and Couscous (Table S2), jacalin staining at the basal
419 pole appears severely reduced or abolished compared to wild type. Jacalin staining was
420 enriched in the center of wild type rosettes in a pattern reminiscent of Rosetteless,
421 which is required for rosette development [3]. Intriguingly, in Jumble and Couscous,
422 Rosetteless localized to the correct pole, but did not become enriched upon induction
423 indicating that the ECM did not properly mature. Rosetteless has mucin-like Ser/Thr
424 repeats that are predicted sites of heavy glycosylation and two C-type lectin domains
425 that would be expected to bind to sugar moieties (Levin et al., 2014). Therefore, it is
426 possible that Rosetteless might be regulated either through direct glycosylation or
427 through the glycosylation of potential binding partners by Jumble and Couscous.

The clumping, rosetteless mutants underscore the difference between cell aggregation and a regulated clonal developmental program, such as embryogenesis or rosette development. Several studies assert that aggregation can only be evolutionarily stable if limited to close relatives (Brunet and King, 2017; Gilbert et al., 2007; Kuzdzal-Fick et al., 2011) and recent studies experimentally evolving multicellularity in yeast and the green alga *Chlamydomonas reinhardtii* resulted in isolates that formed clonally (Ratcliff et al., 2013, 2012). Together this work supports the idea that, despite both aggregation and clonal multicellularity relying on cell-cell adhesion, aggregative multicellularity likely was not an initial step in the evolution of clonal multicellularity. The high co-occurrence of cell aggregation and the loss of rosette development show that aggregative multicellularity might block clonal multicellularity in *S. rosetta* and that the two forms of multicellularity cannot easily be interconverted.

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). Our wild type strain, from which each mutant was generated, was the described strain SrEpac (ATCC PRA-390; accession number SRX365844) –*S. rosetta* 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 first 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, used for mapping crosses (accession number SRX363839) contains *S. rosetta* grown in the presence of *A. machipongonensis* bacteria. 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 either live *A. machipongonensis* (ATCC BAA-2233) (Alegado et al., 2012) or *A. machipongonensis* outer membrane vesicles (OMVs) prepared as in (Woznica et al., 2016); both are referred to as rosette inducing factors (RIFs).

Imaging and quantify 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 1:500 *Algoriphagus* OMVs. Cultures were imaged after 48 hr induction in 8-well glass bottom dishes (ibidi 15 µ-Slide 8 well 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 both uninduced and induced images, 200 µl of cells were plated with a wide bore pipette tip for minimal disruption and allowed to settle for 5 min. For vortexed images, 200 µl of cells were vortexed for 15 second before plating and imaged within 10 mins 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, cells were plated at a density of $1x10^4$ cells/ml in 3 ml HN media with 1:500 *Algoriphagus* OMVs. After 48 hr, an aliquot of cells was vortexed vigorously for 15 secs and fixed with formaldehyde. Rosette formation counted by hemacytometer and rosettes assessed as clusters of 3 or more cells.

Imaging and quantification of clumping

Clumps were quantified using a modified protocol from (Woznica et al., 2017). To prevent spurious sticking, 6 well glass bottom dishes were coated with 1% BSA for 1 hr and washed 3 times with water to remove any remaining BSA. Cells were diluted to $5x10^5$ cells/mL, vortexed for 15 secs to break apart any pre-formed clumps and plated in the BSA pre-treated dishes. For quantification, DIC images were taken using Zeiss Axio Observer.Z1/7 Widefield microscope with a Hammatsu Orca-Flash 4.0 LT CMOS Digital Camera and a 20x objective. Images were collected for each strain from 10 distinct locations throughout the well.

Images were batched processed in ImageJ for consistency. To accurately segment the phase bright cells and limit signal from the phase dark bacteria the following commands were applied: ‘Smooth’, ‘Find Edges’, ‘Despeckle’, ‘Make Binary’, ‘Dilate’, ‘Erode’ and ‘Fill Holes’. Finally, images were analyzed with the ‘Analyze Particles’ command to calculate the area of the clump and only particles larger than 20 μm^2 were kept. Cells equivalents/clump (right y axis) were calculated by dividing the area of the clump by the area of an individual cells (as determined by the average area of the wt cells). Data are presented as violin boxplots, showing the median cell number (middle line), interquartile range (white box), and range excluding outliers (thin line). A

minimum of 630 clumps from two biological replicates were measured for each condition.

Performing mapping crosses

Crosses for each mutant strain (Seafoam, Soapsuds, Jumble, and Couscous) with Mapping Strain (previously described as Isolate B) were attempted through both nutrient limitation for 11 days and addition of 5% *Vibrio fischeri* spent media (Levin and King, 2013; Woznica et al., 2017). Mating was induced, and cells were plated by limiting dilution to isolate diploid clones. Thecate isolates, as the only documented diploid cell type (Levin et al., 2014; Woznica et al., 2017), were genotyped. From each thecate isolate population, we extracted DNA from 75 µl of cells by pelleting cells, resuspending in 10 µl of base solution (25 mM NaOH, 2 mM EDTA), transferring samples into PCR plates, boiling at 100°C for 20 min, followed by cooling at 4°C for 5 min, and then adding 10 µl Tris solution (40 mM Tris-HCl, pH 7.5). We used 2 µl of this sample as the DNA template for each genotyping reaction. We identified heterozygous clones by genotyping PCR at a single microsatellite genotyping marker at position 577,135 on supercontig 1 (Forward primer: GACAGGGCAAAACAGACAGA and Reverse primer: CCATCCACGTTCATTCTCCT) that distinguishes a 25 bp deletion in the Mapping Strain (199 bp) from mutants (217 bp). Isolates containing both markers were saved, and meiosis induced by rapid passaging every day in CG medium. For both Seafoam and Soapsuds, we were able to generate putative outcrossed diploids to the Mapping Strain based on the single genotyping marker, but we only ever able to recover rosette F1 haploids and never isolated any F1 haploids with the clumpy, rosetteless phenotype.

