Insights into the evolution of epithelium from non-metazoan eukaryotes

By

Tess Augusta Linden

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Molecular and Cell Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Nicole King, Chair
Professor David Bilder
Professor Michael Eisen
Professor Sanjay Kumar

Fall 2021

Abstract

Insights into the evolution of epithelium from non-metazoan eukaryotes

By

Tess Augusta Linden

Doctor of Philosophy in Molecular and Cell Biology

University of California, Berkeley

Professor Nicole King, Chair

The ability to form a single body with distinct organs that perform specialized tasks is one of the defining features of metazoan complex multicellularity. Crucial for this ability is the epithelium, a tissue made of a tightly adhered sheet of cells that is capable of compartmentalizing organs within metazoan bodies. Epithelia are recognizable across metazoans by several major characteristics mediated by conserved epithelial proteins. Cells within epithelial sheets adhere to each other through the interaction of cadherin with beta-catenin, creating a seal of reduced permeability to diffusing molecules. Within the sheet, cells are polarized along the same apical-basal axis. At the basal pole, integrins attach the cell to the basement membrane, which is composed of networks of collagen IV and laminin and serves as a scaffold and signaling hub. At the apical pole, the action of actin and myosin networks allows epithelial cells to undergo apical constriction, a cell-shape change that can cause bending of epithelial sheets. Across metazoans, epithelial sheet bending is crucial for morphogenetic rearrangements, such as gastrulation, during embryonic development.

Since epithelia with these characteristics have been identified in all major metazoan phyla from sponges to chordates, it is likely that the last common ancestor of all metazoans already possessed an epithelium. However, key questions remain about the evolution of this crucial tissue type. When did the domains and proteins necessary for the epithelium originate? If they evolved in earlier, single-celled ancestors of metazoans, what might their original functions have been? How did apical constriction and tissue bending evolve? Might there have been precursors for those cell behaviors that predated epithelia and morphogenesis?

To answer these questions, one promising avenue of research is to study the biology of metazoans' extant relatives, which can illuminate events that occurred prior to metazoan origins. The sister taxon of the metazoans is the choanoflagellates, a group of single-celled and colonial microeukaryotes that are found in aquatic environments across the globe. In Chapter 2, we report the discovery of a new choanoflagellate species from Curaçao that forms colonies reminiscent of epithelial sheets: each colony is a cup-shaped monolayer of tightly adhered cells sharing the same apical-basal orientation. Colonies of the new species, which we name *Choanoeca flexa*, exhibit collective contractility, undergoing reversible inversion from a relaxed, flagella-in state to a contracted, flagella-out state in response to light-to-dark transitions. We show that *C. flexa* phototransduction is controlled by a rhodopsin photoreceptor with cGMP as a second messenger, similar to the pathway for vision in many metazoans. Furthermore, *C. flexa* colony inversion

occurs through actomyosin-mediated apical constriction, much like the mechanism of epithelial sheet bending in metazoan morphogenesis. Based on observations of apical contractility in *C. flexa* colonies and in single-celled forms of other choanoflagellate species, we conclude that apical constriction evolved prior to the divergence of choanoflagellates and metazoans, while apical constriction-mediated collective contractility appears to have evolved independently in *C. flexa* and metazoans.

In Chapter 3, we investigate the evolution of collagen, one of the key components of the epithelial basement membrane. Collagens were previously thought to be unique to metazoans until the first choanoflagellate genomes were sequenced and found to encode collagen domains. We conduct the first comprehensive survey of collagens across eukaryotes, and find that collagen domains are not only ubiquitous in choanoflagellates, but in fact are found in almost all major groups of eukaryotes. Furthermore, we find that collagen repeat length and amino acid composition in non-metazoans rivals those of metazoan collagens. Specific types of collagens—including long variants of collagen IV, which scaffolds the epithelial basement membrane—appear to be restricted to metazoans and may therefore explain unique features of metazoan biology.

Table of Contents

List of Figures and Tables

Chapter 1

Chapter 2

Chapter 3

Acknowledgements

I would like to thank my advisor, Nicole King, who made all of this possible, and who has taught me countless invaluable lessons about biology, writing, public speaking, personal responsibility, and what it means to be a good scientific thinker. I would also like to thank my committee members, David Bilder, Michael Eisen, and Sanjay Kumar, for their helpful advice and feedback over the years.

I would like to thank all of the members of the King Lab, past and present: I could not imagine a more collaborative, caring, and generous set of colleagues. Their kindness and support is a model for how scientists should act towards each other. I would especially like to thank my collaborators Thibaut Brunet and Ben Larson for filling my workdays with puns, laughter, and illuminating philosophical discussions.

I would like to thank Dana and Larry Linden for being basically perfect parents and for putting up with all of the stupid things I said in high school (and college, and graduate school, and in the future). I would also like to thank my sister, Maya Linden, whose emotional support has been more important than she knows, and I only hope I can pay it back someday. I would like to thank the entire MCB community for being not just a fascinating and intellectually stimulating bunch but also a rollicking good time. In particular, thank you to my housemates Josh Cofsky, Paige Diamond, Maya Emmons-Bell, Justin Roncaioli, and honorary housemate Arik Shams, who made grad school fun, even on the bad days. Thank you times a thousand to Daniel Mossing, for having more faith in me than I do, and for reminding me in many ways that what matters most in life is love. Finally, I thank Annie and Grinnell and the Cal Falcon Cam team for their constant emotional support.

Chapter 1

Reconstructing the evolution of the epithelium

Epithelium is a defining characteristic of metazoans

All metazoans alive today, from sponges to humans, are descendants of a single common ancestor that lived more than 600 million years ago (Cunningham et al., 2017). By comparing traits shared by most or all extant metazoans, it is possible to reconstruct features of the cell biology, life history, and development of this common ancestor, known as the Urmetazoan (Richter and King, 2013). For example, all metazoans undergo clonal development: a single cell divides into many daughter cells that differentiate into morphologically distinct cell types, which remain adhered to each other and cooperate to form an individual organism. The ubiquity of this life history trait across metazoan phyla indicates that it was a characteristic of the Urmetazoan (Richter and King, 2013; Sebé-Pedrós et al., 2017).

A key feature of metazoan development is cell differentiation, a process that allows individual cells to take on specialized gene expression, physiology and morphology for specialized functions. Which cell types were present in the Urmetazoan is a subject of debate (Arendt et al., 2016; Brunet and King, 2017; King and Rokas, 2017). However, it is clear that one of the earliest and most fundamental cell types to emerge in metazoan evolution is the epithelium (Leys and Riesgo, 2012; Richter and King, 2013; Tyler, 2003). By performing the function of sealing and compartmentalizing tissues and organs, the epithelium is crucial to the development of metazoan body complexity. Epithelial tissues are recognizable across all major metazoan phyla by several defining characteristics (Fig. 1.1). First, epithelial cells are polarized along an apical-basal axis (Gibson and Perrimon, 2003; Tyler, 2003; Vasquez et al., 2021). Second, the cells adhere to each other in sheets, with all cells having the same apicobasal orientation (Gibson and Perrimon, 2003; Tyler, 2003; Vasquez et al., 2021). Third, the cells are in contact with each other via cell-cell junctions, creating a seal with restricted permeability to diffusing molecules (Adams et al., 2010; Leys and Riesgo, 2012; Tyler, 2003). Finally, the cells adhere at their basal pole to a specialized extracellular matrix (ECM), also known as a basal lamina or basement membrane, which performs structural and signaling roles (Yurchenco, 2011).

While these universal characteristics of epithelia were first recognized morphologically across diverse metazoans using microscopy and histology, later molecular research revealed that homologous proteins likely underlie many of these morphological characteristics in distantly related metazoan phyla (Leys et al., 2009; Leys and Riesgo, 2012; Tyler, 2003). For example, the Par, Crumbs, and Discs large protein complexes appear to be involved in the establishment of apical-basal polarity in epithelia across metazoans (Fahey and Degnan, 2010; Richter and King, 2013; Srivastava et al., 2010). Similarly, cadherins and beta-catenin form adherens junctions between epithelial cells (Leys and Riesgo, 2012; Tyler, 2003), while collagen IV and laminin scaffold the basement membrane underlying the epithelium across metazoan phyla (Boute et al., 1997; Fidler et al., 2017).

Crucial to understanding the evolution of epithelia is determining whether an epithelium meeting the characteristics above was present in the Urmetazoan, or whether it evolved later in metazoan evolution. This understanding hinges on whether such a tissue can be identified in the most basally branching metazoans (Leys and Riesgo, 2012). Sponges (phylum Porifera)—which along with ctenophores (phylum Ctenophora) are considered candidates for the most basally

branching metazoan phylum (King and Rokas, 2017)—were historically thought not to possess epithelia with a capacity for sealing and compartmentalization of tissues (Tyler, 2003). In accordance with a general conceptualization of sponges as the least complex metazoans (rather "loose associations of transiently differentiated cells" (Leys et al., 2009)), it was thought that because sponge epithelia appeared to lack the tight or septate junctions typical of bilaterian epithelial tissues, they were "probably quite permeable, almost mesenchymal in nature" (Tyler, 2003). However, through functional experiments, Adams et al. (2010) found that epithelia in freshwater sponges (Demospongiae) in fact do restrict passage of ions across the epithelial layer. Thus, epithelia with the ability to create a chemically regulated internal environment appear to be a feature of all metazoan phyla, suggesting that it may also have been a feature of the Urmetazoan. Furthermore, while sponges were historically thought not to have basement membranes, Boute et al. (1997) found that the epithelia of the homoscleromorph sponge *Pseudocorticium jarrei* has a basement membrane scaffolded by collagen IV, similar to those of eumetazoans. Similarly, in Ctenophora, the other candidate sister phylum of eumetazoans (King and Rokas, 2017), collagen IV-based basement membranes are found in some, but not all, species (Fidler et al., 2017). Thus, based on their presence in all major metazoan phyla, sealing epithelia with a collagenous basement membrane are thought to have been present in the Urmetazoan, potentially allowing early compartmentalization of tissues. As a crucial tool for the separation of distinct tissues and the organization of metazoan bodies, the evolution of epithelium may have been a key innovation that enabled complex multicellular body plans to evolve in the metazoan lineage.

The epithelial basement membrane is a scaffold and signaling hub

One of the defining characteristics of epithelium is a basement membrane, the specialized basal ECM that serves as a physical scaffold and signaling hub at the basal pole of epithelial cells (Boute et al., 1997; Fidler et al., 2017; Tyler, 2003; Yurchenco, 2011). The key structural components of the basement membrane across metazoans are the proteins collagen IV and laminin (Boute et al., 1997; Fidler et al., 2017). Collagen IV is a large, rope-like protein composed of many short units of the "collagen triple helix," a rigid, rod-like structure resulting from repetition of the collagen motif Gly-X-Y, where X and Y can be any amino acid (Ricard-Blum, 2011). The non-collagenous domains of collagen IV promote oligomerization, resulting in the formation of strong, flexible mesh-like networks that scaffold the basement membrane (Ricard-Blum, 2011). Laminin, the other major structural protein of metazoan basement membranes, is a large heterotrimeric protein composed of three subunits, termed alpha, beta, and gamma. Together, these subunits form a cross-like structure with three short arms and one long arm, where each of the short arms corresponds to an N-terminal region of one of the three subunits, and the long arm corresponds to the C-terminal regions of the three subunits (Aumailley, 2013; Engvall and Wewer, 1996). The short arms mediate polymerization of many laminin proteins into a laminin network that helps scaffold the assembly of the basement membrane (Engvall and Wewer, 1996; Leys et al., 2009), and they also mediate interactions between laminin and other ECM proteins (Aumailley, 2013). Meanwhile, the C-terminal region of the laminin alpha subunit interacts with cell-surface receptors such as integrins and is thereby involved in dynamic regulation of cell behavior, such as adhesion, migration, and differentiation (Aumailley, 2013).

In addition to collagen IV and laminin, other key components of the basement membrane in eumetazoans are nidogen and perlecan (Hynes, 2012; Yurchenco, 2011). Nidogen is a glycoprotein that binds strongly to laminin and also forms associations with collagen IV and perlecan, and is thus thought to facilitate the assembly of the basement membrane (Dziadek, 1995). Perlecan is a heparan-sulfate proteoglycan that binds to many other factors and thus plays diverse roles in assembly and signaling within eumetazoan basement membranes (Knox and Whitelock, 2006; Whitelock et al., 2008; Yurchenco, 2011). Interestingly, although these molecules are key BM components across eumetazoa (Hynes, 2012; Yurchenco, 2011), they do not appear to be present in ctenophores or sponges, many species of which nevertheless construct functional basement membranes (Fidler et al., 2017). This raises the question of whether ctenophores and sponges have found alternative evolutionary solutions for the construction of basement membranes, or whether nidogen and perlecan may have replaced more ancestral basement membrane components at some point after the divergence of ctenophores/sponges and eumetazoans. Further research into basement membrane structure and function in basal metazoans may shed light on this evolutionary history.

The basement membrane is not only a structural scaffold for epithelia, but also a signaling hub that dynamically interacts with epithelial cells. Integrins are key mediators of this interaction (Hynes, 2012; Paulsson, 1992; Yurchenco, 2011). Integrins are a family of heterodimeric transmembrane proteins found in all metazoan phyla (Hynes, 2012, 2002). One of the functions of integrins in epithelia is to allow the cell to physically adhere to the basement membrane through binding of laminin and collagen IV (Yurchenco, 2011). On top of that structural, adhesive function, integrins play diverse signaling roles in epithelia. Through their intracellular attachments to the cytoskeleton and extracellular attachments to laminin and collagen networks, integrins dynamically regulate epithelial cell shape (Domínguez-Giménez et al., 2007; Mateos et al., 2020). In addition, integrin-BM interactions can regulate epithelial cell differentiation into specialized epithelial cell types (Sheppard, 2003; Streuli, 2016; Streuli et al., 1991) or into mesenchymal cells via the epithelial-to-mesenchymal transition (Yeh et al., 2012). Integrin-BM interactions are also required for the proper establishment of apical-basal polarity in epithelia due to their role in recruiting proteins to the basal pole of epithelial cells (Lee and Streuli, 2014). Other cell behaviors affected by integrin-BM signaling include cell migration, apoptosis, and proliferation (Sheppard, 1996). While integrins are not the only proteins responsible for BM-related signaling, the remarkable diversity of cell behaviors and processes that are affected by integrin-BM interactions demonstrates the importance of the epithelial basement membrane as a dynamic signaling hub.

Epithelial sheet bending is important for metazoan morphogenesis

In addition to their passive role in shaping metazoan bodies through the compartmentalization of tissues, epithelia can play physically active roles in morphogenesis through contraction and deformation. Arguably the most dramatic remodeling event of metazoan embryonic development is the process known as gastrulation (Gilbert, 2013; Leptin, 2005). During this process, which is conserved across all major metazoan phyla (King and Rokas, 2017), the embryo goes from a simple ball of cells to an organized metazoan body plan, having at least two distinguishable germ layers (Leptin, 2005). In eumetazoans, gastrulation also establishes a defined body axis and gut (Gilbert, 2013; Leptin, 2005). In many metazoans,

epithelial sheet bending is crucial for gastrulation to occur. Perhaps one of the most dramatic examples is the gastrulation of calcareous sponges: the blastula-stage embryo, which is essentially a hollow ball of epithelial cells with internally-facing cilia, undergoes eversion, i.e., it turns completely inside out, resulting in the cilia facing outward (Leys and Eerkes-Medrano, 2005a). Epithelial deformations are also required for the well-studied gastrulation process of the fruit fly *Drosophila melanogaster*, during which invagination of the blastoderm epithelium drives the internalization of presumptive mesodermal and endodermal cells into the interior of the embryo (Leptin, 1999). In chordates, two fundamental morphogenetic events, gastrulation and neurulation, both depend on epithelial contractility. Like gastrulation, neurulation begins with the invagination of epithelial cells to form an internal furrow; the internal cells then pinch off to form the neural tube (Vijayraghavan and Davidson, 2017). Thus, epithelial deformations play a crucial role in the morphogenesis of diverse metazoans.