518 For the successful Jumble cross, we took fast growing, regularly passaged
519 strains and set up starvation conditions for the cross. We pelleted $2x10^6$ cells/mL of
520 each strain together and resuspended in 10mL of ASW. After 11 days of starvation in
521 ASW, we pelleted all cells and resuspended in 100% CG media. After 3 days of
522 recovery, we isolated clones by limiting dilution in 10% CG media in ASW (vol/vol). The
523 probability of clonal isolation in this step was 0.91-0.93. Three heterozygous isolates,
524 each in the thecate life history, were identified and carried forward. To induce meiosis,
525 isolates were passaged 1:2 in 25% CG media in ASW (vol/vol) every 1-2 days for 8
526 days. As soon as rosettes and swimming cells were identified, we repeated clonal
527 isolation (probability of clonal isolation 0.85-0.98). We collected any isolates that formed
528 purely rosettes or chains and ignored any thecate wells assuming that these
529 represented diploid cells. 56% of all swimming isolates were clumpy chains, consistent
530 with Mendelian segregation of a single locus. Isolates were genotyped with the marker
531 on supercontig 1 to ensure that independent assortment of the genotype and the
532 phenotype indeed occurred. In total, 30 clumpy F1s were grown up for bulk segregation
533 analysis.
534 For the successful Couscous cross, a mixture of $1x10^6$ Couscous and the
535 Mapping Strain cells at stationary phase were induced to mate in 5% *Vibrio fischeri*
536 spent media in ASW (vol/vol) (Woznica et al., 2017). After 24 hr, the cells were pelleted
537 and resuspended in 5% HN media in ASW (vol/vol) and allowed to recover for 24 hr.
538 Then we isolated clones by limiting dilution in 10% CG media in ASW (vol/vol). The
539 probability of clonal isolation in this step was between 0.97-0.98. We extracted DNA as
540 described above and identified heterozygous clones by genotyping PCR at a single

microsatellite genotyping marker on supercontig 1. Four heterozygous isolates, each in the thecate stage, were identified and carried forward. To induce meiosis, isolates 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 identified, we repeated clonal isolation (probability of clonal isolation 0.78-0.97). We collected any isolates that formed purely rosettes or chains and ignored any thecate wells assuming that these represented diploid cells. Only 14.6% of swimming isolates were clumpy chains; this deviation from a Mendelian ratio may be indicative a potential fitness defect of the mutant phenotype. Isolates were genotyped with the marker on supercontig 1 to ensure that independent assortment indeed occurred. In total, 22 haploids were collected for bulk segregant analysis.

Whole genome re-sequencing

Jumble and Couscous were re-sequenced individually in order to identify mutation carried in each strain. To do this, Jumble and Couscous cells were grown to stationary phase in 500 mL of 5% HN media in ASW (vol/vol). To generate bulk segregant samples, we grew up $5x10^6$ Rosetteless×Mapping Strain (Levin et al., 2014) and Jumble×Mapping Strain and $1x10^7$ Couscous×Mapping Strain mutant F1s cells of each isolate. For all samples, we extracted DNA by phenol chloroform extraction and used a CsCl gradient to separate *S. rosetta* DNA from contaminating *E. pacifica* DNA by GC content (King et al., 2008).

Multiplexed, 150 bp paired-end libraries were prepared and sequenced on an Illumina HiSeq 4000. Raw reads were trimmed with TrimmomaticPE (Bolger et al., 2014) to remove low quality base calls. Trimmed reads were mapped to the *S. rosetta* reference genome (Fairclough et al., 2013) using Burrows-Wheeler Aligner (Li and

564 Durbin, 2009), and we removed PCR duplicates with Picard
565 (http://broadinstitute.github.io/picard/). We realigned reads surrounding indel calls using
566 GATK (Depristo et al., 2011) and called variants using SAMtools and bcftools (Li et al.,
567 2009).

568 **Bulk segregant sequencing analysis**

569 No large region of the genome (i.e. haplotype block) was found to co-segregate
570 with the mutant phenotype in any of the crosses, likely because of the sparse, uneven
571 distribution of genetic markers. To find segregating variants in the pooled samples,
572 variant call files were intersected with vcftools vcf-isec (Danecek et al., 2011) by: (1)
573 finding any variants shared with the mutant strain in the cross, (2) removing any variants
574 shared with the Mapping Strain (Isolate B), wild type (previously Isolate C), and the
575 unmutagenized control from the Rosetteless mutagenesis (C2E5) (Levin et al., 2014;
576 Levin and King, 2013). The remaining variants were filtered by quality: depth >2, quality
577 score >10, and reference allele not N. The remaining list represents high quality variants
578 in the pooled population that are shared with the mutant to the exclusion 3 different
579 strains competent to form rosettes. Segregating variants were determined by dividing
580 the number of reads that map to the alternative allele by the total number of high quality
581 reads determined by SAMtools and bcftools (Li et al., 2009); any variants with >99% of
582 reads that map to the alternative allele were considered variants that segregated
583 perfectly with the mutant phenotype.

584 **Backcrosses**

585 For Jumble, a cross was set up with $1x10^6$ mutant F1 and the Mapping Strain.
586 Cells were mixed, pelleted, resuspending in 10 mL of 5% *Vibrio fischeri* spent media in

587 ASW (vol/vol). After 24 hr, media was replaced with 25% CG media in ASW (vol/vol)
588 and cells were plated to limiting dilution. Thecate isolates were genotyped by
589 genotyping marker on supercontig 1; 4 heterozygous diploids were isolated (probability
590 of clonal isolation 0.79-0.95). Isolates were rapidly passaged for 2 weeks to induce
591 meiosis before being plated for clonal isolation (probability of clonal isolation 0.95-0.98).
592 12 clumps and 9 rosettes were isolated, DNA extracted using Base-Tris method
593 described above, and the region around the causal mutation was amplified. The
594 resultant PCR product was diagnostically digested for 4 hr with BfaI, which cleaves the
595 mutant allele but not the wt, and products were distinguished by agarose gel
596 electrophoresis.

597 For Couscous, two crosses were set up with $2.5x10^5$ mutant F1s from two
598 different isolates and the Mapping Strain. Cells were mixed, pelleted, resuspending in
599 0.5 mL of 2.5% *Vibrio fischeri* spent media in ASW (vol/vol). After 24 hr, media was
600 replaced with 25% CG media in ASW (vol/vol) and cells were plated to limiting dilution
601 (probability of clonal isolation 0.85-0.97). Thecate isolates were genotyped with the
602 same genotyping marker on supercontig 1; 3 heterozygous diploids were isolated from
603 each cross. Isolates were rapidly passaged for 2 weeks to induce meiosis before being
604 plated for clonal isolation (probability of clonal isolation 0.88-0.97). 51 chains and 38
605 rosettes were isolated, DNA extracted using Base-Tris method described above, the
606 region around the causal mutation was amplified, and the resultant PCR product was
607 Sanger sequenced.

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

609 Jumble domains were predicted using Interpro (Finn et al., 2017), pfam (Finn et
610 al., 2016), and the NCBI Conserved Domain Search (Marchler-Bauer et al., 2017).
611 Structural homology was performed with Phyre2 (Kelly et al., 2015) and HHphred
612 (Zimmermann et al., 2017). The alignment from the top hit on HHphred, human GlcNAc
613 T4 catalytic domain structure was aligned to the predicted Jumble structure using the
614 PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC. Other
615 choanoflagellate homologs were determined by reciprocal BLAST of the 20 sequenced
616 choanoflagellate transcriptomes (Richter et al., 2018) and alignment were performed
617 with ClustalX (Larkin et al., 2007). Couscous domains were predicted using Interpro
618 (Finn et al., 2017) and pfam (Finn et al., 2016), and alignment to yeast MNN2
619 glycosyltransferase domains were performed with ClustalX (Larkin et al., 2007).

620 **Generating transgenic constructs**

621 Jumble (GenBank accession EGD72416/NCBI accession XM_004998928) and
622 Couscous (GenBank accession EGD77026/ NCBI accession XM_004990809) were
623 cloned from wild type cDNA prepared as described in (Booth et al., 2018). Jumble^{lw1}
624 was cloned from cDNA prepared from the Jumble mutant. Couscous^{lw1} could not be
625 cloned from cDNA directly (possibly because of low mRNA levels due to nonsense
626 mediate decay or simply because of high GC content of the gene). However, the 1 bp
627 deletion in *Couscous*^{lw1} was confirmed by Sanger sequencing of genomic Couscous
628 DNA. Site directed mutagenesis of the wild type gene was used to generate the mutant
629 allele.

630 For complementation, constructs were generated from a plasmid with a pUC19
631 backbone with a 5' *S. rosetta* elongation factor L *(efl)* promoter, monomeric teal

fluorescent protein (*mTFP)*, and the 3' UTR from actin (Addgene ID NK633) (Booth et al., 2018). A puromycin resistance gene was synthesized as a gene block and codon optimized for *S. rosetta*. The puromycin resistance gene (puro) was inserted after the *efl* promoter and separated from fluorescent reporters by self-cleaving 2A peptide from the porcine virus (P2A) (Kim et al., 2011). Copies of *jumble, jumble*^{lw1}*, couscous,* and *couscous*^{lw1}were inserted either 5' or 3' of the mTFP and separated from mTFP by a flexible linker sequence (SGGSGGS) through Gibson cloning.

For fluorescent localization, constructs were generated from a pUC19 backbone with a 5' *S. rosetta* elongation factor L (*efl)* promoter, mWasabi, and 3' UTR from actin. Copies of *jumble, jumble*^{lw1}*,* and *couscous* were inserted either 5' of the mWasabi separated by a flexible linker sequence (SGGSGGS) through Gibson cloning. Plasma membrane and ER markers from (Booth et al., 2018) were used as previously described (Addgene ID NK624 and NK644).

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 if 1:500 *Algoriphagus* OMVs 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

678 grown an additional 24 hr after transfection to allow for transgene expression, and then
679 40 µg/ml puromycin was added for selection. Selection occurred for 48 hr before rosette
680 induction was counted by hemocytometer. After vortexing for 15 sec and fixing with
681 formaldehyde, 200 cells of each transfection well were counted on a hemocytometer to
682 determine percentage of cells in rosettes. Rosettes were assessed as clusters of 3 or
683 more cells. Complementation was repeated on 2 biological replicates with 3 technical
684 transfection replicates each. Representative rosette images were taken on by confocal
685 microscopy using Zeiss Axio Observer LSM 880 a C-Apochromat 40x/NA1.20 W Korr
686 UV-Vis-IR water immersion objective.

687 **Live cell imaging**

688 Glass-bottom dishes for live cell imaging were prepared by corona-treating and
689 poly-D-lysine coating as described in (Booth et al., 2018). Transfected cells were
690 prepared for microscopy by pelleting 1-2 ml of cells and resuspend in 200 µl of 4/5 ASW
691 with 100 mM LiCl to slow flagellar beating. Cells were plated on glass-bottom dishes
692 and covered by 200 µl of 20% (w/v) Ficoll 400 dissolved in 4/5 ASW with 100 mM LiCl.
693 Confocal microscopy was performed on a Zeiss Axio Observer LSM 880 with an
694 Airyscan
695 detector and a 63x/NA1.40 Plan-Apochromatic oil immersion objective.
696 Confocal stacks were acquired in super-resolution mode using ILEX
697 Line scanning and two-fold averaging and the following settings: 35 nm x 35 nm pixel
698 size, 100 nm z-step, 0.9-1.0 µsec/pixel dwell time, 850 gain, 458 nm laser operating at
699 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 initially 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. 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, 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 1:500 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

723 collar and the cell body meet on one side of the cell around the cell to the other and the
724 fluorescence intensity was measured along the line. To compare between cells, the
725 lines drawn around the cell body were one-dimensional interpolated in R to include 150
726 points and normalized to the length of the line. The average fluorescence intensity was
727 plotted over the length of the line drawn around the cell body for Jumble, Couscous, and
728 wild type induced and uninduced with a 95% confidence interval. Measurements were
729 taken from two biological replicates with at least 59 cells in total from each condition.

730 **Wild type and mutant growth curves**

731 All cells strains were plated at a density of $1x10^4$ cells/ml in 3 ml HN media.
732 Every 12 hr an aliquot of cells was vortexed vigorously for 15 sec, fixed with
733 formaldehyde, and counted by hemacytometer. Curves were generated from the
734 average ± SD from 2 biological replicates with 3 technical replicates each.

735 **Wild type and mutant clumping assays**

736 Wild type cells expressing cytoplasmic mWasabi fused to a puromycin resistance
737 cassette under the *efl* promoter were generated and maintained in 40 µg/mL puromycin
738 to enrich for positive transformants. For clumping assays, equal numbers of mWasabi-
739 wt cells either uninduced or induced to form rosettes were mixed with either Jumble or
740 Couscous, vortexed, and plated on BSA treated 8-well glass bottom dishes. DIC and
741 fluorescent images were obtained after 30 mins using Zeiss Axio Observer.Z1/7
742 Widefield microscope with a Hammatsu Orca-Flash 4.0 LT CMOS Digital Camera and a
743 40x/NA1.40 Plan-Apochromatic lens.

744 **Rosetteless immunofluorescence staining and imaging**

745 Immunofluorescence was performed previously described (Levin et al., 2014)
746 with the modifications for better cytoskeleton preservation described in (Booth et al.,
747 2018). Two mL of dense wild type, Jumble, and Couscous cells, that were either
748 uninduced or induced with 1:500 *Algoriphagus* OMVS for 24 hr, were allowed to settle
749 on poly-L-lysine coated coverslips (BD Biosciences) for 30 min. Cells were fixed in two
750 steps: 6% acetone in cytoskeleton buffer (10 mM MES, pH 6.1; 138 KCl, 3 mM $MgCl_2$; 2
751 mM EGTA; 675 mM Sucrose) for 5 and 4% formaldehyde with diluted in cytoskeleton
752 buffer for 20 min. The coverslips were gently washed three times with cytoskeleton
753 buffer. Cells were permeabilized with permeabilization buffer [100 mM PIPES, pH 6.95;
754 2 mM EGTA; 1 mM $MgCl_2$; 1% (w/v) bovine serum albumin-fraction V; 0.3% (v/v Triton
755 X-100)] for 30 min. Cells were stained with the anti-Rosetteless genomic antibody at
756 3.125 ng/µl (1:400), E7 anti-tubulin antibody (1:1000; Developmental Studies
757 Hybridoma Bank), Alexa fluor 488 anti-mouse and Alexa fluor 647 anti-rabbit secondary
758 antibodies (1:1000 each; Molecular Probes), and 6 U/ml rhodamine phalloidin
759 (Molecular Probes) before mounting in Prolong Gold antifade reagent with DAPI
760 (Molecular Probes).
761 Images were acquired on a Zeiss LSM 880 Airyscan confocal microscope with a
762 63x objective (as described for live cell imaging) by frame scanning in the super-
763 resolution mode with the following settings: 30 nm x 30 nm pixel size; 100 nm z-step;
764 561 nm laser operating at 1.5% power with 700 master gain, and 488 nm laser
765 operating at 2.0% power with 800 master gain. Wild type rosettes were imaged with 633
766 nm laser operating at 0.3% laser power and 650 master gain to prevent overexposure of

767 Rosetteless, but all other conditions were operating at 2% laser power and 650 master
768 gain in the 633 nm channel.