The primary mechanism through which epithelia undergo deformation is apical constriction (Martin and Goldstein, 2014; Sawyer et al., 2010). Apical constriction refers to the narrowing of the apical side of the cell, which often results in a change in the shape of the cell from columnar or cuboidal to wedge-shaped or bottle-shaped. Since epithelial cells are tightly adhered to each other, local apical constriction of some or all cells within an epithelial sheet can cause bending of the entire sheet as cells pull on one another (Martin and Goldstein, 2014). Across metazoans, apical constriction involves a conserved set of machinery. Type II myosin functions as a motor that pulls on the network of cortical actin filaments at the apical pole of the cell, where they are anchored to the membrane by association with adherens junctions (Martin and Goldstein, 2014). In addition to actin and myosin, many of the upstream regulators of apical constriction are conserved from bilaterians to the most basally branching metazoans. One example is Rho-associated coiled-coil kinase (ROCK), which activates myosin by phosphorylating the myosin regulatory light chain (Martin and Goldstein, 2014). ROCK homologs have been identified in sponges and cnidarians, where ROCK inhibition disrupts morphogenesis and budding, respectively, by interfering with cell shape change (Beljan et al., 2020). Thus, epithelial bending through actomyosin-mediated apical constriction is a defining characteristic of metazoan morphogenesis, and was likely present in the Urmetazoan.

Non-metazoan eukaryotes can provide insight into the evolution of epithelium

Because an epithelium was already present in the Urmetazoan, this leaves open the question of how the epithelium evolved in the lineage leading to Metazoa. In order to illuminate this history, one promising approach is to study the closest living relatives of metazoans and use their biology to reconstruct earlier ancestors that predate the Urmetazoan. Metazoa resides within a group of eukaryotes known as the Holozoa (Fig. 1.2); within the Holozoa, the sister taxon of metazoans is the choanoflagellates, a group of free-living, aquatic, bacterivorous microeukaryotes that can be found as single cells or as colonies (Leadbeater, 2015). When the first choanoflagellate genome was sequenced, it was discovered that many genes required for metazoan epithelia are also present in choanoflagellates. Specifically, the solitary choanoflagellate *Monosiga brevicollis* was found to encode integrin alpha (one of the two subunits of the integrin heterodimer) and cadherins (King et al., 2008). In addition, *M. brevicollis* was found to encode several key domains necessary in epithelial proteins, such as collagen triple helix domains, laminin G domains, and laminin N terminal domains, without encoding a

complete ortholog of collagen IV or laminin (King et al., 2008). Thus, these epithelium-related genes and domains were likely present in the last common ancestor of choanoflagellates and metazoans.

As more non-metazoan holozoans continued to be sequenced, an increasing number of proteins and domains relevant to epithelium were discovered to predate metazoan origins. Richter et al. (2018) sequenced the transcriptomes of 19 choanoflagellate species spanning the choanoflagellate phylogeny. Consistent with their finding that the taxon of Choanoflagellata contains as much internal genetic diversity as its sister taxon Metazoa, Richter et al. were able to identify many genes relevant to metazoan biology with variable patterns of gains and losses within the choanoflagellate lineage. For instance, while the previously sequenced laboratory model organisms *M. brevicollis* and *Salpingoeca rosetta* do not encode integrin beta, the species *Didymoeca costata* does (Richter et al., 2018), which (based on the current understanding of choanoflagellate phylogenetic relationships) suggests that integrin beta was lost independently many times over the course of choanoflagellate evolution. Investigations of other groups of holozoans has also been fruitful. As one example, Grau-Bové et al. (2017) found that the filasterean *Ministeria vibrans* encodes the NC1 domain crucial to the metazoan basement membrane scaffolding protein collagen IV, pushing back the timeline of this domain's evolution to the last common ancestor of metazoans and filastereans. The *M. vibrans* protein containing the NC1 domain also contains collagen triple helix repeats, which raises the possibility that collagen IV itself predates metazoan origins, a hypothesis interrogated further in Chapter 3.

Investigation of the cell biology of metazoan relatives also stands to illuminate the history of epithelial evolution, as well as to uncover novel cell-biological phenomena not found within Metazoa. Many aspects of holozoan biology can be identified that bear resemblances to, and potentially may be homologous with, characteristics of epithelia. Like epithelial cells, choanoflagellates are apicobasally polarized, with an apical flagellum homologous to the apical cilium of metazoan ciliated epithelium (Brunet and King, 2017). This evolutionary relationship enabled Sigg et al. (2017) to discover a previously unknown ciliary protein, ENKUR, which localizes to the flagellum/cilium of both choanoflagellates and metazoans, and plays a role in situs inversus (a developmental disorder of disrupted left-right patterning) in humans. Cell-cell associations in choanoflagellates also share characteristics with epithelial sheets: when some choanoflagellates form multicellular colonies, the cells are typically uniformly apicobasally oriented (Dayel et al., 2011; Leadbeater, 2015, 1977) and adhere to a basally secreted ECM (Larson et al., 2020; Levin et al., 2014). Though even further removed from Metazoa than choanoflagellates, the ichthyosporean *Sphaeroforma arctica* was recently described to exhibit a mode of cell division with a striking resemblance to epithelial cellularization during metazoan embryonic development (Dudin et al., 2019). An *S. arctica* single cell undergoes several rounds of nuclear division to form a coenocyte, which then undergoes actomyosin-mediated plasma membrane invaginations to form a layer of polarized cells—all words that could equally apply to the process of cellularization in *Drosophila melanogaster* or *Caenorhabditis elegans* embryonic epithelia (Dudin et al., 2019).

Whether the epithelium-like characteristics of choanoflagellate colonies and ichthyosporean cellularization are the products of evolutionary convergence, or whether they are ancestral traits passed down from a stem holozoan, is difficult to determine. Some researchers have even argued that epithelial multicellular sheets are so ancient as to predate the last common ancestor of amoebozoans and opisthokonts (Dickinson et al., 2012a). Dickinson et al. (2011) found that aggregative multicellular colonies of the amoebozoan *Dictyostelium discoideum* form

sheets of polarized cells and that *D. discoideum* homologs of alpha- and beta-catenin are required for proper sheet formation and cell polarization, as is true in metazoan epithelium. Unlike in metazoans, however, cell-cell adhesion in *D. discoideum* colonies is cadherin-independent (Dickinson et al., 2011), since cadherins are not present outside holozoans (Nichols et al., 2012). Nevertheless, Dickinson et al. (2012a) posit that the similarities between *D. discoideum* and metazoan epithelium are best explained by a scenario in which a metazoan-like multicellular epithelium was present in an ancient eukaryotic ancestor of amoebozoans and opisthokonts. This contrasts with the view in which multicellularity evolved many times independently within amoebozoans and opisthokonts (including multiple times within fungi and once at the base of metazoans)—a view supported by the fact that *D. discoideum* multicellularity is aggregative rather than clonal, and that most species spanning the evolutionary distance between *D. discoideum* and Metazoa are unicellular (Knoll, 2011; Parfrey and Lahr, 2013; Sebé-Pedrós et al., 2017). Further investigation of the molecules underlying *D. discoideum* epithelium formation, as well as further investigation of the diversity of multicellular life stages in non-metazoans, may provide additional evolutionarily relevant information to differentiate between the two models.

In general, studying the gene content and cell biology of holozoans and of more distant metazoan relatives stands to illuminate the evolutionary history of the epithelium, and thus the history of metazoan complex multicellularity. Many outstanding questions remain with regard to the evolution of epithelia. How did apical constriction and epithelial tissue bending evolve? How and when (and how many times) did collagen, which underlies the epithelial basement membrane, evolve? In Chapter 2, we describe the discovery of a new species of choanoflagellate that forms multicellular sheet colonies capable of collective, actomysosin-dependent apical constriction, a behavior resembling epithelial tissue bending during embryonic development. In Chapter 3, we take advantage of recently sequenced choanoflagellate, filasterean, and icthyosporean genomes and transcriptomes, as well as publicly available proteomes of diverse eukaryotes, to investigate the evolutionary history of collagen, and find that collagens are more widespread, and perhaps more ancient, than previously appreciated. Both of these studies demonstrate that research into the biology of diverse relatives of Metazoa has the potential to shed light on the history of epithelial components and behaviors—and to uncover novel biological phenomena that are interesting in their own right.

FIGURES

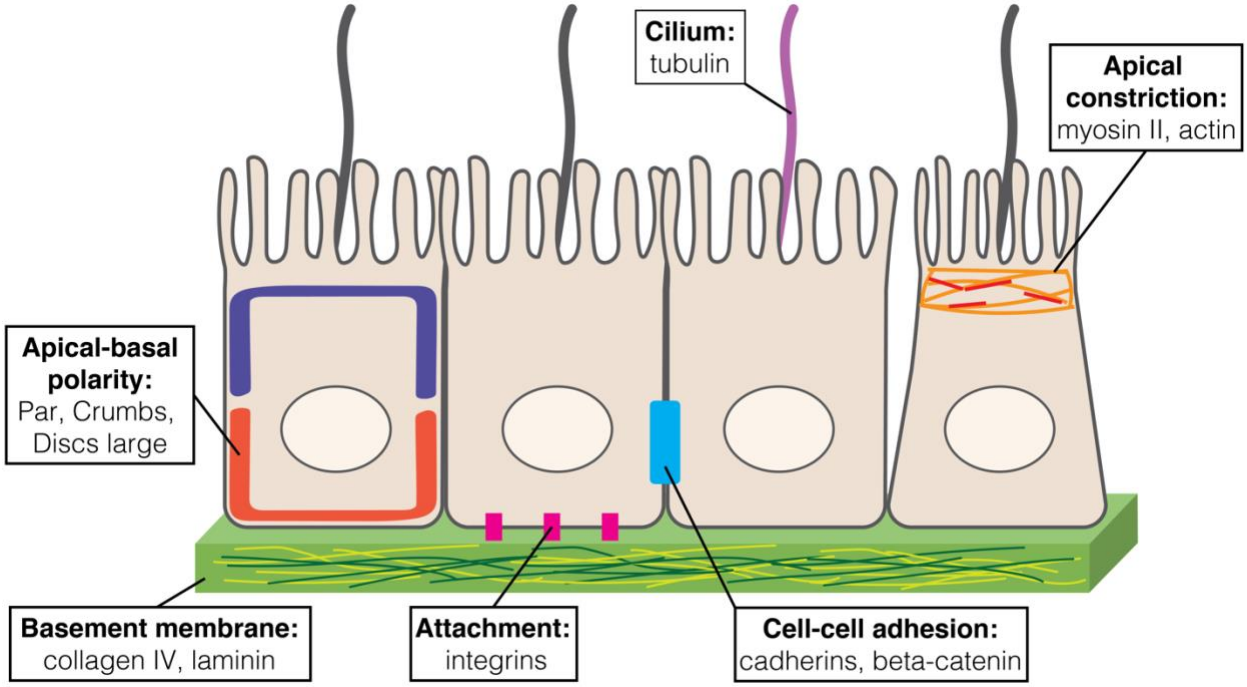


Figure 1.1 Conserved proteins underlie the characteristics of epithelia across Metazoa. Diagram of cells in an epithelial sheet showing the molecular and cellular characteristics of epithelia that are conserved across Metazoan phyla and were therefore likely present in the Urmetazoan. (From left to right) **Apical-basal polarity:** The Par, Crumbs, and Discs large protein complexes, which are required for the establishment of apical-basal polarity in epithelial cells, are shared across metazoans. **Basement membrane:** A specialized ECM known as the basement membrane underlies the basal side of epithelial sheets. The conserved proteins collagen IV and laminin assemble into networks that are the physical scaffold of the basement membrane. **Attachment:** Epithelial cells attach to the basement membrane using integrins, transmembrane proteins whose extracellular regions bind to laminin and collagen. **Cilium:** Many epithelial cells have a cilium, an organelle located at the apical pole of the cell that has a conserved microtubule-based morphology. **Cell-cell adhesion:** The tight adhesion of cells in epithelial sheets through cell-cell junctions is important for their ability to seal and compartmentalize tissues. Cadherins and beta-catenin are required for epithelial cell-cell adhesion across metazoans. **Apical constriction:** Epithelial sheets across metazoans can bend via the cell shape change of apical constriction, which is mediated by myosin II and actin networks.

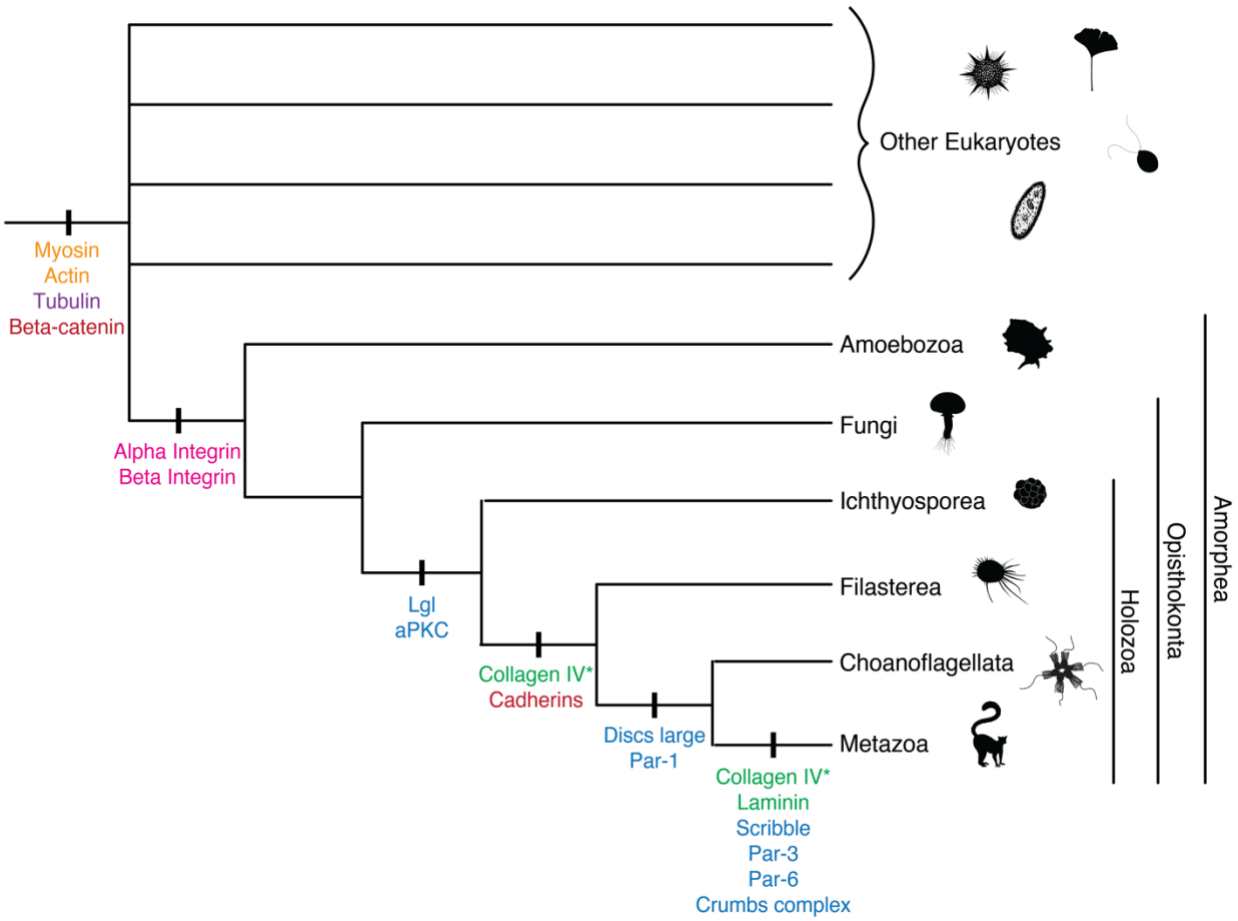


Figure 1.2 Genes relevant to metazoan epithelia predate metazoan origins. Phylogenetic tree representing the evolutionary relationships between metazoans and their eukaryotic relatives. Tick marks represent the evolution of epithelia-related genes; note that subsequent losses of these genes in some lineages are not represented here. Two key proteins involved in apical constriction (orange), myosin and actin, were likely present in the last common ancestor of all eukaryotes (Sebé-Pedrós et al., 2014; Velle and Fritz-Laylin, 2019). The same is true of tubulin, the major cytoskeletal protein of the cilium (purple) (Pollard and Goldman, 2018). Beta-catenin, which is important for epithelial cell-cell adhesion (red), may also predate eukaryotic origins, while cadherins are a more recent innovation of the Holozoa (Nichols et al., 2012). The conserved set of protein complexes relevant to the establishment of apical-basal polarity (blue) includes the Discs large complex (Lgl, Discs large, Scribble), the Par complex (aPKC, Par-1, Par-3, Par-6), and the Crumbs complex (Fahey and Degnan, 2010); of these, Lgl, aPKC, Discs large, and Par-1 appear to predate the Urmetazoan (Fahey and Degnan, 2010; Sebé-Pedrós et al., 2012). Of the conserved basement membrane proteins (green), laminin appears to be a metazoan innovation (Hynes, 2012), while the evolutionary history of collagen IV is less clear. *: Collagen IV may have evolved prior to the divergence of filastereans and metazoans, or may be a metazoan innovation; for further discussion, see Chapter 3.