769 **Jacalin western blot**

770 Whole cell lysates were made from pelleting $1x10^7$ cells at 4C at 3,000 x g and
771 resuspending in lysis buffer (20 mM Tris-HCl, pH 8.0; 150 mM KCl; 5 mM MgCl2; 250
772 mM Sucrose; 1 mM DTT; 10 mM Digitonin; 1 mg/ml Sodium Heparin; 1 mM Pefabloc
773 SC; 0.5 U/µl DNaseI; 1 U/µl SUPERaseIN). Cells were incubated in lysis buffer for 10
774 min on ice and passed through 30G needle 5x. Insoluble material was pelleted at 6,000
775 x g for 10 min at 4C. Lysate ($1x10^6$ cells/sample) was run on 4-20% TGX mini-gel (Bio-
776 Rad) for 45 min at 200 V and transferred onto 0.2 µm nitrocellulose membrane using
777 Trans-Blot Turbo Transfer System (Bio-Rad) with semi-dry settings 25V for min. The
778 blot was blocked for 30 min with Odyssey PBS Block (Li-cor). The blot was probed with
779 biotinylated jacalin (1:4,000) and E7 anti-tubulin antibody (1:10,000; Developmental
780 Studies Hybridoma Bank) diluted in block for 1 hr, and then with IRDye 800 streptavidin
781 (1:1,000; Li-cor) and IRDye 700 mouse (1:1,000; Li-cor) in PBST [PBS with %1 Tween
782 20 (v/v)]. Blot was imaged on Licor Odyssey.

783

784 **ACKNOWLEDGEMENTS**

785 Hannah Elzinga, Lily Helfrich, and Max Coyle helped with experiments and
786 reagent preparation. We thank members of the King lab for helpful discussions,
787 research support, mutant naming suggestions, and comments on the manuscript,
788 especially Kayley Hake, Ben Larson, Tess Linden, and Thibaut Brunet. This work used

789 the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by

790 NIH S10 OD018174 Instrumentation Grant.

791

792 **REFRENCES**

793 Alegado RA, Brown LW, Cao S, Dermenjian RK, Zuzow R, Fairclough SR, Clardy J,
794 King N. 2012. A bacterial sulfonolipid triggers multicellular development in the
795 closest living relatives of animals. *Elife* **2012**:1–16. doi:10.7554/eLife.00013

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

798 Bonner JT. 1998. The origins of multicellularity. *Integr Biol Issues, News, Rev* **1**:27–36.
799 doi:10.1002/(SICI)1520-6602(1998)1:1<27::AID-INBI4>3.0.CO;2-6

800 Booth DS, Middleton H, King N, Szmidt-middleton H, King N. 2018. A robust method for
801 transfection in choanoflagellates illuminates their cell biology and the ancestry of
802 animal septins. *bioRxiv*. doi:https://doi.org/10.1101/343111

803 Brown PJ, Gill AC, Nugent PG, McVey JH, Tomley FM. 2001. Domains of invasion
804 organelle proteins from apicomplexan parasites are homologous with the Apple
805 domains of blood coagulation factor XI and plasma pre-kallikrein and are members
806 of the PAN module superfamily. *FEBS Lett* **497**:31–38. doi:10.1016/S0014-
807 5793(01)02424-3

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

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

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

818 Dayel MJ, Alegado RA, Fairclough SR, Levin TC, Nichols SA, McDonald K, King N.
819 2011a. Cell differentiation and morphogenesis in the colony-forming
820 choanoflagellate Salpingoeca rosetta. *Dev Biol* **357**:73–82.
821 doi:10.1016/j.ydbio.2011.06.003

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

827 Doitsidou M, Poole RJ, Sarin S, Bigelow H, Hobert O. 2010. *C. elegans* mutant

identification with a one-step whole-genome-sequencing and SNP mapping 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 Enzymol* **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 Biol* **14**:1–15. doi:10.1186/gb-2013-14-2-r15

Fairclough SR, Dayel MJ, King N. 2010. Multicellular development in a choanoflagellate. *Curr Biol* **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 Res* **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 Res* **44**:D279–D285. doi:10.1093/nar/gkv1344

Gilbert OM, Foster KR, Mehdiabadi NJ, Strassmann JE, Queller DC. 2007. High relatedness maintains multicellular cooperation in a social amoeba by controlling cheater mutants. *Proc Natl Acad Sci* **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. *Nat Med* **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. *J Biol Chem* **284**:34986–34997. doi:10.1074/jbc.M109.060806

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

Ho DH, Badellino K, Baglia FA, Walsh PN. 1998. A Binding Site for Heparin in the Apple 3 Domain of Factor XI. *J Biol Chem* **273**:16382–16390.

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

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

King N. 2004. The unicellular ancestry of animal development. *Dev Cell* **7**:313–325. doi:10.1016/j.devcel.2004.08.010

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

Knoll AH. 2011. The Multiple Origins of Complex Multicellularity. *Annu Rev Earth Planet Sci* **39**:217–239. doi:10.1146/annurev.earth.031208.100209

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

Lairson LL, Henrissat B, Davies GJ, Withers SG. 2008. Glycosyltransferases: Structures, Functions, and Mechanisms. *Annu Rev Biochem* **77**:521–555. doi:10.1146/annurev.biochem.76.061005.092322

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

Larsen ISB, Narimatsu Y, Joshi HJ, Siukstaite L, Harrison OJ, Brasch J, Goodman KM, Hansen L, Shapiro L, Honig B, Vakhrushev SY, Clausen H, Halim A. 2017. Discovery of an O-mannosylation pathway selectively serving cadherins and protocadherins. *Proc Natl Acad Sci* **114**:201708319. 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*. 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 Res* **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*. *Curr Biol* **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. *Curr Opin Plant Biol* **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 Res* **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. *J Biol Chem* **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*. *Int J Syst Evol Microbiol* **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. *Proc Natl Acad Sci U S A* **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. *Eukaryot Cell* **10**:724–733. doi:10.1128/EC.00016-11

940 Ratcliff WC, Denison RF, Borrello M, Travisano M. 2012. Experimental evolution of
941 multicellularity. *Proc Natl Acad Sci* **109**:1595–1600. doi:10.1073/pnas.1115323109