Chapter 2

Light-regulated collective contractility in a multicellular choanoflagellate

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

Brunet, T.*, Larson, B.T.*, Linden, T.A.*, Vermeij, M.J.A., McDonald, K., King, N. (2019) Light-regulated collective contractility in a multicellular choanoflagellate. *Science* 366: 326-334.

Abstract

Collective cell contractions that generate global tissue deformations are a signature feature of animal movement and morphogenesis. However, the origin of collective contractility in animals remains unclear. While surveying the Caribbean island of Curaçao for choanoflagellates, the closest living relatives of animals, we isolated a previously undescribed species (here named *Choanoeca flexa* sp. nov.) that forms multicellular cup-shaped colonies. The colonies rapidly invert their curvature in response to changing light levels, which they detect through a rhodopsin–cyclic guanosine monophosphate pathway. Inversion requires actomyosin-mediated apical contractility and allows alternation between feeding and swimming behavior. *C. flexa* thus rapidly converts sensory inputs directly into multicellular contractions. These findings may inform reconstructions of hypothesized animal ancestors that existed before the evolution of specialized sensory and contractile cells.

Introduction

The evolution of animals from their single-celled ancestors involved several major evolutionary innovations, including multicellularity, spatial cell differentiation, and morphogenesis (Brunet and King, 2017; Sebé-Pedrós et al., 2017). Efforts to reconstruct the origin of animal multicellularity have benefited from the study of choanoflagellates, the closest living relatives of animals (King et al., 2008; Leadbeater, 2015; Ruiz-Trillo et al., 2008). Choanoflagellates are microbial eukaryotes that feed on bacteria and live in aquatic environments around the world; many species differentiate over their life history into diverse cell types, including unicellular and multicellular forms (Dayel et al., 2011; Laundon et al., 2019; Leadbeater, 2015, 1983). Comparative genomics has revealed many gene families once thought to be unique to animals (such as cadherins, C-type lectins, and receptor tyrosine kinases) in choanoflagellates (Abedin and King, 2008; King et al., 2008, 2003; Levin et al., 2014; Richter et al., 2018). Moreover, the model choanoflagellate *Salpingoeca rosetta* (Dayel et al., 2011) exhibits diverse responses to environmental cues, including pH-taxis (Miño et al., 2017), aerotaxis (Kirkegaard et al., 2016), and bacterial regulation of multicellular development (Alegado et al., 2012) and mating (Woznica et al., 2017). However, *S. rosetta* is only one of ~380 known species (Carr et al., 2017), and choanoflagellates are at least as genetically diverse as animals (Richter et al., 2018). Choanoflagellate diversity thus represents a largely untapped opportunity to investigate environmental regulation of cell behavior, the principles that broadly

underpin multicellularity, and the evolution of animal cell biology. We report here on the discovery of light-regulated collective cell contractility in a recently isolated multicellular choanoflagellate. This finding reveals that apical cell contractility evolved before the origin of animal multicellularity.

Results and Discussion

Photic cues induce sheet inversion in a colonial choanoflagellate

During a survey of choanoflagellate diversity on the Caribbean island of Curaçao (Fig. 2.1, A and B), we collected large, cup-shaped colonies of protozoa (~100 mm in diameter) from shallow splash pools above the tide line of a rocky coastal area (Fig. 2.1B). Each colony was composed of a monolayer ("sheet") of up to hundreds of flagellated cells (Fig. 2.1C and movie 2S1). The cells bore the characteristic collar complex of choanoflagellates (Brunet and King, 2017; Leadbeater, 2015), in which a "collar" of microvilli surrounds a single apical flagellum (Fig. 2.1D). However, unlike in most choanoflagellate colonies (Leadbeater, 2015), the apical flagella pointed into the interior of the colony (Fig. 2.1E), resembling the orientation of collar cells in the choanocyte chambers of sponges (Nielsen, 2012). The colonies inverted rapidly (within ~30 s from initiation to completion) while maintaining their cell topology, such that the flagella pointed outward along the radius of curvature of the colony (Fig. 2.1, F and G, and movie 2S2). The colonies tended to remain in the inverted form ("flagella-out") for several minutes before reverting to their initial, relaxed conformation ("flagella-in") in a similarly rapid process (Fig. 2.1H and movie 2S3).

To start laboratory cultures of this choanoflagellate, we manually isolated representative colonies away from the other microbial eukaryotes in the splash pool sample (for example, presumptive Oxyrrhis sp. dinoflagellates) (Fig. 2.1C and movie 2S1) and transferred them into nutrient-supplemented artificial seawater along with co-isolated environmental bacteria (which choanoflagellates need as a food source). Analysis of the 18S ribosomal DNA (rDNA) sequence indicated that the choanoflagellate is the sister species of the previously described *Choanoeca perplexa* (also known as *Proterospongia choanojuncta*) (Fig. 2.1J and Fig. 2S1), whose life history includes both single cells and colonies (Leadbeater, 1983, 1977). Inspired by the sheet-bending behavior of the Curaçao choanoflagellate, we named it *Choanoeca flexa*. [A similar behavior was mentioned in a 1983 study of *C. perplexa* (Leadbeater, 1983), but the culture subsequently stopped forming colonies, preventing further study (Leadbeater, 2015).] *C. flexa* sheet inversion, representing a global change in multicellular form, is reminiscent of concerted movement and morphogenesis in animals (such as muscle contraction or gastrulation). Because of the potential evolutionary implications of rapid shape change in *C. flexa*, we investigated (i) how colony inversion is regulated, (ii) the mechanisms underlying colony inversion, and (iii) the ecological consequences of colony inversion.

Several lines of evidence indicated that light levels regulate *C. flexa* colony inversion. While imaging *C. flexa* sheets for >1 hour under constant illumination, colony inversions became less frequent. By contrast, after the microscope illumination was turned off, the colonies inverted almost immediately (movie 2S4). To quantify light-to-dark–induced *C. flexa* inversion, we established an assay based on the observation that the projected area of a *C. flexa* sheet decreases by as much as 50% during inversion (Fig. 2.2, A to C, and movie 2S5). Using this assay, we

confirmed that a rapid decrease in illumination reliably induces inversion of *C. flexa* colonies within 30 s (Fig. 2.2D and movie 2S6). Choanoflagellates had previously been thought to be insensitive to light (Jékely, 2009).

A rhodopsin-cGMP pathway regulates colony inversion in response to light-to-dark transitions

We next investigated how *C. flexa* colonies sense and respond to changing light levels. Although choanoflagellates are unpigmented and transparent, at least four species encode a choanoflagellate-specific rhodopsin-phosphodiesterase fusion protein (Fig. 2S2) (Richter et al., 2018), RhoPDE, that has been investigated as a potential optogenetic tool (Lamarche et al., 2017; Tian et al., 2018; Watari et al., 2019; Yoshida et al., 2017). RhoPDEs consist of an N-terminal type I rhodopsin [a photosensitive transmembrane protein broadly involved in light detection (Spudich, 2006)] fused to a C-terminal phosphodiesterase (PDE) that catalyzes cyclic nucleotide hydrolysis (Fig. 2.3A). In *in vitro* studies (Lamarche et al., 2017; Tian et al., 2018; Watari et al., 2019; Yoshida et al., 2017), *S. rosetta* RhoPDE appears capable of converting a photic stimulus into a biochemical signal within seconds, similar to the time scale of the *C. flexa* response to light-to-dark transitions.

To search for RhoPDE or other candidate photosensitive proteins in *C. flexa*, we sequenced and assembled the *C. flexa* transcriptome (Brunet, 2019), which we found to encode four RhoPDE homologs (Fig. 2S2) (GenBank accession numbers MN013138, MN013139, MN013140, and MN013141). No other rhodopsins were detected in the *C. flexa* transcriptome. The only other candidate photoreceptor protein domain found was a cryptochrome transcription factor (Chaves et al., 2011), which is predicted to act on the time scale of transcriptional regulation [at least several minutes (Coulon et al., 2013; Pérez-Ortín et al., 2007)] and, therefore, is unlikely to mediate the light-to-dark transition response. Thus, we focused our attention on the RhoPDEs.

If RhoPDE regulates the light-to-dark transition response, depletion of the rhodopsin chromophore, retinal (Arendt, 2003; Ernst et al., 2014), should abolish rhodopsin activity and thereby prevent the response. Moreover, artificially increasing the cellular concentration of cyclic guanosine monophosphate (cGMP) or cyclic adenosine monophosphate (cAMP) (which are degraded by the enzymatic activity of PDEs) should mimic the effect of darkness and be sufficient to trigger sheet inversion.

Plants and some bacteria synthesize retinal and other carotenoids, but the *C. flexa* transcriptome lacks a key enzyme in the retinal biosynthesis pathway (Fig. 2S3). Therefore, like animals, *C. flexa* must receive retinal or its precursor, beta-carotene, from its food. It is possible that *C. flexa* acquires retinal or beta-carotene from its bacterial prey, with which it is cultured. To test whether bacterially produced carotenoids are required for light-regulated colony inversion, we established a culture containing only *C. flexa* and a co-isolated bacterium, *Pseudomonas oceani,* that lacks genes in the retinal biosynthesis pathway (Fig. 2S3) (García-Valdés et al., 2018). This culture, named "ChoPs" (for *Choanoeca* + *Pseudomonas*), was expected to be devoid of carotenoids, thereby abolishing rhodopsin activity. As predicted, *C. flexa* sheets in ChoPs cultures did not invert in response to darkness (Fig. 2.3B). Inoculating ChoPs cultures with a mixture of co-isolated environmental bacteria restored the light-to-dark response, demonstrating that a bacterial factor is necessary for the response. Addition of exogenous retinal to ChoPs cultures restored the wild-type light-to-dark response (Fig. 2.3C), indicating that the

absence of retinal explains the absence of colony inversion. The requirement of retinal for the inversion response and the fact that all rhodopsin-encoding genes in the *C. flexa* are RhoPDEs suggest that one (or more) RhoPDE mediates the light-to-dark–induced inversion.

We next investigated whether cyclic nucleotide signaling influences *C. flexa* inversion. Treatment of *C. flexa* sheets with two inhibitors of PDE activity, caffeine(Boswell-Smith et al., 2006) and 3-isobutyl-1-methylxanthine (IBMX) (Weishaar et al., 1985), induced colony inversion in the absence of a photic stimulus (Fig. 2.3D). Moreover, incubating *C. flexa* sheets with a cell-permeant analog of cGMP induced colony inversion in a dose-responsive manner, whereas a cell-permeant analog of cAMP had no effect (Fig. 2.3E), suggesting that cGMP acts as a second messenger in phototransduction and thereby triggers colony inversion.

Together, these results indicate that the *C. flexa* response to light-to-dark transitions relies on a rhodopsin as a photoreceptor and cGMP as a second messenger. The simplest interpretation of these findings is that a RhoPDE controls *C. flexa* phototransduction. However, direct validation will require targeted disruption of all four *C. flexa* RhoPDE homologs (Fig. 2S2).

Sheet inversion mediates a trade-off between feeding and swimming

What are the functional and ecological roles of sheet inversion in *C. flexa*? Flagella-in sheets showed little to no motility, either slowly drifting or settling to the bottom of flasks (movie 2S7), whereas inverted (flagella-out) sheets swam rapidly (Fig. 2.4, A and B; Fig. 2S5; and movie 2S4). By contrast, cells in flagella-in sheets fed efficiently (>75% cells per sheet internalizing beads), whereas cells in flagella-out sheets did not (~10% cells per sheet internalizing beads on average) (Fig. 2.4, C to G). Moreover, in relaxed sheets, fluid flow converged toward the center of the colony (carrying bacterial prey toward the cells), whereas in inverted sheets, the flow was directed away from the colony, allowing swimming but not feeding (Fig. 2S6). Thus, sheet inversion increases motility, which may allow escape from environmental hazards (including predators), whereas sheet relaxation allows enhanced feeding efficiency.

Because sheets swim slowly when relaxed and rapidly when inverted, we suspected that darkness-induced inversion might allow sheets to accumulate in bright areas, effectively undergoing phototaxis. Using chambers illuminated with directional light, we found that *C. flexa* sheets tended to accumulate near the illumination source compared with a control in which no illumination was provided (Fig. 2.4H and Fig. 2S7). Further, neither sheets from ChoPs cultures nor single cells from dissociated sheets were capable of phototaxis, suggesting that rhodopsin activity, multicellularity, and sheet inversion are all required for phototaxis (Fig. 2.4H). Thus, sheet inversion mediates an ecologically relevant trade-off between feeding and swimming (Fig. 2.4I).

Sheet inversion requires apical actomyosin contractility

How do cells in sheets interact, and what mechanisms allow sheet inversion? We found that cells in *C. flexa* sheets form direct contacts between their collar microvilli (Fig. 2.5, A to C, and Fig. 2S8) but not through the intercellular bridges, shared extracellular matrix, or filopodial contacts that mediate multicellularity in other choanoflagellate species (Dayel et al., 2011; Leadbeater, 1983; Levin et al., 2014; Wetzel et al., 2018). Collar morphology differs between

relaxed and inverted sheets. In relaxed sheets (flagella-in; Fig. 2.5C), the microvilli on each cell assemble into a barrel-shaped collar whose diameter varied little from base to tip. In inverted sheets (flagella-out; Fig. 2.5C), the microvilli form a flared, cone-shaped collar whose diameter increases from base to tip. Active "opening out" of the collar, by increasing the surface area of the apical side of the sheets relative to their basal side, might force a change in sheet curvature. Consistent with this, the collar of dissociated *C. flexa* cells treated with caffeine opened into a conical shape, straightened, and slid down toward the base of the cell, whereas untreated controls maintained a barrel-shaped collar (Fig. 2.5, D to G, and Fig. 2S9). These data suggest that the changes in collar geometry during inversion are actively generated by individual cells and do not require interactions among neighboring cells.

Sheet bending in animal epithelia is frequently mediated by apical constriction, in which contraction of an apical actomyosin network reduces the surface area of the apical side of the cell (Martin and Goldstein, 2014; Sawyer et al., 2010). Apical constriction is mediated by molecular motors belonging to the myosin II family, which predates the diversification of modern eukaryotes (Richards and Cavalier-Smith, 2005) and is found in all choanoflagellate genomes (Sebé-Pedrós et al., 2014) and transcriptomes (Richter et al., 2018) published to date. *C. flexa* encodes homologs of the myosin II regulatory light chain (GenBank accession MK787241) and heavy chain (GenBank accession MK787240) (Fig. 2S4), whose protein sequences are 78 and 63% similar, respectively, to their human counterparts.

In choanoflagellates and animal epithelial cells, the apical pole is defined by the presence of a flagellum or cilium, respectively, and/or microvilli, and the apicobasal axis of both types of cells is broadly considered to be homologous (Richter and King, 2013). *C. flexa* sheets contain a pronounced F-actin ring at the apical pole of each cell, from which the microvillar collar extends (Fig. 2.5H). Extending from this ring, we detected a small number of longitudinal actin fibers (usually two or three) pointing toward the basal pole. Diameter measurements showed that the F-actin ring was consistently smaller in inverted sheets compared with relaxed sheets (Fig. 2.5, I to L). The same was true for dissociated, caffeine-treated cells compared with the corresponding negative controls (Fig. 2.5, I to J and L), consistent with the ring actively constricting during sheet inversion. During inversion and in response to caffeine, some (but not all) cells transiently acquired a "bottle cell" morphology with a narrow apex and a bulbous base (Fig. 2S9), reminiscent of animal cells undergoing pronounced apical constriction (Lee and Harland, 2007; Magie et al., 2007). Caffeine treatment also induced shortening of the longitudinal actin fibers (Fig. 2S9, G to H), suggesting that fiber contraction pulls the collar toward the basal pole.

Using an antibody raised against *C. flexa* myosin II as well as five different commercial myosin II antibodies (Fig. 2.5, M to S, and Fig. 2S10), we found that *C. flexa* cells contain myosin that overlaps in regions with the apical actin ring (Fig. 2.5, M to P, and Fig. 2S10) and longitudinal fibers (Fig. 2.5, Q to S), consistent with the idea that the apical actin network is contractile. Inhibition of myosin II activity with blebbistatin (Kovács et al., 2004) abolished ring constriction in caffeine-treated dissociated cells (Fig. 2S11A) and prevented sheet inversion (Fig. 2.5T), as did inhibition of dynamic actin polymerization with latrunculin B (Wakatsuki et al., 2001) and inhibition of phosphorylation of the myosin regulatory light chain with ML-7 (Saitoh et al., 1987) (Fig. 2.5T). None of these drugs affected flagellar beating, consistent with them specifically targeting actomyosin. Although compaction is normally associated with inversion, drug treatment resulted in sheet compaction in the absence of inversion (Fig. 2S11B), suggesting that a baseline level of tension is needed to maintain spacing between cells. Together, these

results suggest that sheet inversion requires apical constriction of an actomyosin network at the base of the collar (Fig. 2.5U).