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

945 Rayner JC, Munro S. 1998. Identification of the MNN2 and MNN5
946 mannosyltransferases required for forming and extending the mannose branches of
947 the outer chain mannans of *Saccharomyces cerevisiae*. *J Biol Chem* **273**:26836–
948 26843. doi:10.1074/jbc.273.41.26836

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

951 Rokas A. 2008. The Origins of Multicellularity and the Early History of the Genetic
952 Toolkit For Animal Development. *Annu Rev Genet* **42**:235–251.
953 doi:10.1146/annurev.genet.42.110807.091513

954 Ruiz-Trillo I, Roger AJ, Burger G, Gray MW, Lang BF. 2008. A phylogenomic
955 investigation into the origin of Metazoa. *Mol Biol Evol* **25**:664–672.
956 doi:10.1093/molbev/msn006

957 Sawaguchi S, Varshney S, Ogawa M, Sakaidani Y, Yagi H, Takeshita K, Murohara T,
958 Kato K, Sundaram S, Stanley P, Okajima T. 2017. O-GlcNAc on NOTCH1 EGF
959 repeats regulates ligand-induced Notch signaling and vascular development in
960 mammals. *Elife* 1–30. doi:10.7554/eLife.24419

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

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

968 Strahl-Bolsinger S, Gentzsch M, Tanner W. 1999. Protein O-mannosylation. *Biochim*
969 *Biophys Acta - Gen Subj* **1426**:297–307. doi:10.1016/S0304-4165(98)00131-7

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

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

Tu L, Banfield DK. 2010. Localization of Golgi-resident glycosyltransferases. *Cell Mol Life Sci* **67**:29–41. doi:10.1007/s00018-009-0126-z

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

Wang Y, Wang F, Wang R, Zhao P, Xia Q. 2015. 2A self-cleaving peptide-based multi-gene expression system in the silkworm Bombyx mori. *Sci Rep* **5**:16273. doi:10.1038/srep16273

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

Wiggins C a, Munro S. 1998. Activity of the yeast MNN1 alpha-1,3-mannosyltransferase requires a motif conserved in many other families of glycosyltransferases. *Proc Natl Acad Sci U S A* **95**:7945–7950. doi:10.1073/pnas.95.14.7945

Woznica A, Cantley AM, Beemelmanns C, Freinkman E, Clardy J, King N. 2016. Bacterial lipids activate, synergize, and inhibit a developmental switch in choanoflagellates. *Proc Natl Acad Sci* **113**:7894–7899. doi:10.1073/pnas.1605015113

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

Zhang L, Zhang Y, Ten Hagen KG. 2008. A mucin-type O-glycosyltransferase modulates cell adhesion during *Drosophila* development. *J Biol Chem* **283**:34076–34086. doi:10.1074/jbc.M804267200

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

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

TABLES

Table 1. Phenotypes of wild type and Class C strains.

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

Table 2. Segregating variants in Rosetteless mapping cross.

Supercontig	Location	Position relative to genes	Type	Coverage†
4	516,051	intron	INDEL*	8
6	1,139,589	5' UTR	INDEL*	40
8	**427,804**	**splice donor**	**SNV****	**253**
11	524,974	intron	INDEL*	12
11	1,660,350	intron	INDEL*	6
Average genome-wide coverage: 187				

†Number of high quality reads determined by SAMtools (Li et al., 2009) at nucleotide position; *Insertion or deletion; **Single nucleotide variant; ‡Highlighted sequence variant indicates known causative lesion (Levin et al., 2014).

Table 3. Segregating variants in Jumble mapping cross.

Supercontig	Location	Position relative to genes	Type	Coverage†
1	**1,919,681**	**coding sequence**	**SNV****	**165**
20	530,561	intron	INDEL*	3
22	65,983	intron	INDEL*	37
32	134,832	intron	INDEL*	5
49	3,863	intron	SNV**	2
Average genome-wide coverage: 187				

†Number of high quality reads determined by SAMtools (Li et al., 2009) at nucleotide position; *Insertion or deletion; **Single nucleotide variant; ‡Highlighted sequence variant indicates predicted causative lesion.

Table 4. Segregating variants in Couscous mapping cross.

Supercontig	Location	Position relative to genes	Type	Coverage†
3	1,812,030	splice acceptor	INDEL*	2
4	475,982	intron	INDEL*	10
4	518,253	intron	INDEL*	12
5	533	intron	INDEL*	3
9	141,246	intron	INDEL*	3
13	698,752	intron	INDEL*	6
22	110,265	intron	INDEL*	5
22	**462,534**	**coding sequence**	**INDEL***	**128**
Average genome-wide coverage: 72				

†Number of high quality reads determined by SAMtools (Li et al., 2009) at nucleotide position; *Insertion or deletion; ‡Highlighted sequence variant indicates predicted causative lesion.

Figure 1. Class C mutant cells aggregate and fail to form rosettes. (A) Wild type cells form linear chains in the absence of rosette inducing factors (RIFs) and develop into organized spherical rosettes with the apical flagellum of each cell pointing outward in response to RIFs. Rosettes are resistant to shear force and survive vortexing. Four class C mutants—Seafoam, Soapsuds, Couscous, and Jumble—form disorganized clumps of cells in the presence and absence of RIFs. The clumps are not resistant to vortexing and fall apart. **(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. ⌀ indicates that no rosettes were observed. **(C)** Automated image analysis allowed quantification of the area of clumps 30 minutes after cells were vortexed in the absence of RIFs. 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). **(D)** Clumping occurs within minutes after vortexing in the Class C mutants without RIFs. DIC images obtained at 0, 15, and 30 minutes post-vortexing.