The ancestry of apical constriction

The discovery of sheet bending driven by apical constriction in a multicellular choanoflagellate has several potentially important evolutionary implications. Epithelial sheet bending is a fundamental mechanism underlying animal embryonic development (Darren Gilmour et al., 2017; Heisenberg and Bellaïche, 2013; Martin and Goldstein, 2014), and multicellular contractility also plays a fundamental role in the behavior of adult animals by allowing fine-tuned body deformations (Marieb and Hoehn, 2015). As both embryonic and adult tissue contractility are found in nearly all animal lineages, including sponges (Leys and Eerkes-Medrano, 2005b; Nickel et al., 2011), ctenophores (Dayraud et al., 2012; Pang and Martindale, 2008), placozoans (Armon et al., 2018), cnidarians (Magie et al., 2007; Steinmetz et al., 2012), and bilaterians (Brunet et al., 2016; Martin and Goldstein, 2014; Sawyer et al., 2010), both were likely present in the last common animal ancestor. By contrast, collective contractility and apical constriction were hitherto unknown in close relatives of animals, making their origin mysterious.

The existence of actomyosin-mediated apical constriction in *C. flexa* raises the possibility that this cellular module might have been present in the last common ancestor of choanoflagellates and animals [which together compose the choanozoans (Brunet and King, 2017)]. In addition to *C. flexa*, in which collar contractions occur in sheets (Fig. 2.5G), dissociated cells from sheets (Fig. 2.5, D and E), and naturally solitary "thecate" cells (movie 2S8), collar contractions have also been reported in unicellular stages from three other choanoflagellates: *Codosiga pulcherrima* (James-Clark, 1867), *Monosiga gracilis* (Kent, 1878), and *C. perplexa* (Leadbeater, 1983). We found that four other choanoflagellate species, *Monosiga brevicollis*, *S. rosetta*, *Salpingoeca urceolata*, and *Diaphanoeca grandis,* which together cover both main branches of the choanoflagellate phylogenetic tree (Carr et al., 2017), display spontaneous changes in collar geometry occurring at the scale of a few seconds. *S. urceolata* (movie 2S9) and *M. brevicollis* (movie 2S10) showed spontaneous and reversible opening and closing of the collar (similar to *C. flexa*), whereas *S. rosetta* (movie 2S11) and *D. grandis* (movie 2S12) displayed subtler shape changes (reorientation of individual microvilli and modulation of collar curvature, respectively) (Fig. 2.6A and Fig. 2S12). Like animal epithelial cells and *C. flexa*, all four species have an apical actomyosin ring at the base of the collar (Fig. 2.6B). This suggests that the apical actomyosin ring is a conserved feature of choanoflagellate biology (Fig. 2.6C) and that unicellular apical constriction was present in the last common ancestor of choanoflagellates and animals.

What is the function of apical constriction in single cells? In some sessile choanoflagellates, including in the thecate form of *C. perplexa*, collar contraction in response to mechanical stimulation allows retraction of the cell inside an extracellular structure called a theca (Leadbeater, 1977), suggesting it represents a defensive withdrawal reflex from predators or other threats. In free-swimming cells, collar contraction might fine-tune the hydrodynamics of swimming or feeding. For example, a closed collar might reduce drag and facilitate locomotion, whereas a flared collar could slow down swimming and increase collar area, thereby facilitating prey capture. Validation of these functional hypotheses will require direct testing.

In contrast to single-cell apical constriction, the multicellular sheet bending observed in *C. flexa* and *C. perplexa* (Leadbeater, 1983) has not been reported in other choanoflagellates. This suggests that apical constriction was present in solitary cells in the last choanozoan common ancestor and was independently converted into multicellular sheet bending through the evolution of intercellular junctions in animals(Nichols et al., 2012) and the evolution of microvillar adhesions in *C. flexa*. Multicellular inversion has been proposed to have been part of the developmental repertoire of ancient animals (Arendt et al., 2015; Brunet and King, 2017), on the basis of the existence of whole-embryo inversion (from flagella-in to flagella-out) during calcareous sponge development (Franzen, 1988). A similar inversion (but much slower, about an hour long, and irreversible) takes place during the development of the alga *Volvox* (Höhn et al., 2015). Given the large evolutionary distance between choanoflagellates and volvocalean green algae, along with the absence of inversion in intervening branches, inversion likely evolved independently in both groups (Brunet and King, 2017).

These observations suggest that apical actomyosin-mediated cell constriction evolved on the choanozoan stem lineage (Fig. 2.6B). Could polarized actomyosin contractility be even more ancient? Polarized actomyosin contractions have been implicated in multicellular morphogenesis in the fruiting body of the slime mold *Dictyostelium* (Dickinson et al., 2012b) and may be homologous to those observed in choanoflagellates and animals. However, the absence of comparable processes in the intermediate branches between *Dictyostelium* and choanozoans raises the possibility that polarized cell contractions in *Dictyostelium* and apical constriction in choanozoans evolved independently (Dickinson et al., 2012a). The ichthyosporean *Sphaeroforma arctica*, a close relative of choanozoans, forms large multinucleated spores that partition into distinct cells in an actomyosin-dependent process (Dudin et al., 2019), providing an independent example of actomyosin-dependent multicellular development.

In animals, the control of multicellular contractions invariably relies either on the cooperation of multiple cell types [as in adult organisms (Brunet et al., 2016; Steinmetz et al., 2012; Varoqueaux et al., 2018)] or on complex programmed signaling cascades [as in embryos (Bailles et al., 2019; Darren Gilmour et al., 2017; Heisenberg and Bellaïche, 2013; Martin and Goldstein, 2014; Merle and Farge, 2018)]. By contrast, *C. flexa* directly converts sensory stimuli into collective contractions, without observable spatial cell differentiation, and evokes some hypotheses of early animal evolution that envisioned the first contractile tissues as homogeneous myoepithelia of multifunctional sensory-contractile cells (Mackie, 1970).

The fact that contractility in *C. flexa* can be controlled by light represents another intriguing parallel to animal biology. Indeed, rhodopsin-cGMP pathways similar to that in *C. flexa* also underlie phototransduction in some animal cells [for example, bilaterian ciliary photoreceptors (García-Valdés et al., 2018; Zhang and Cote, 2005) and cnidarian photoreceptors (Koyanagi et al., 2008; Plachetzki et al., 2010)], as well as in fungal zoospores (Avelar et al., 2014). In contrast with choanoflagellates, however, phototransduction in animal photoreceptors relies on a type II (eukaryotic) rhodopsin that activates a separate PDE through a G-protein intermediary (Arendt, 2003; Zhang and Cote, 2005) (Fig. 2S13). Meanwhile, fungal zoospores use a distinct rhodopsin fusion protein (a type I rhodopsin fused to a guanylyl cyclase) to increase cellular cGMP in response to light (Avelar et al., 2014) (Fig. 2S13). If a RhoPDE fusion protein controls *C. flexa* phototransduction, this would represent a third independent solution to the problem of transducing information from a change in illumination into a change in cyclic nucleotide signaling.

Much remains to be discovered concerning the ecological function, mechanical underpinnings, and molecular mechanisms of phototransduction and apical constriction in *C. flexa*. A deeper understanding will require the development of molecular genetic tools, which have only recently been established in *S. rosetta* (Booth et al., 2018; Wetzel et al., 2018) and *D. grandis* (Li et al., 2018). Nonetheless, *C. flexa* demonstrates how the exploration of choanoflagellate diversity can reveal biological phenomena and provides an experimentally tractable model for studying multicellular sensory-contractile coupling.

Materials and Methods

Species description

Order Craspedida Cavalier-Smith 1997 (Cavalier-Smith, 1997; Ellis, 1930)
Genus *Choanoeca* Ellis 1930 (Ellis, 1930)
Genus *Choanoeca* Ellis 1930 (Ellis, 1930)
Choanoeca flexa sp. nov. Brunet, Larson, Linden et King

Etymology: from the Latin 'flexa', nominative feminine singular of 'flexus' which means 'bending' or 'transition, change'.
Type material: The name-bearing hapantotype consists of fixed sheet specimens that were sputter-coated and mounted on stubs for scanning electron microscopy. This material is deposited with the California Academy of Sciences Invertebrate Zoology Collections (San Francisco) with accession number CASIZ 197968. This material also contains bacteria belonging to the species *Pseudomonas oceani*, which are explicitly excluded from the hapantotype.
Description: cell body is about 4-5 μm long. Microvillous collar is about 10 μm long. Microvilli of resting cells appear markedly curved. Cells form two-dimensional curved sheets linked by direct contact between their collars. Resting sheets are cup-shaped hemispheres with flagella pointing on the inner side of the curvature. Cells in their thecate form (Movie 2S8) are characterized by a theca with a short stalk, a bulbous cell base, a narrow and pointed apex, and a widely spread collar without a flagellum, as in the thecate form of *Choanoeca perplexa* (Leadbeater, 1977).
Type locality: splash pool of the Curaçao rocky coast (12°13'38.9" N 69°00'47.0" W), from which water was collected in April 2018.

Initial isolation and culture of *Choanoeca flexa*

Samples containing *Choanoeca flexa* were first isolated by the three co-first authors from multiple splash pools on the northern rocky coast of Curaçao (12°13'38.9" N 69°00'47.0" W). Individual colonies were manually isolated by pipetting 0.1 μL with a P2 micropipette under a Leica DMIL transmitted light microscope. Isolation was confirmed by pipetting the resulting droplet onto a microscopy slide and visually confirming the presence of single colonies, which were then transferred into 24-well plates containing 1 mL culture medium. Each well was seeded with 3 or 4 individual colonies. Wells were visually inspected 24 to 48 hours after transfer, and those containing colonies were transferred into a T25 culture flask containing 10 mL culture medium. Sheets were cultured in 1% to 10% Cereal Grass Medium (hereafter CGM3) in artificial seawater (hereafter ASW; as in (King et al., 2009a)). Cultures were maintained at 22°C under a light-dark cycle (12:12 hours) in a VWR Scientific model 2005 low temperature

incubator equipped with a lamp (Venoya Full Spectrum 150W Plant Growth LED) controlled by a programmable timer (Leviton VPT24-1PZ Vizia). For long-term storage, *C. flexa* cultures were frozen and kept in a liquid nitrogen dewar following (King et al., 2009b). Cells could be successfully revived by thawing from frozen stocks (King et al., 2009b).

Establishment and propagation of the ChoPs (*Choanoeca-Pseudomonas*) monoxenic strain

We define the ChoPs strain (Fig. 2.3B) as a monoxenic culture of *C. flexa* with the co-isolated bacterium *Pseudomonas oceani* (*P. oceani*). ChoPs was established by elimination of all environmental bacteria from the original polyxenic isolate by antiobiotic treatment, followed by re-seeding with *P. oceani*. More specifically, a polyxenic isolate was treated with a combination of 6 antibiotics to prevent growth of environmental bacteria: carbenicillin (100 µg/mL), erythromycin (160 µg/mL), kanamycin (100 µg/mL), lincomycin (200 µg/mL), rifampicin (20 µg/mL), and streptamycin (200 µg/mL).

To confirm that the antibiotic combination prevented proliferation of all co-isolated bacteria, we inoculated 100 µL of the original polyxenic culture, with or without antibiotics, into 5 mL 1% CMG3 with 20 µg/mL cycloheximide (to prevent eukaryotic growth). Cells were grown in a bacterial incubator at 22°C under 300 rpm agitation and cell growth was monitored by measuring 600 nm optical density (Fig. 2S14).

To isolate *P. oceani*, 70 µL of the original polyxenic strain were spread on a 4% agar plate supplemented with CGM3 culture medium. A single colony was picked, grown in 5 mL 100% CGM3 for 3 days, and identified by 16S sequencing as *P. oceani*. We then passaged the antibiotic-treated *C. flexa* polyxenic line 5 times at 1:10 dilution in the 6-antibiotic combination, re-seeding every time with 1:20 volume of *P. oceani* culture (to progressively dilute out all other bacterial species, and feed choanoflagellates with a constant supply of pre-grown bacteria). After 5 passages, antibiotic treatment was replaced by treatment with 20 µg/mL rifampicin only (to which *P. oceani* is resistant) and active re-seeding with novel *P. oceani* was stopped.

Monoxenicity of the resulting ChoPs line was validated by PCR amplification of the 16S locus followed by sequencing of 50 individual clones, as well as by iTag sequencing (Table 2S1 and section "Characterization of bacterial communities by iTag sequencing"). ChoPs was cultured in the constant presence of rifampicin to prevent contamination.

C. flexa DNA extraction, rDNA locus amplification and phylogenetic analysis

C. flexa was grown to 10^6 cells/mL in 200 mL culture medium in 6-layer culture flasks, and pelleted by centrifugation for 15 minutes at 2,000 rpm at 4°C. DNA was extracted using the DNEasy Blood & Tissue Kit (Qiagen 69504). The 18S locus was amplified by PCR with Q5 DNA Polymerase (New Englands Biolab M0491L) and degenerate primers (sequences in Table 2S2).

The following PCR programs were used: (1) for the first round of PCR: 1. 98 °C, 2 minutes; 2. 98 °C, 10 seconds; 3. 48 °C, 30 seconds; 4. 55 °C, 30 seconds; 5. 72 °C, 3 minutes; 6. repeat steps 2-5, 30x total; 7. 72 °C, 11 minutes 11 seconds; (2) for the nested PCR: 1. 98 °C, 2 minutes; 2. 98 °C, 10 seconds; 3. 48 °C, 30 seconds; 4. 52 °C, 30 seconds; 5. 72 °C, 1 minute; 6. repeat steps 2-5, 35x total; 7. 72 °C, 7 minutes 21 seconds.

rDNA sequences from *C. flexa* and other opisthokonts were aligned with ClustalX 2.0 (Larkin et al., 2007), trimmed with gBlocks (http://phylogeny.lirmm.fr/phylo_cgi/one_task.cgi?task_type=gblocks) under minimally stringent parameters, and a Maximum Likelihood phylogenetic tree was produced with SeaView

v. 4.7 (Gouy et al., 2010) using PhyML with the GTR model, empirical nucleotide equilibrium frequencies, optimized proportion of invariable sites, optimal rate variation across sites, and best of NNI & SPR tree search method. Trees were visualized and edited with FigTree v. 1.4.4 (http://tree.bio.ed.ac.uk) and further edited with Adobe Illustrator CC 2018.

C. flexa RNA extraction, transcriptome sequencing and assembly

C. flexa was grown to 10^6 cells/mL in 200 mL culture medium in 6-layer culture flasks and pelleted by centrifugation for 15 minutes at 2,000 rpm at 4°C. Cells were lysed by selective lysis as in (Booth et al., 2018) and RNA was extracted using a RNEasy kit (Qiagen 7404). 150 paired-end RNAseq libraries were prepared after poly-A selection by the QB3 Functional Genomics Laboratory at UC Berkeley with poly-A selection and sequenced on a HiSeq 4000 sequencer (Illumina, San Diego, California, United States) at the Vincent J. Coates Genomics Sequencing Laboratory at the California Institute for Quantitative Biosciences (Berkeley, California, United States). Read quality was assessed with FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and transcriptome assembly was performed as in (Richter et al., 2018) with Trinity v. 2.5.1, with the '--trimmomatic' option and '- min_contig_length' set to 150. Predicted protein sequences were generated with Transdecoder (Haas et al., 2013) with a predicted minimum protein sequence length of 50 aminoacids and redundant protein sequences were eliminated using CD-HIT (Li and Godzik, 2006). The reads generated were uploaded onto the NCBI website with the BioSample accession number SAMN11533889 and BioProject ID/SRA accession number PRJNA540068. Transcriptome and non-redundant predicted proteome are available on Figshare at DOI: 10.6084/m9.figshare.8216291. Phylogenetic trees for proteins of interest were generated following the same procedure as for rDNA (see above '*C. flexa* genome extraction, rDNA locus amplification and phylogenetic analysis') with aminoacid sequences instead of nucleotide sequences.