Figure 1

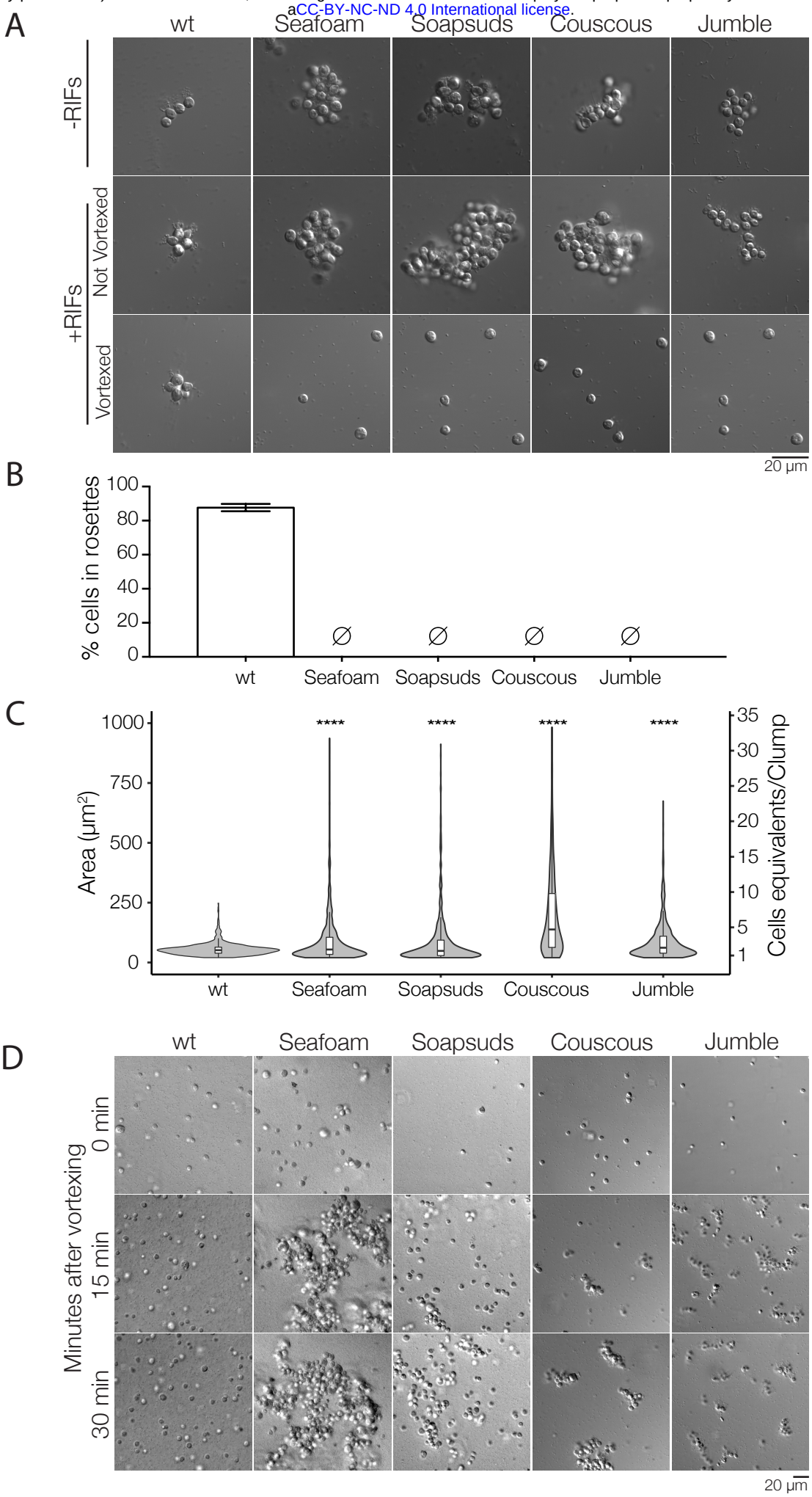
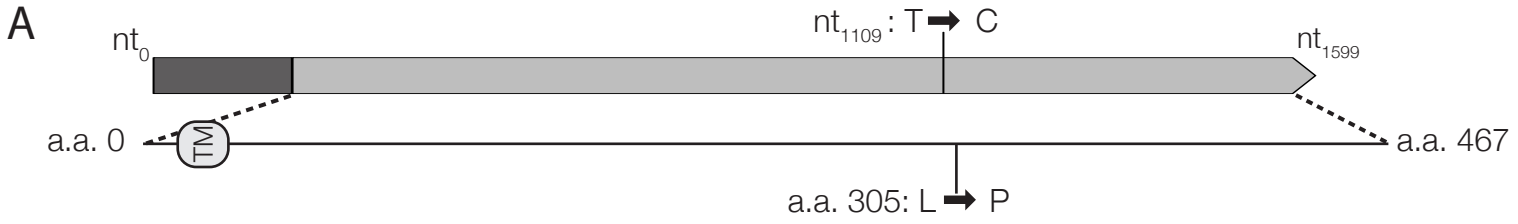


Figure 2. Jumble maps to a predicted glycosyltransferase that localizes to the Golgi apparatus. (A) Jumble has a predicted transmembrane domain (marked TM), but no conserved domains by sequence identity. The mutant gene has a T to C mutation at nucleotide 1109 that causes a missense mutation from proline to leucine at amino acid position 305. **(B)** A single mutant F1×Mapping Strain backcross confirms the tight linkage of the *jumble*^{lw1} to the clumpy, rosetteless phenotype (Chi-squared test, p<0.001). **(C)** Transgene expression of *jumble-mTFP* and *mTFP-jumble* rescued rosette formation in the Jumble mutant, but *jumble*^{lw1}*-mTFP*, *mTFP-jumble*^{lw1}, or *mTFP* alone did not. RIFs were immediately after transfection, but 40 µg/mL puromycin was added 24 hours post-transfection. Representative rosette expressing *jumble-mTFP* shown. Rosette development was measured as the % of cells in rosettes 72 hr post-transfection and shown as mean ± SD. ⌀ indicates that no rosettes were observed (n=200 cells from 2 biological replicates with 3 technical replicates each). Transgenes of membrane marker-mCherry (magenta) and Jumble-mWasabi (green) under the *efl* promoter were expressed in wild type *S. rosetta* **(D)** without RIFs and **(E)** with RIFs. Jumble-mWasabi localizes to the basal pole of cells, consistent with the localization of the Golgi apparatus. Expression of the mutant Jumble^{lw1}-mWasabi transgene in **(F)** without RIFs and **(G)** with RIFs wild type cells incorrectly localizes to the ER and food vacuole. Boxes indicate the presumed location of the Golgi apparatus at the basal pole. Asterisks denote the food vacuole which may be fluorescent due to autofluorescence from ingested bacteria or through accumulation of the fluorescent markers.

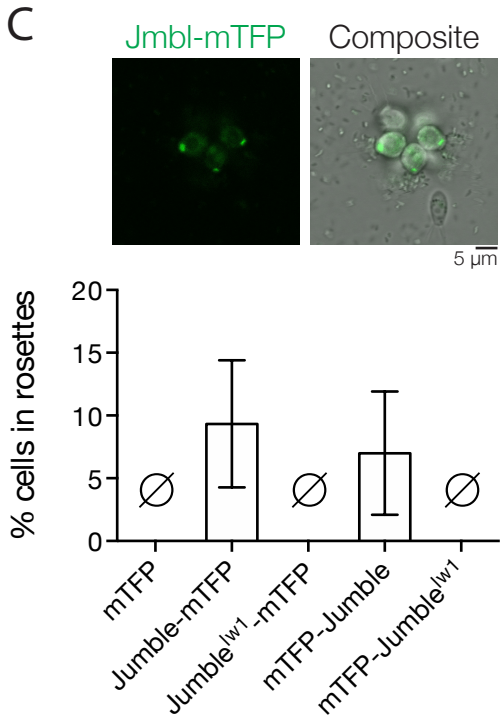
Figure 2



B

Jumble		
	Rosettes	Clumps
T	9	0
C	0	12

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



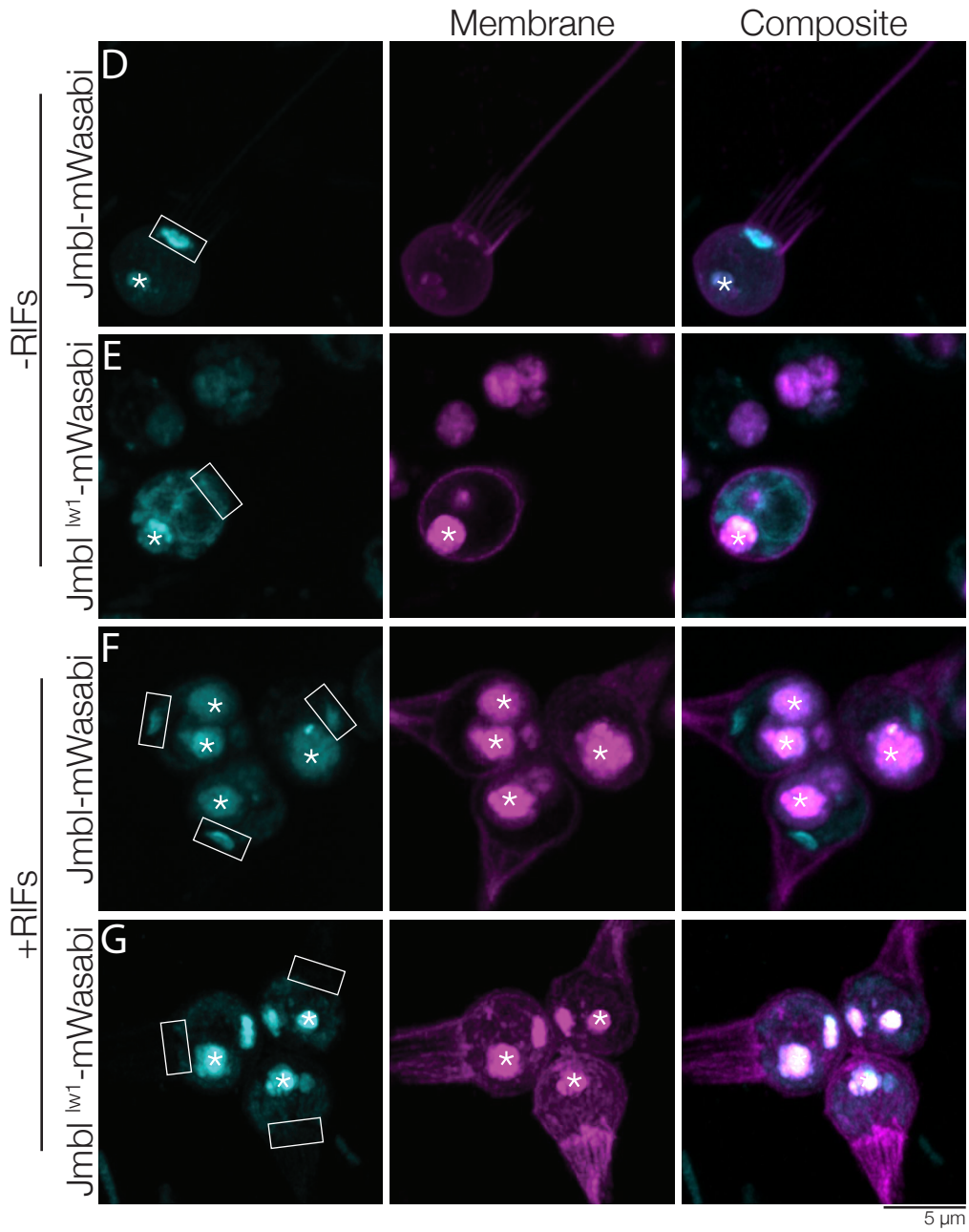
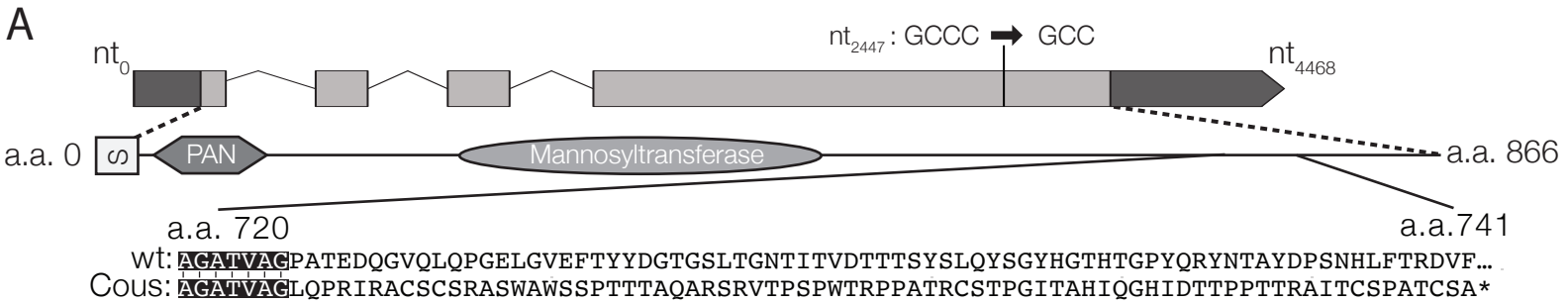


Figure 3. Couscous maps to a predicted mannosyltransferase with a PAN/Apple domain. (A) Couscous has a predicted signal sequence (marked S), a PAN/Apple domain (marked PAN), and a mannosyltransferase domain. The causative lesion is a 1-base pair deletion at position 2447 to cause a frameshift that leads to an early stop codon at amino acid 741. **(B)** Two F1 mutant×Mapping Strain backcrosses confirmed the tight linkage between the mutation and the rosetteless, clumpy phenotype (Chi-squared test, $p<0.001$). **(C)** Transgene expression of *couscous-mTFP* or *mTFP-couscous* could rescue rosette formation in Jumble, but not *couscous^{lw1}-mTFP*, *mTFP-couscous^{lw1}*, or *mTFP* alone. RIFs were immediately after transfection, but 40 µg/mL puromycin was added 24 hours post-transfection. Representative rosette shown. **(D)** Rosette development was measured as the % of cells in rosettes after 72 hr after and is shown as mean ± SD. ⌀ indicates that no rosettes were observed (n=200 cells from 2 biological replicates with 3 technical replicates each).