Characterization of bacterial communities by iTag sequencing

A *C. flexa* culture was pelleted at 4500xg and bacterial DNA was extracted using a DNeasy Kit (Qiagen, Hilden, Germany) following the manufacturer's Gram-Positive Bacteria Protocol. Library construction and sequencing were then performed by the UC Berkeley Functional Genomics Laboratory. Amplicon library construction was carried out in two distinct PCR steps: briefly, PCR1 used modified gene-specific primers to create amplification products of the V4 region of the bacterial 16S gene flanked by stub sequences (to provide priming templates for PCR2); PCR2 primers added dual-matched index sequences, sequencing primer binding sites, and Illumina p5 and p7 adapter sequences to the 5' and 3' ends of each amplicon. PCR1 was carried out in 25 µl reactions including 1 µl genomic DNA, 0.5 µl of each 10 µM primer, 10 µl 2.5x 5PRIME HotMasterMix (QuantaBio, Beverly, MA), 1µl BSA (New England Biolabs, Ipswitch, MA) (to a final concentration of 10µg/µL) and 12 µl nuclease free water. We amplified the 16S V4 hypervariable region using the primer set 515f from Parada et. al (Parada et al., 2016) (GTG**Y**CAGCMGCCGCGGTAA) and 806r from Apprill et al. (2015) (GGACTAC**N**VGGGTWTCTAAT). All PCR reactions were set up on ice and using Hot Start polymerase master mix to minimize non-specific amplification and primer dimerization. PCR conditions were: denaturation at 94°C for 3 mins; 30 amplification cycles of 45 sec at 94°C, 1 min at 50°C and 90 sec at 72°C; followed by a 10 min final extension at 72°C. PCR products were visualized for successful amplification and correct sizing using gel electrophoresis.

PCR2 was carried out in 25 µl reactions including 5 µl genomic DNA, 2 µl of combined, µM forward and reverse indexing primers, 10 µl 2.5x 5PRIME HotMasterMix (QuantaBio, Beverly, MA), and 8 µl nuclease free water. PCR2 primers added dual-matched 8bp unique barcodes to each end of the amplicons, such that each forward and reverse primer pair carried the same index sequence. PCR conditions were: denaturation at 94°C for 3 mins; 8 amplification cycles of 45 sec at 94°C, 1 min at 52°C and 90 sec at 72°C; followed by a 10 min final extension at 72°C. PCR products were visualized and quantified using an Advanced Analytical Fragment Analyzer (Agilent, Santa Clara, CA), and individual samples were pooled equimolarly based on Fragment Analyzer concentrations.

The final pool of all individually indexed amplicon libraries was cleaned with Agencourt AMPure XP magnetic beads using a 0.8X bead ratio. The cleaned pool was quantified with qPCR using the KAPA Illumina Library Quant Kit and Universal qPCR Mix (KAPA Biosystems, Wilmington, MA). Amplicon libraries were sequenced on an Illumina MiSeq 300PE v3 run spiked with 10% PhiX in order to achieve sufficient sample heterogeneity.

Light microscopy

Sheets were imaged in FluoroDishes (World Precision Instruments FD35-100) by differential interference contrast (DIC) microscopy using a 40x (water immersion, C-Apochromat, 1.1 NA), 63x (oil immersion, Plan-Apochromat, 1.4 NA), or 100x (oil immersion, Plan-Apochromat, 1.4 NA) Zeiss objective mounted on a Zeiss Observer Z.1 with a Hamamatsu Orca Flash 4.0 V2 CMOS camera (C11440-22CU).

Transmission electron microscopy (TEM)

Sheets were concentrated by centrifugation (200xg for 5 min) and then resuspended in 5% BSA in artificial seawater. The resuspended sheets were then high pressure frozen using a Leica EM PACT2 and fixed by freeze substitution in 0.01% OsO_4 + 0.2% uranyl acetate in acetone. Samples were resin embedded in Epon Araldite (McDonald, 2014), cut into 80 nm sections, and then imaged using an FEI Tecnai 12 transmission electron microscope.

Scanning electron microscopy (SEM)

C. flexa sheets were concentrated by pelleting 6 mL of culture for 15 minutes at 200 g and gently resuspending the pellet in 600 µL final volume. Sheets were then pipetted onto silicon wafers coated with poly-D-lysine (Sigma Aldrich P6407-5MG), fixed for 2 hours in 2% glutaraldehyde in 0.1M Sodium cacodylate buffer pH 7.2, rinsed 3 times (for 15 minutes each) in 0.1M sodium cacodylate buffer pH 7.2 (for 15 minutes each), post-fixed for 2 hours in 1% osmium tetroxide in 0.1M sodium cacodylate buffer pH 7.2, and rinsed 3 times (for 5 minutes each) in 0.1M sodium cacodylate buffer, pH 7.2. Samples were dehydrated in the following steps: 35% ETOH (5 min), 50% ETOH (5 min), 70% ETOH (5 min), 80% ETOH (10 min), 95% ETOH (10 min), 100% ETOH (10 min) and 100% ETOH (10 min). Samples were critical point dried for 60 minutes on a Tousimis AutoSamdri 815 Critical Point Dryer, mounted on stubs using conductive carbon tape, and sputter coated before imaging on a Hitachi S-5000 Scanning electron microscope.

Antibody design and validation

Four polyclonal antibodies were raised in rabbits against *C. flexa* myosin II (two against recombinant aminoacids 1-377 of the heavy chain (Cf-MHC) and two against the recombinant

full-length light chain sequence (Cf-MRLC)) and affinity-purified by the manufacturer (GenScript, Piscataway, New Jersey, USA). Antibodies were validated by Western blot on protein extracts from *C. flexa* monoxenic ChoPs cultures, as well as on protein extracts from the bacterial species *Pseudomonas oceani* (co-cultured with *C. flexa* in the ChoPs line). Western blots were generated using the following protocol: protein extracts were run on a 4-20% Tris glycine gel (Invitrogen), transferred into a 0.2 µm nitrocellulose membrane (Bio-Rad), blocked with Odyssey Block (Licor), stained with the antibody of interest at 0.37 µg/mL, stained with a Licor secondary antibody 800CW goat anti-rabbit, and imaged with a Licor Odyssey Imager. Choanoflagellates were lysed with a selective lysis protocol adapted from (Booth et al., 2018). All four antibodies recognized a band of the predicted size, but three of them also stained additional non-specific bands, with only one (raised against Cf-MHC) giving a fully specific Western Blot staining (Fig. 2S10B). For this reason, this antibody was chosen for immunofluorescence. Immunostainings were performed with 1.23 µg/mL anti-Cf-MHC antibody according to the protocol in the following section ("Sheet fixation and FM143-FX/phalloidin/immunofluorescence stainings") with the following modification: fixation was performed for 15 minutes (instead of 2 hours) at room temperature, as longer fixation times resulted in weaker staining (possibly due to epitope degradation after extended fixation).

Sheet fixation and FM143-FX/phalloidin/immunofluorescence stainings

FluoroDishes (World Precision Instruments FD35-100) were pre-treated with a handheld Corona surface treater (Electro-Technic Products BD-20AC), coated with poly-D-lysine (Sigma Aldrich P6407-5MG) diluted to the provider's specifications, and rinsed twice with ASW. Sheet colonies were then transferred into the treated dishes and left to adhere to the bottom surface for 30 minutes. Fixation was performed by adding 1:3 volume ice-cold 16% paraformaldehyde (PFA; Fischer Scientific 50-980-487) to the FluoroDish (to a final concentration of 4%) for 2 hours at room temperature (or for 15 minutes for immunofluorescence with the anti-Cf-MHC antibody; see above). Following steps followed one of the three specific stainings protocol below (FM143-FX, phalloidin or immunofluorescence).

FM143-FX staining: 5 µg/mL FM143-FX (ThermoFischer Scientific F35355) was included together with PFA directly in the fixation solution. This procedure was chosen because FM143-FX was observed to quickly trigger dissociation of live sheets, but to efficiently stain samples if included during fixation. The fixation solution was then washed out three times carefully with ASW and the samples were imaged directly in FluoroDishes. 3D renderings (Fig. 2.1E and F) were generated with Imaris v3.8 (Bitplane, Belfast).

Phalloidin staining: fixed sheets were washed 3 times with ASW. Cells were then stained with 0.66 units/mL Alexa 488-phalloidin (ThermoFischer Scientific A12379) or rhodamine-phalloidin (ThermoFischer Scientific R415) in 0.3% Triton X/ASW overnight under agitation at 4°C. The fixation solution was then washed out three times carefully with ASW and the samples were imaged directly in FluoroDishes.

Immunofluorescence: samples were blocked for 30 minutes at room temperature with blocking solution (1% bovine serum albumin in PEM buffer (100 mM PIPES pH 6.9, 1 mM EGTA, 1 mM MgSO4)/0.3% Triton-X), and then stained with antibodies diluted in blocking buffer (see below).

We selected commercial antibodies against myosin II (full-length or heavy chain) that were reported to bind both smooth and striated myosin isoforms, based on the following reasoning: as the divergence between both myosin paralogs predates the choanoflagellate/animal

divergence (choanoflagellates have only retained the smooth isoform; Fig. 2S4B), we expected any antibody that recognizes both smooth and striated myosin of animals to be likely to bind to the choanoflagellate myosin heavy chain as well. Additionally, we investigated antibodies targeting the myosin regulatory light chain (either in all its forms or in its phosphorylated form only), motivated by its high level of sequence conservation (>80% identity between *C. flexa* and *H. sapiens*) and by conservation of the Ser19 phosphorylation residue. The antibodies tested are listed in Table 2S3.

Immunostainings in choanoflagellate species other than *C. flexa* (Fig. 2.6B) were generated using the Sigma Aldrich M7648 antibody (*D. grandis*, *S. urceolata* and *M. brevicollis*) or the Developmental Hybridoma Bank CMII23 antibody (*S. rosetta*).

Primary antibodies were diluted in blocking solution to the manufacturer's specifications. Primary antibody incubation was performed overnight at 4°C under agitation. Samples were then washed three times in PEM buffer, incubated with secondary antibodies diluted 1:300 in blocking solution (Goat anti-Mouse IgG1 Cross-Adsorbed Secondary Antibody Alexa Fluor 488; ThermoFischer Scientific A-21121) together with 0.66 units/mL rhodamine-phalloidin for 2 hours at room temperature, washed three times with PEM, and imaged directly in FluoroDishes. All samples were imaged using a Zeiss LSM 880 AxioExaminer with Airyscan and a 63x, 1.4 NA C Apo oil immersion objective (Zeiss) and excitation provided by a 488 nm laser (Zeiss).

Light-induced sheet inversion assays

Light-controlled inversion assays were performed in 24-well plates (Fischer Scientific 09-761-146) containing 2 mL sheet culture per well and imaged with a Leica DMIL LED transmitted light microscope coupled to a Leica DFC 350FX camera. Images were acquired with the MicroManager software at a frame rate of 1 image/second. The "light on" condition corresponded to maximal illumination by the microscope lamp and imaging with 1 ms exposure time. The histogram of detected light intensities was monitored with MicroManager and was found, under these conditions, to be a narrow bell curve centered around 50% light intensity. The "light off" condition was established by raising the exposure time to 400 ms and manually decreasing light intensity until the histogram of detected light intensities closely matched the one of the "light on" condition (sharp peak at 50%) – thus indicating that light intensity had been reduced down to close to 1:400 of its initial value. Each sheet was imaged for 120 seconds (55 to 60 seconds with the light off and 60 to 65 seconds after the light was switched off). The light reduction procedure itself lasted about 1 second (resulting in one white slice (see Movie 2S6) which was excluded from the area quantification pipeline – see below). This protocol was found to reliably induce inversion of polyxenic sheets during the sheets' subjective day.
While sheets in laboratory cultures significantly vary in size (ranging from about 20 to more than 500 cells), sheets of variable sizes displayed no evident difference in their inversion behavior, as shown by Movie 2S13 (showing inversion of a 24-cell sheet) and Movie 2S14 (showing inversion of a >565-cell sheet.)

Sheet area quantification and image analysis

Sheet projected area was quantified in ImageJ 1.46r. Stacks produced by live imaging of sheets undergoing light-induced inversion were cropped around the sheet of interest, split into 2 parts (before and after switching off the light, to correct for minimal differences in exposure), and processed with the "Make Binary" command followed by 2 to 5 iterations of the "Close" command and "Dilate" command in alternation. Iterations were stopped when the resulting shape

contained no gaps, and were performed an identical number of times for both parts of each movie. Areas were quantified by measuring the mean gray value on each slice with the measureStack plugin (http://www.optinav.info/MeasureStack.htm). Resulting measurements were stored in .xls files and combined into a single table using a Python or R script [citation TK – see SOM]. Area=f(time) curves were plotted using the ggplot2 package in R studio (https://ggplot2.tidyverse.org/). Measurements were smoothened using a rolling average over a 5-second time window. For each individual sheet, areas were normalized by either (1) for the "light on" time window, the time average of areas measured over the entire window or (2) the initial area (for the "light off" time window).

Drug treatments and all-trans-retinal rescue assays

Drugs used together with stock and working concentrations are listed in Table 2S4. Sheets were pre-treated with actomyosin inhibitors for 30 minutes before behavioral assays. Y-27632 did not observably affect inversion. 30-minute treatment with actomyosin inhibitors tended to result in a more compact sheet morphology (Fig. 2S11A), suggesting that some baseline level of actomyosin tension is necessary to maintain proper collar angle and consequently proper spacing between cells within sheets. Treatment did not otherwise visibly affect sheet structure or conformation.

Caffeine, IBMX, and 8-Br-cGMP were observed to induce inversion in a few seconds after addition and active mixing by gentle swirling of the culture. For quantification of the number of inverted sheets, sheets were fixed by addition of 1:3 volume ice-cold 16% PFA, and manually counted under a Leica DMIL LED transmitted light microscope.

Bacterial rescue protocol

Isolation of bacteria from monoxenic and polyxenic cultures: To isolate bacteria from the monoxenic ChoPs culture *(P. oceani)* or polyxenic culture ("env. bacteria"), 20 µl of either culture was inoculated into 100% CGM3 with 20 µg/ml cycloheximide (to inhibit eukaryotic growth). Cultures were grown shaking at 22 degrees for 48 hours. Bacteria were then pelleted at 20,000xg, washed with ASW, and resuspended to 1 O.D.600/ml in ASW.

Pre-treatment of recipient culture: To reduce bacterial load, the recipient ChoPs culture was treated for 24 hours with an antibiotic cocktail (20 µg/ml rifampicin, 200 µg/ml streptomycin, 100 µg/ml kanamycin, 160 µg/ml erythromycin, 100 µg/ml carbenicillin, and 200 µg/ml lincomycin). After 24 hours, choanoflagellates were pelleted at 1000xg, washed once in ASW, then resuspended in 10%CGM3.

Bacterial rescue: Pre-treated Chops cultures were seeded into 10% CGM3 at 4E4 cells/ml; isolated bacteria (either *P. oceani* or "env. bacteria") were diluted into the culture at 1:1000 and grown for 48 hours. Cultures that received *P. oceani* were subsequently treated with 20ng/ul rifampicin for one week to ensure absence of contaminants. Photic response was then quantified as described below (see "Light-induced sheet inversions assays").

Retinal rescue protocol

Monoxenic ChoPs culture was grown in 10% CGM3 + 20 ng/µl rifampicin + 0 nM, 125 nM or 500 nM all-trans-retinal. After one week treatment, photic response was quantified as described above (see "Light-induced sheet inversion assays").

Sheet dissociation

A dense (1E6 cells/mL) sheet culture was reconcentrated 10x by centrifugation (2,000 rpm for 15 min at 4°C) and dissociated by vortexing for 30 seconds followed by filtration through a Tisch Scientific syringe-top 5 μm filter (SPEC18191). Dissociated cells were immediately transferred into an Ibidi 8-well plate imaging chamber (80826) coated with poly-D-lysine and washed twice with ASW (200 μL of cell suspension per chamber).

Quantification of actin ring diameter

Imaging of phalloidin labeled samples (both single cells and intact sheets) for actin ring quantification was performed as detailed in the "Sheet fixation and FM143-FX/phalloidin/immunofluorescence stainings" section. Actin ring quantification for both intact and dissociated sheets was performed using Imaris v3.8 (Bitplane, Belfast). Images were first segmented with a global threshold chosen manually for each condition on one image and then held constant for all subsequent images. Diameters were measured using the "Measure Points" function, with three diameters manually chosen and the longest measured distance chosen as the ring diameter. Only completely intact rings with associated intact collars were chosen for analysis.

Sheet tracking

Sheets were imaged on a Zeiss Axio Zoom.V16 (generously lent by Zeiss to the 2018 Physiology course at the Marine Biology Laboratory in Woods Hole). Individual sheets were tracked on Fiji v. 2.0 (Schindelin et al., 2012) using the Manual Tracking option of the Tracking plugin.

Movement quantification

Movies of sheets swimming were produced on a Leica DMIL LED transmitted light microscope coupled to a Leica DFC 350FX camera, and movement was quantified by measuring inter-frame Pearson correlation using RStudio (RStudio Team, 2020). Sheet inversion was induced by a rapid (ca. 1 sec transition time) reduction in light intensity (see "Light-induced sheet inversion" section).

Microbead feeding assays, flow visualization and quantification

0.2 μm green fluorescent microbeads were used for both feeding assays and flow visualization (FluoSpheresTM Carboxylate-Modified Microspheres, 0.2 μm, red fluorescent (580/605), 2% solids ; ThermoFischer Scientific F8810) at a 1:100 dilution. For flow visualization, live sheets were observed in FluoroDishes on a Zeiss Z.1 Observer (see "Light microscopy" section). Flow fields around sheets were measured by particle image velocimetry (PIV) using PIVlab software in MATLAB (Ryerson and Schwenk, 2012). Settings for PIV included four direct Fourier transform correlation passes with window sizes of 64, 32, 16 and 8 pixels and corresponding step sizes of 32, 16, 8, and 4 pixels. To reject noise and erroneous velocities, filters of 7 standard deviations about the mean and local median filters with a threshold of 5 and epsilon of 0.1 were applied.

For feeding assays, 1 mL sheet culture were pipetted into wells in a 24-well plate with microbeads at 1:100 dilution, and put on a Reliable Inc. rocking shaker under level 10 agitation for 1 hour. Sheets were then fixed by addition of 1:3 volume 16% PFA at 4°C (to a final concentration of 4% PFA) and carefully transferred into poly-D-lysine-coated FluoroDishes for observation, using a P1000 micropipette with a truncated pipet tip (to limit shear forces). Sheets

were imaged on a Zeiss Z.1 Observer in both DIC (to observe cells) and green fluorescence (to reveal beads), and cells having phagocytosed beads were manually counted for each sheet.

Phototaxis assays and analysis

Phototaxis assays were performed in Ibidi 8-well plate imaging chambers (80826). Wells were completely filled with dense *C. flexa* cultures, and a coverslip (FisherScientific, 12-545-D) was used to cover the well in order to both prevent evaporation and flatten the background illumination intensity profile over the full extent of the well. Samples were then allowed to settle in ambient overhead light conditions for 30min. Next, a phase contrast reference image of the entire well was acquired using a Zeiss Z.1 Observer and Hamamatsu Orca Flash 4.0 V2 CMOS camera (C11440-22CU) by tiling with a 10x, 0.45 NA Plan-Apochromat (Zeiss) objective. The room was then darkened by turning off or blocking all sources of light. Except for the negative control condition, samples were then immediately exposed to directional illumination supplied by a white LED (Elegoo) powered by an Elegoo Uno R3 Arduino supplying 5V through across a 10 kΩ (Elegoo) resistor and focused onto one edge of the well containing the sample using a 15x magnification lens (TV-15 Triview, Carson Optical). This illumination scheme created a conical gradient of decreasing light intensity along the well in the direction of illumination due to increasing distance from the focal point at the edge of the well. Negative controls received no directional illumination. After one hour, a final image of the entire well was acquired as before. For single cell phototaxis experiments, prior to the settling step, sheets were dissociated by vortexing in a 1.5 mL Eppendorf tube using a Vortex Genie 2 (Scientific industries) for one minute.

All image processing was performed using Fiji (Schindelin et al., 2012). First, images were background subtracted using Gaussian blurred (σ = 100 pixels) duplicate images. Images were then cropped to remove the edges of well, and a five-pixel median filter was applied. Images were then imported into MATLAB release 2016a (Mathworks, Natick) where subsequent analysis was performed. For each pair of reference and final images (see "Phototaxis assays" section), a phototaxis index (PI) was calculated, with PI defined as:

$$PI = (I_{pf} / I_{df}) - (I_{pr} / I_{dr})$$

where I_{pr} is the integrated intensity of the first 1/3 of the reference image proximal to the phototaxis light source, I_{dr} is the integrated intensity of the 1/3 of the reference image most distal to the phototaxis light source, I_{pf} is the integrated intensity of the first 1/3 of the final image proximal to the phototaxis light source, and I_{df} is the integrated intensity of the 1/3 of the final image most distal to the phototaxis light source. Because choanoflagellate cells are bright under phase contrast, PI is a readout of the relative change in abunadnce of choanoflagellate cells in the region proximal versus distal to the light source over the course of the phototaxis assay.

Live imaging of *Diaphanoeca grandis*, *Salpingoeca rosetta*, *Salpingoeca urcoelata* and *Monosiga brevicollis*

D. grandis, *S. rosetta*, *S. urceolata* and *M. brevicollis* cultures were obtained by thawing frozen stocks stored in liquid nitrogen in the King lab (following (King et al., 2009b)) and maintained in 1% CGM3/ASW medium. Live cells were mounted in FluoroDishes treated with a handheld Corona surface treater, coated with poly-D-lysine, and rinsed three times with ASW. Imaging was performed in DIC optics at 100x magnification on a Zeiss Observer Z.1 platform using a Hamamatsu Orca Flash 4.0 V2 CMOS camera (C11440-22CU).

FIGURES

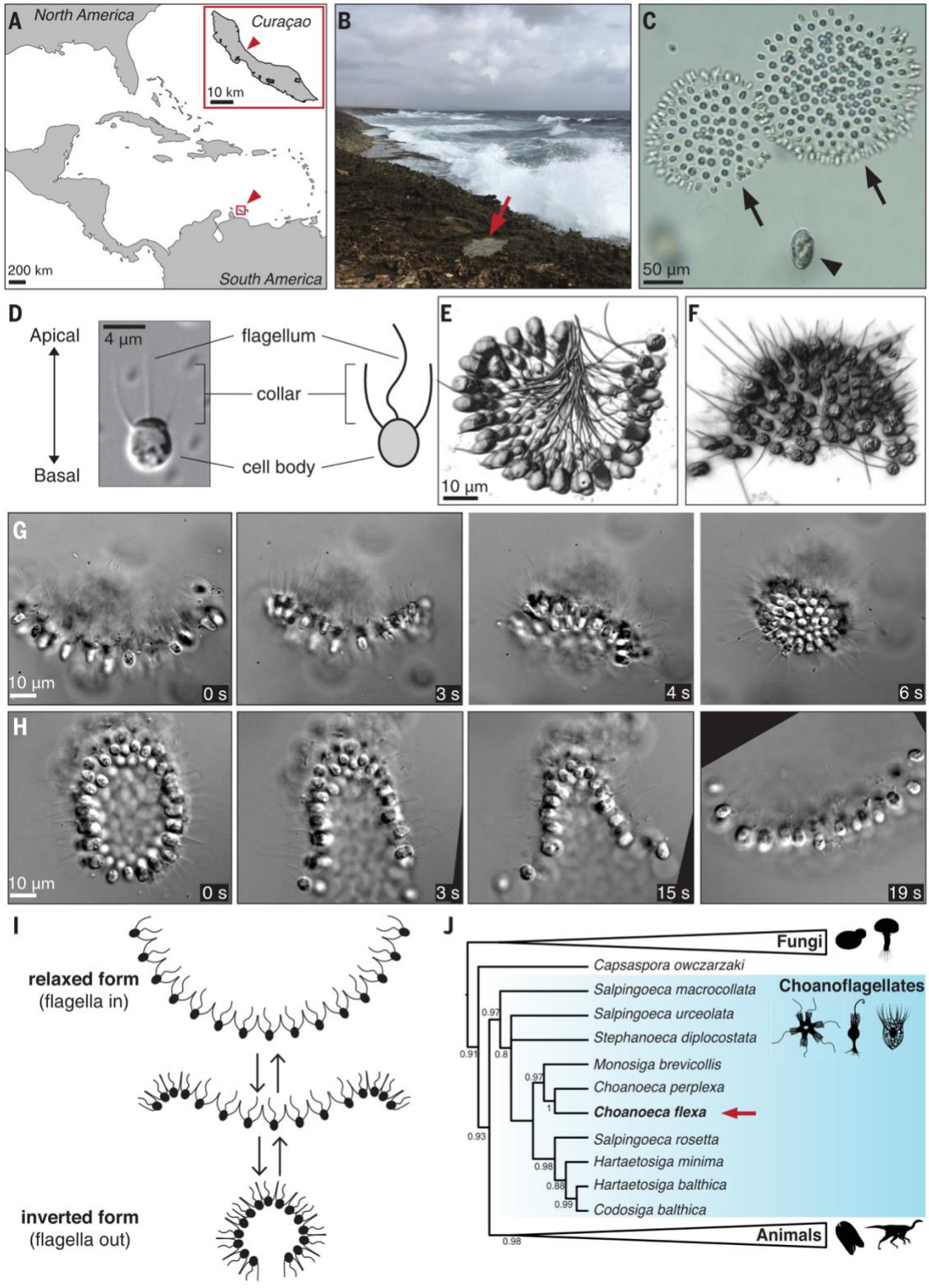

A
North America
Curaçao
10 km
200 km
South America
B
C
50 µm
D
Apical
Basal
4 µm
flagellum
collar
cell body
E
10 µm
F
G
10 µm
0 s
3 s
4 s
6 s
H
10 µm
0 s
3 s
15 s
19 s
I
relaxed form
(flagella in)
inverted form
(flagella out)
J
Fungi
Capsaspora owczarzaki
Salpingoeca macrocollata
Salpingoeca urceolata
Stephanoeca diplocostata
Monosiga brevicollis
Choanoeca perplexa
Choanoeca flexa
Salpingoeca rosetta
Hartaetosiga minima
Hartaetosiga balthica
Codosiga balthica
Choanoflagellates
Animals
0.97
0.91
0.8
0.97
1
0.93
0.98
0.88
0.99
0.98

Figure 2.1 Multicellular sheets of a colonial choanoflagellate from Curaçao rapidly and reversibly invert in curvature. (A to C) *C. flexa* was isolated from splash pools on the northern shore of Curaçao. (A) Map of the Caribbean Sea. Arrowhead: Curaçao. (Inset) Map of Curaçao. Inset arrowhead: sampling site (12°13′38.9″ N 69°00′47.0″ W). (B) Photograph of the sampling site. Arrow: splash pool. (C) Microscopy of freshly collected splash pool sample revealed a diverse microeukaryotic community, including dinoflagellates (*Oxyrrhis* sp.; arrowhead) and cup-shaped choanoflagellate colonies (*C. flexa*; arrows) that rapidly inverted their curvature. Still frame from movie 2S1. (D) Diagnostic features of choanoflagellate cell morphology shown by differential interference contrast (DIC) micrograph and sketch of a *C. flexa* cell. (E and F) *C. flexa* colonies alternate between two conformations, flagella-in (E) and flagella-out (F). (G and H) *C. flexa* colonies rapidly and reversibly invert their curvature while maintaining contacts among neighboring cells. (G) Flagella-in colony inverts to the flagella-out orientation (movie 2S2). (H) Flagella-out colony reverts to the flagella-in orientation (movie 2S3). Movie frames were rotated to facilitate tracking individual cells between images. (I) Summary of the inversion and relaxation processes. (J) Phylogenetic analysis of 18S rDNA sequences revealed that *C. flexa* (bold) is a sister to the choanoflagellate *C. perplexa* (Leadbeater, 1977). (Scaled branch lengths are in Fig. 2S1.)

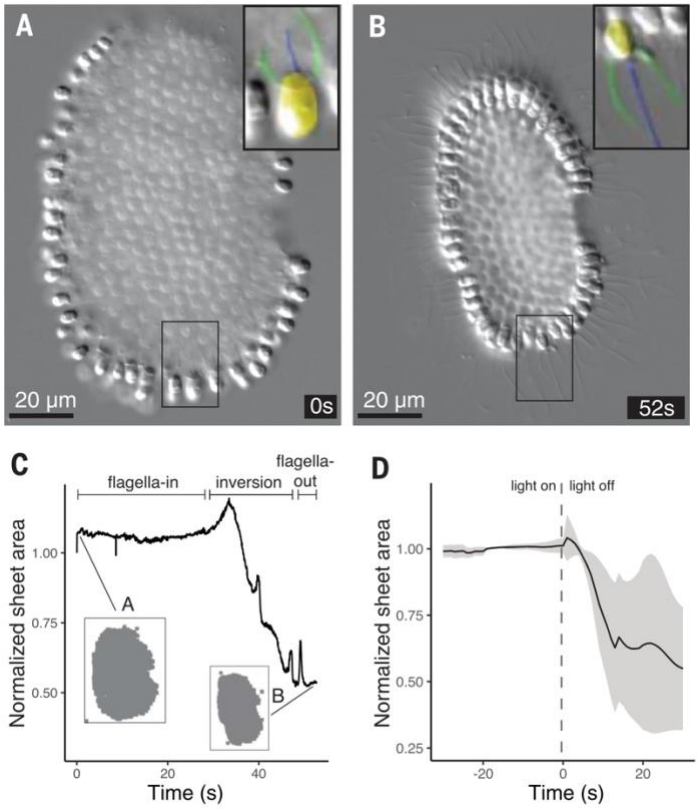


Figure 2.2 Light-to-dark transitions induce *C. flexa* colony inversion. (A to C) Colony inversion correlates with a decrease in the projected area of the colony. (A and B) *C. flexa* colony spontaneously inverts from the flagella-in (A) to the flagella-out (B) orientation (movie 2S5). (Insets) Pseudocoloring highlights the orientations of the cells in the boxed regions. Cell orientation relative to colony curvature inverts without breaking contacts with neighboring cells. (C) The projected area of the colony from (A) and (B) decreased by ~50% during inversion. Sheet area was normalized to its preinversion projected area. (D) Colonies reliably undergo inversion in response to light-to-dark transitions. Normalized projected area of $n = 5$ colonies over time before and after light reduction (vertical dashed line). Line: mean projected area (rolling average over 5-s windows); ribbon: standard deviation.

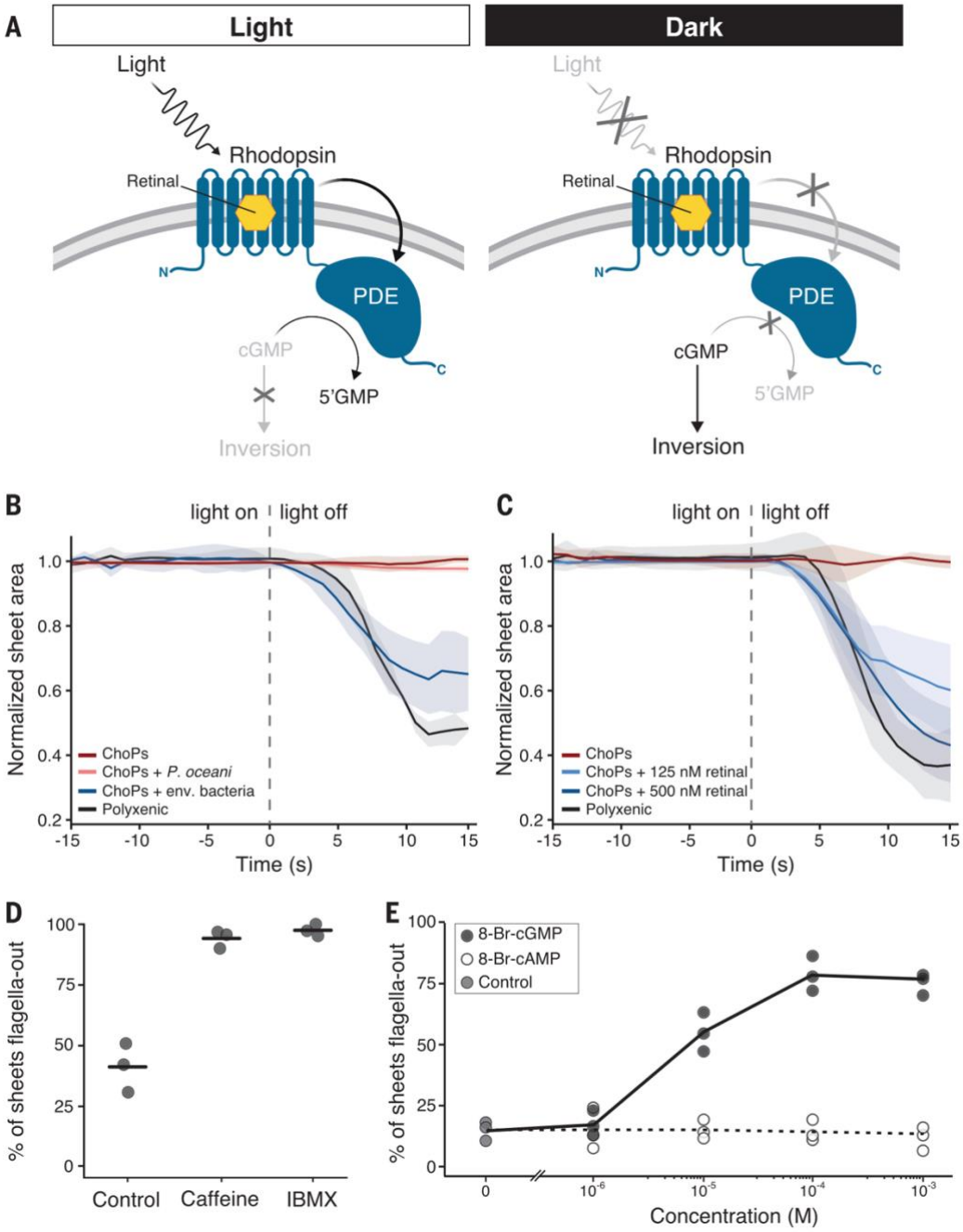


Figure 2.3 *C. flexa* cells transduce light stimuli through a rhodopsin-cGMP pathway using bacterial carotenoids. (A) RhoPDE (blue), a choanoflagellate-specific enzyme rhodopsin. *C. flexa* encodes four copies of RhoPDE (Fig. 2S2), each comprising an 8-transmembrane-pass type I (bacterial) rhodopsin fused to a cyclic nucleotide PDE. Photodetection by rhodopsin requires binding of the chromophore retinal. Under illumination (left panel), the rhodopsin domain activates the PDE domain, which hydrolyzes cGMP to 5′GMP, thereby reducing cellular cGMP levels. Upon light reduction (right panel), the PDE domain is inactivated, allowing cellular cGMP levels to rise. (B) A bacterial factor is required for light-regulated sheet inversion.