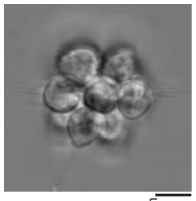
Figure 3



B

Couscous		
	Rosettes	Clumps
GCCC	38	0
GCC	0	51

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



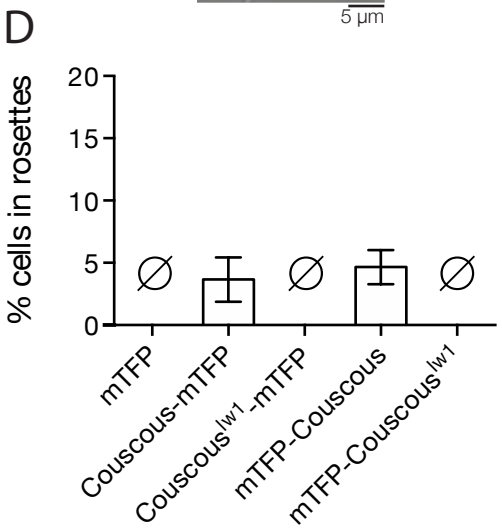
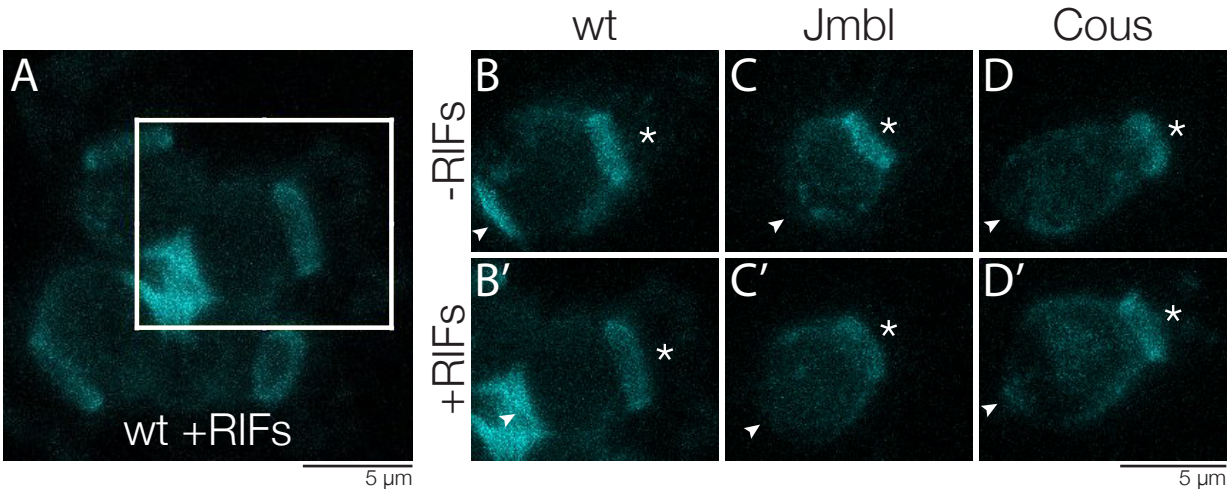


Figure 4. Jumble and Couscous mutants lack sugar modifications at the basal pole. Jacalin recognizes a sugar modification at the basal pole of wild type cells in single cells **(B)** and becomes enriched at ECM in the center of rosettes **(A, B')**. Jumble (**C, C')** and Couscous **(D, D')** have severely reduced or abolished jacalin at the basal poles of cells, both without and with RIFs. Asterisks mark the apical pole and arrow mark the basal pole. **(E)** Jacalin fluorescence was measured by drawing an arc from one edge of the collar around the cell body to the other edge of the collar, and the line was normalized for cell size and background intensity. **(F)** The average normalized fluorescence intensity from at least 59 cells (n=2 biological replicates) from each condition is graphed against the normalized length of the cell body. Gray shadows indicate 95% confidence intervals.

Figure 4



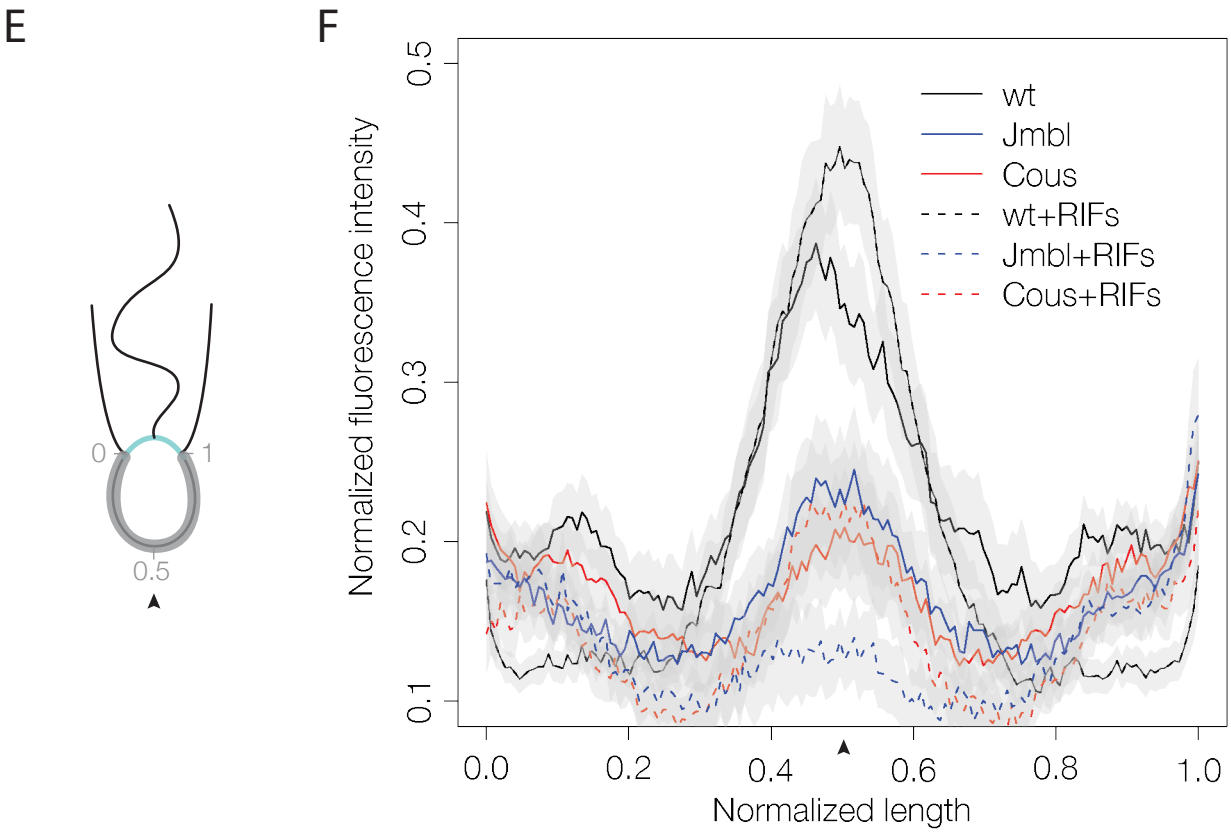


Figure 5. Model for promiscuous clumping in rosette defective Class C mutants.

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 observed at the basal pole of wild type cells before induction. The altered cell surface could lead to clumping through either mis-regulating cell adhesion molecules or exposing an adhesive cell surface. The same alteration to the cell surface of Class C mutants may prevent the proper formation of ECM required for rosette formation. Potentially, the disrupted glycan modification could be on the Rosetteless protein or on one of its interaction partners.

Figure 5