In the presence of diverse "polyxenic" bacteria, *C. flexa* sheets inverted to flagella-out in response to decreased illumination *(n* = 3 colonies). By contrast, *C. flexa* sheets grown only with the bacterium *P. oceani* (ChoPs culture, *n* = 4 colonies) did not respond to changes in illumination. The photic response of the ChoPs culture was restored by inoculating with environmental bacteria from the polyxenic culture (ChoPs + env. bacteria, *n* = 3 colonies) but not by inoculating with *P. oceani* bacteria (ChoPs + *P. oceani*, *n* = 3 colonies). (C) Retinal (or one of its carotenoid precursors) is the bacterial molecule required for the photic response. The photic response of the ChoPs culture, which normally does not respond to light stimuli (*n* = 4 colonies), was restored by treatment with 125 nM or 500 nM retinal (*n* = 5 colonies each). Thus, retinal is required for inversion in response to reduced illumination. For (B) and (C), photic response was quantified as for Fig. 2.2. (D) PDE activity suppresses sheet inversion in *C. flexa*. Treatment with the PDE inhibitors caffeine (10 mM) or IBMX (1 mM) caused *C. flexa* colonies to invert to the flagella-out orientation in the absence of a photic stimulus (*n* = 3 independent trials; *N* = 52, 55, and 38 sheets for controls; *N* = 23, 31, and 40 sheets for caffeine; *N* = 42, 37, and 27 sheets for IBMX). (E) cGMP induces sheet inversion. Treating the ChoPs culture with a cell-permeant cGMP analog (8-Br-cGMP) caused sheets to invert into the flagella-out orientation in a dose-dependent manner in the absence of a photic stimulus. Treating with 8-Br-cAMP had no effect.

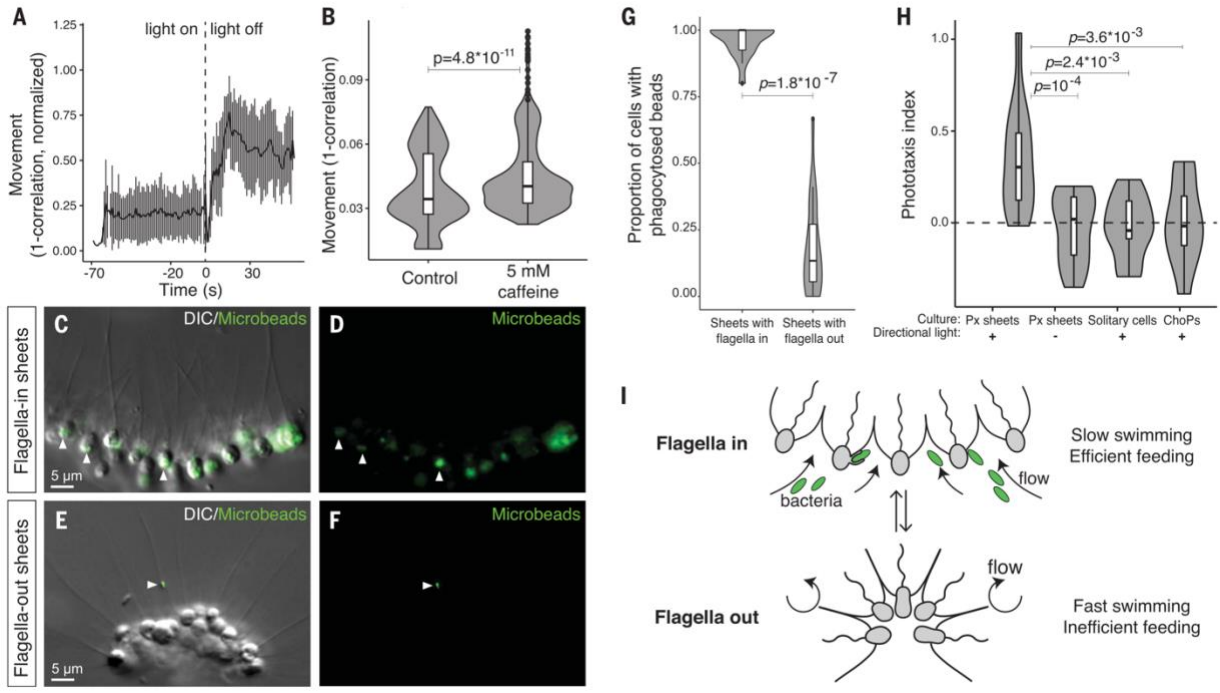


Figure 2.4 Sheet inversion mediates a trade-off between swimming and feeding. (A and B) Flagella-out sheets swim faster than flagella-in sheets. (A) After light-to-dark–induced inversion, swimming speed increased quickly (movie 2S4), as quantified by an increase in the amount of movement (defined as 1-correlation, where "correlation" is the Pearson correlation between two consecutive frames). Movement was normalized between 0 and 1 for each of $n = 9$ movies. Gray shading: error bars representing standard deviation. (B) Sheets swim faster after caffeine-induced inversion under constant light. Movement quantified as in (A). $n = 9$ time-lapse movies for controls (relaxed sheet populations under constant light) and $n = 10$ movies for caffeine-treated samples. (C to G) Flagella-in sheets feed more efficiently than flagella-out sheets. Internalization of 0.2-mm fluorescent beads was used to quantify phagocytic activity. (C to F) Detection of fluorescent microbeads phagocytosed by flagella-in sheets (untreated) and flagella-out sheets (caffeine-treated), fixed after 1 hour. *C. flexa* cells were visualized by DIC (C and E) and beads by green fluorescence (C to F). Arrowheads: fluorescent beads [inside the cells in (C); stuck to a flagellum in (E)]. (G) Proportion of cells having phagocytosed beads in $n = 17$ flagella-in sheets compared with $n = 21$ flagella-out sheets. P value: χ^2 test. (H) *C. flexa* phototaxes in a retinal- and multicellularity-dependent manner, as measured by the phototaxis index, which quantifies directional accumulation of the sheets toward a localized light source (materials and methods). Light-responsive polyxenic (Px) sheets migrated toward a lateral light source over 1 hour ($n = 12$ experiments). By contrast, no directional accumulation was observed in Px sheets without directional light ($n = 12$ experiments), in dissociated single cells ($n = 9$ experiments), or in retinal-deprived light-insensitive monoxenic cultures (ChoPs; $n = 10$ experiments) that do not undergo inversion. P values: analysis of variance with Dunnett's correction. (I) In flagella-in sheets, flagellar beating generates a feeding flow that carries bacteria toward the basal side of the cells (Fig. 2S7), allowing phagocytosis. In flagella-out sheets, flagellar beating causes rapid swimming, whereas the basal side of the cells faces inward, preventing prey capture.

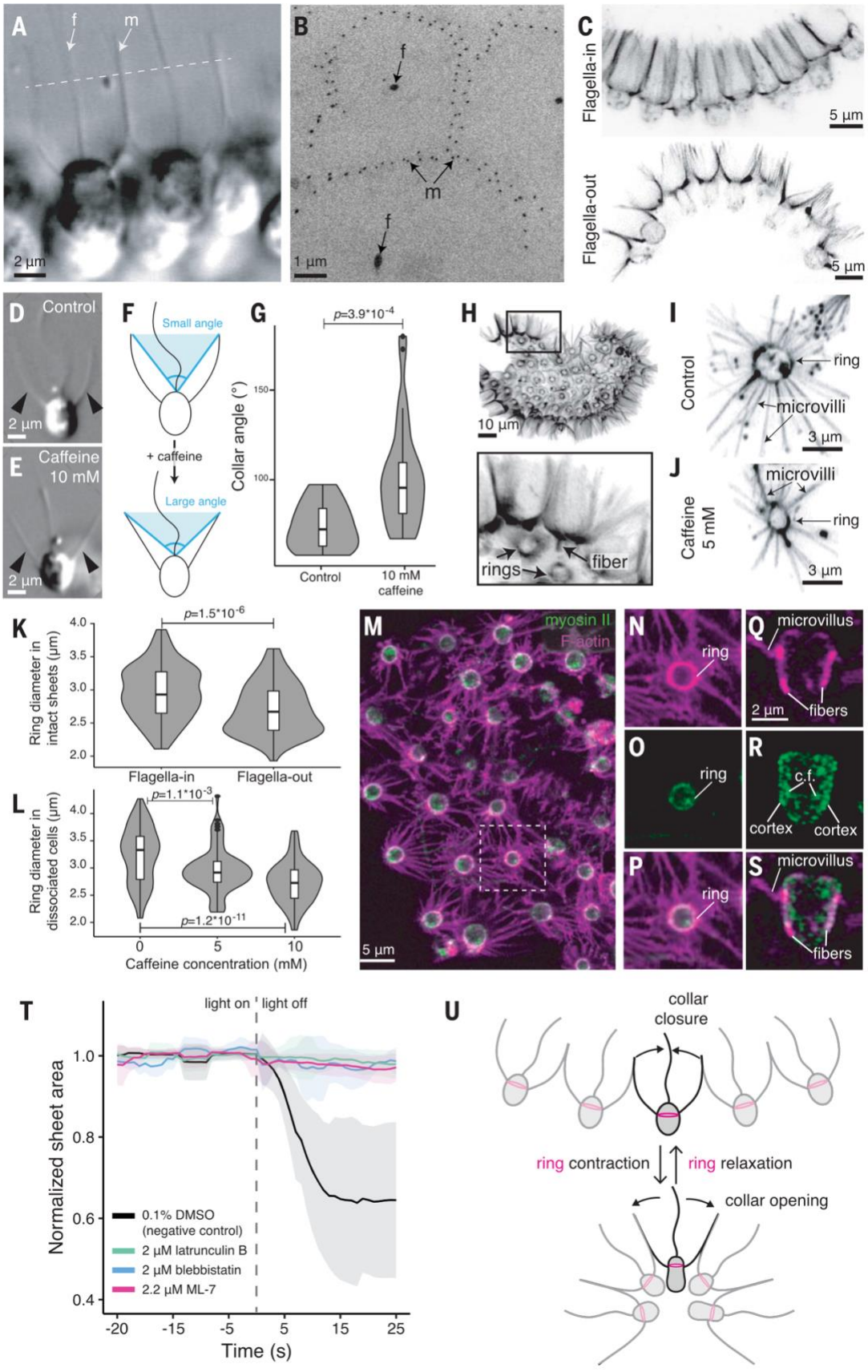

A
f
m
2 µm
B
f
m
f
1 µm
C
Flagella-in
5 µm
Flagella-out
5 µm
D
Control
2 µm
E
Caffeine
10 mM
2 µm
F
Small angle
+ caffeine
Large angle
G
p=3.9*10-4
Collar angle (°)
150
100
Control
10 mM caffeine
H
10 µm
rings
fiber
I
Control
ring
microvilli
3 µm
J
Caffeine 5 mM
microvilli
ring
3 µm
K
p=1.5*10-6
Ring diameter in intact sheets (µm)
4.0
3.5
3.0
2.5
2.0
Flagella-in
Flagella-out
L
p=1.1*10-3
p=1.2*10-11
Ring diameter in dissociated cells (µm)
4.0
3.5
3.0
2.5
2.0
0
5
10
Caffeine concentration (mM)
M
myosin II
F-actin
5 µm
N
ring
O
ring
P
ring
Q
microvillus
2 µm
fibers
R
c.f.
cortex
cortex
S
microvillus
fibers
T
light on
light off
Normalized sheet area
1.0
0.8
0.6
0.4
0.1% DMSO (negative control)
2 µM latrunculin B
2 µM blebbistatin
2.2 µM ML-7
-20
-15
-5
5
15
25
Time (s)
U
collar closure
ring contraction
ring relaxation
collar opening

Figure 2.5 Sheet inversion requires apical actomyosin cell contraction. (A and B) Cells in a sheet are linked by their collars. (A) DIC micrograph showing direct collar contacts between neighboring cells. f, flagellum; m, microvilli. Dotted line: approximate plane of section in (B). (B) Transmission electron micrograph of a transverse section through the collars of neighboring cells. (C) Collar morphology differs between flagella-in sheets (top) and flagella-out sheets (bottom). Cells in flagella-in sheets have barrel-shaped collars, whereas cells in flagella-out sheets have flared, conical collars. Staining: phalloidin. (D to G) Caffeine treatment of dissociated cells causes the collar to flare. (D) In the absence of caffeine, the collar of a cell from a dissociated colony curves upward. (E) In the presence of 10 mM caffeine, the collar of a cell from a dissociated colony straightened and flared open. Arrowheads: microvilli. (F) Differences in collar morphology where quantified in control and caffeine-treated single cells by measuring the collar angle (defined by the tip of two opposite microvilli and the base of the flagellum) (Fig. 2S9). (G) Collar angles in caffeine-treated cells (n = 28) are wider than in control cells (n = 16). P value from Mann-Whitney U test. (H) Actin ring connected to longitudinal fibers is present at the base of each collar. (Inset) Higher magnification showing actin rings and a longitudinal fiber. Staining: phalloidin. (I to L) Actin ring constricts during inversion in intact sheets and in response to caffeine in isolated cells. (I and J) Actin ring observed with phalloidin in an untreated cell (I) and in a cell treated with 5 mM caffeine (J). (K) Ring diameter is larger in flagella-in sheets (n = 8 sheets, N = 124 cells) than in flagella-out sheets (n = 7 sheets, N = 110 cells). P value from Mann-Whitney U test. (L) Ring diameter is smaller in 10 mM caffeine-treated dissociated cells (n = 74) and 5 mM caffeine-treated dissociated cells (n = 82) than in untreated dissociated cells (n = 89). P values are from Dunnett's test for comparing several treatments with a control. (M to S) *C. flexa* myosin II localizes to the actin ring and longitudinal fibers. (M) Immunostained sheet showing apical rings of actin and myosin II at the apical poles of all cells. (N to P) Closer views showing the apical ring. (Q to S) Side view of a stained *C. flexa* cell (apical side up) showing cytoplasmic foci of myosin II (c.f.), as well as cortical staining (cortex) that overlaps with the longitudinal fibers. Green: anti-*C. flexa* myosin II antibody, magenta: rhodamine-phalloidin (Fig. 2S10). (T) Treatment with inhibitors of actin polymerization (latrunculin B, n = 6), myosin contractility (blebbistatin, n = 6), or myosin activation by phosphorylation (ML-7, n = 9) prevented sheet inversion in response to light-to-dark transitions. n = 13 DMSO-treated controls. Photic response is quantified as in Fig. 2.2. DMSO, dimethyl sulfoxide. (U) Proposed role of actomyosin contractility in inversion.

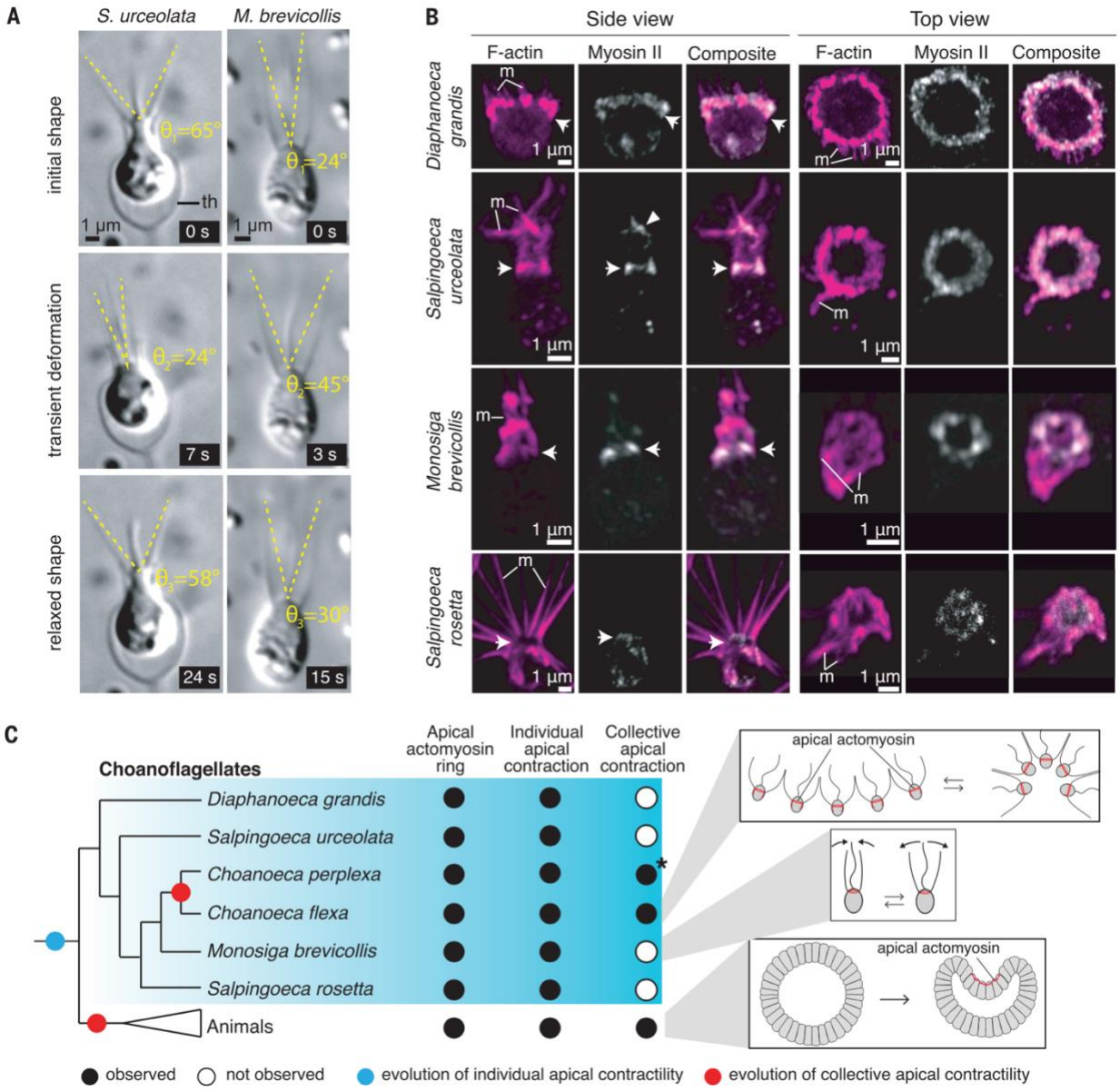


Figure 2.6 Apical constriction is conserved in choanoflagellates. (A) Spontaneous collar contractions observed in *S. urceolata* and *M. brevicollis* (movies S8 and S9). Collar angles measured as in Fig. 2.5F. th, theca. (B) Apical actomyosin ring (arrows) is detected at the base of the collar in four representative choanoflagellate species. Myosin II was also detected in microvilli (m) in *S. urceolata* (arrowhead). (C) Apical constriction of individual cells was present in the last common ancestor of choanozoans and independently gave rise to multicellular apical constriction in *C. flexa* and in animals. **C. perplexa* was previously reported (Leadbeater, 1983) to undergo transient inversions of colony curvature. On the basis of our study of *C. flexa,* we hypothesize that these inversions reflect collective apical constriction.

SUPPLEMENTARY MATERIALS

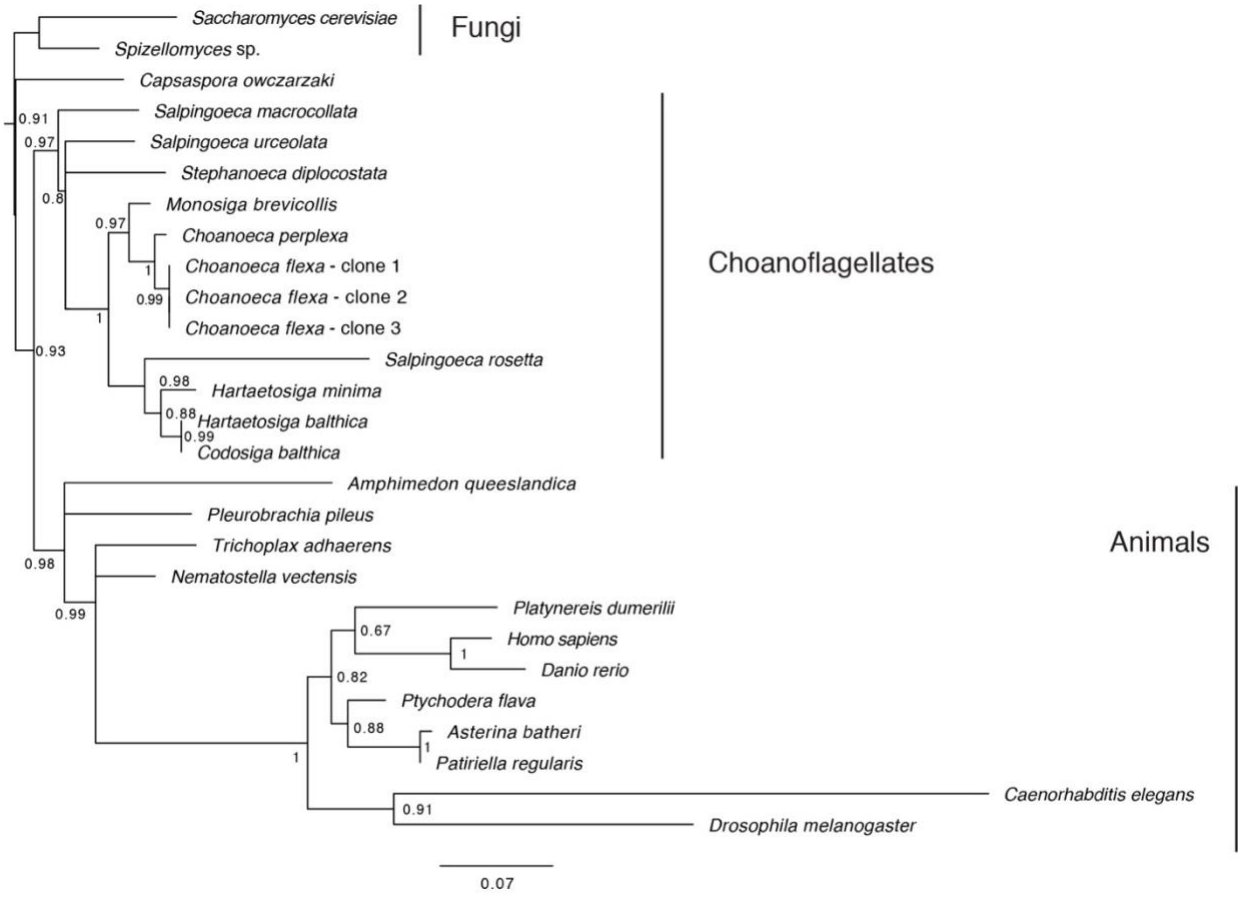
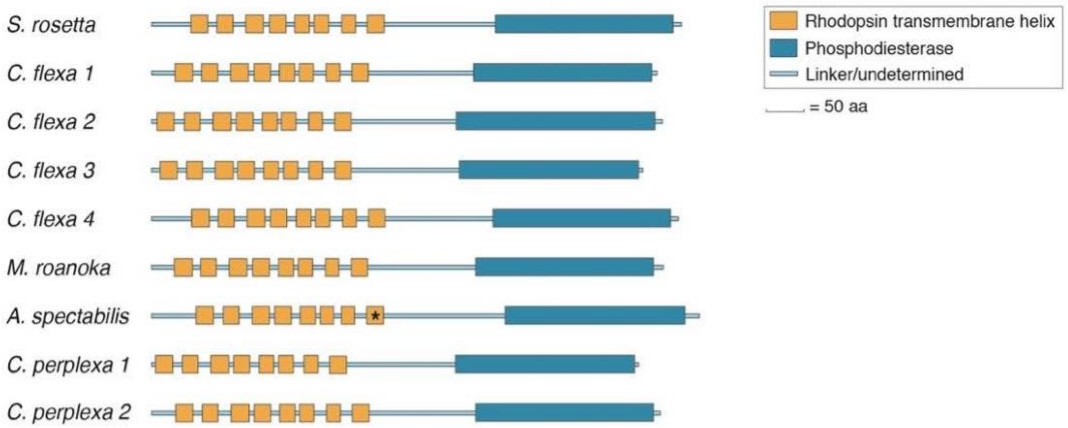


Figure 2S1 Full 18S rDNA phylogenetic tree of *C. flexa,* several other choanoflagellates, and representative outgroup opisthokonts (including animals and fungi). Maximum Likelihood phylogenetic tree of 18S rDNA sequences from *C. flexa* and other opisthokonts. Support values on nodes: approximate likelihood ratio test (aLRT) (Anisimova and Gascuel, 2006). Nodes with support values <0.5 were collapsed into polytomies (e.g. the branching order between the sponge *Amphimedon queeslandica*, the ctenophore *Pleurobrachia pileus* and other metazoans, which represents a currently uncertain point of higher-order metazoan phylogeny (King and Rokas, 2017)).

A



B

	No. in *S. rosetta* RhoPDE	161	164	165	168	172	175	264	267	268	271	292	295	296
	No. in BR	82	85	86	89	93	96	182	185	186	189	212	215	216
	Transmembrane Helix	**4**	**4**	**4**	**4**	**4**	**4**	**7**	**7**	**7**	**7**	**8**	**8**	**8**
	***S. rosetta* RhoPDE**	R	E	W	T	L	W	W	F	P	E	D	A	K
	***C. flexa* RhoPDE 1**	R	E	W	T	L	W	W	F	P	E	D	A	K
	***C. flexa* RhoPDE 2**	R	E	W	T	L	W	W	H	P	E	D	A	K
	***C. flexa* RhoPDE 3**	R	E	W	T	L	V	W	F	P	E	D	A	K
	***C. flexa* RhoPDE 4**	R	E	W	T	L	W	W	F	P	E	D	A	K
H+ pump	BR	R	D	W	T	L	D	W	Y	P	W	D	A	K
Cl- pump	HR	R	T	W	S	I	A	W	Y	P	W	D	A	K
Na+ pump	KR2	R	N	W	D	L	Q	W	Y	P	Y	D	S	K
cation channel	ChR2	R	E	W	T	I	H	W	F	P	F	D	S	K
anion channel	ACR1	R	S	W	T	M	L	W	Y	P	W	D	C	K
sensor	SRII	R	D	W	T	I	F	W	Y	P	W	D	T	K
	ASR	R	D	W	T	L	S	W	Y	P	W	P	S	K
guanylate cyclase	Rhod-GC	R	E	W	T	L	L	W	F	P	W	D	A	K
	Rhod-GC Allomyces	R	E	R	T	L	V	W	V	P	A	D	A	K
histidine kinase	HKR1	R	M	W	T	M	L	W	F	P	W	D	G	K
	HKR (cop6)	R	Q	W	T	M	I	W	F	P	W	N	A	K

C

	No. in *S. rosetta* RhoPDE	492	605	609	616	620	623	624	644	656	657	660	661	693
	No. in PDE10A	564	674	678	685	689	692	693	713	725	726	729	730	762
	***S. rosetta* RhoPDE**	D	D	P	N	S	L	Q	F	G	E	F	I	W
	***C. flexa* RhoPDE 1**	D	D	P	T	A	L	Q	F	G	E	F	I	W
	***C. flexa* RhoPDE 2**	D	D	P	A	A	L	Q	F	G	E	F	I	Y
	***C. flexa* RhoPDE 3**	N	D	N	A	A	L	Q	F	A	Q	H	F	W
	***C. flexa* RhoPDE 4**	D	D	P	T	A	L	Q	F	G	E	F	I	W
	1B	D	D	P	H	T	L	M	L	S	Q	F	I	W
	2A	D	D	Q	T	A	I	Y	M	L	Q	F	M	W
Both	3A	D	D	P	H	T	I	V	F	L	Q	F	I	W
	10A	D	D	V	T	A	I	Y	M	G	Q	F	Y	W
	11A	D	D	V	S	A	V	T	I	L	Q	W	I	W
	4A	D	D	P	Y	T	I	M	M	S	Q	F	I	A
cAMP	7A	D	D	P	S	S	V	T	L	I	Q	F	M	W
	8B	D	D	P	C	A	I	S	V	S	Q	F	I	W
	5A	D	D	I	Q	A	V	A	L	M	Q	F	I	W
cGMP	6A	D	D	I	Q	A	V	A	M	L	Q	F	I	W
	9A	D	D	E	A	V	L	L	F	A	Q	F	I	Y

Figure 2S2 *C. flexa* encodes four putative homologs of RhoPDE. (A) Domain architectures of RhoPDE proteins across choanoflagellates. Shown are all choanoflagellate genes found to

contain a phosphodiesterase domain and the 7 helices characteristic of bacterial (type I) rhodopsins. Phosphodiesterase domains were annotated using PFAM. Rhodopsin transmembrane helices of the *S. rosetta* protein were annotated following Lamarche et al. (2017); those of the other homologs were determined by alignment with the *S. rosetta* protein. *The 8th transmembrane helix of *A. spectabilis* RhoPDE lacks the conserved lysine residue required to covalently bind retinal, and therefore likely lacks rhodopsin activity. (B and C) Comparison of conserved amino acids between *S. rosetta* RhoPDE, *C. flexa* RhoPDEs, and various type I rhodopsins and phosphodiesterases. Proteins and residues were selected as described by Yoshida et al. (2017). Red and blue letters indicate acidic and basic amino acids, respectively. (B) Comparison of conserved amino acids between choanoflagellate RhoPDEs and selected type I rhodopsins. Notably, all of the *C. flexa* homologs contain the conserved lysine residue of the 8th transmembrane helix, which is required for covalent binding of the chromophore retinal. (C) Comparison of conserved amino acids between choanoflagellate RhoPDEs and selected cyclic nucleotide phosphodiesterases. The substrate specificities of the phosphodiesterases are indicated in the leftmost column. *S. rosetta* RhoPDE hydrolyzes cGMP ~10-fold more efficiently than cAMP based on *in vitro* studies (Yoshida et al., 2017).

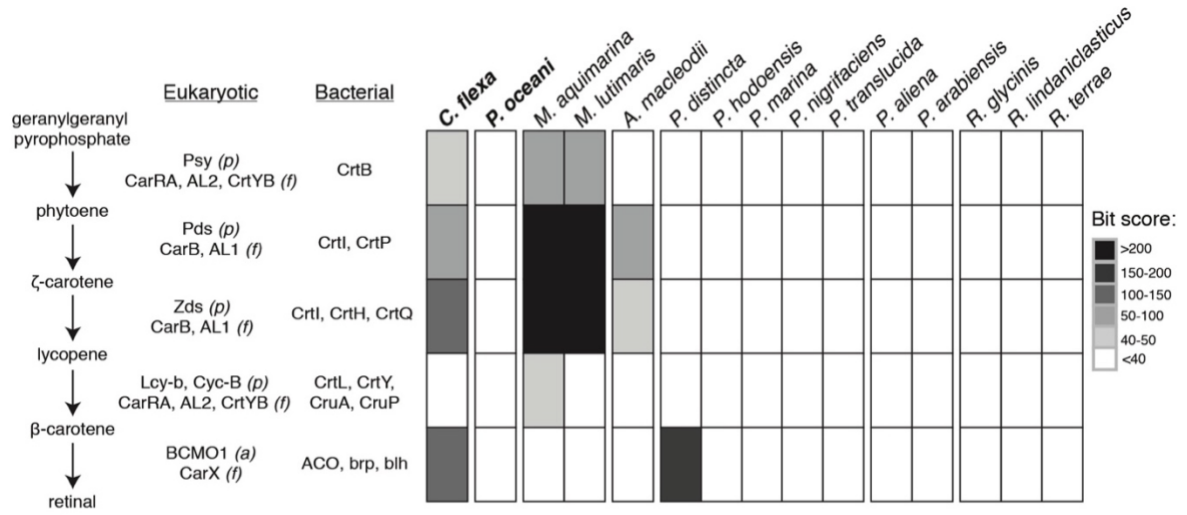
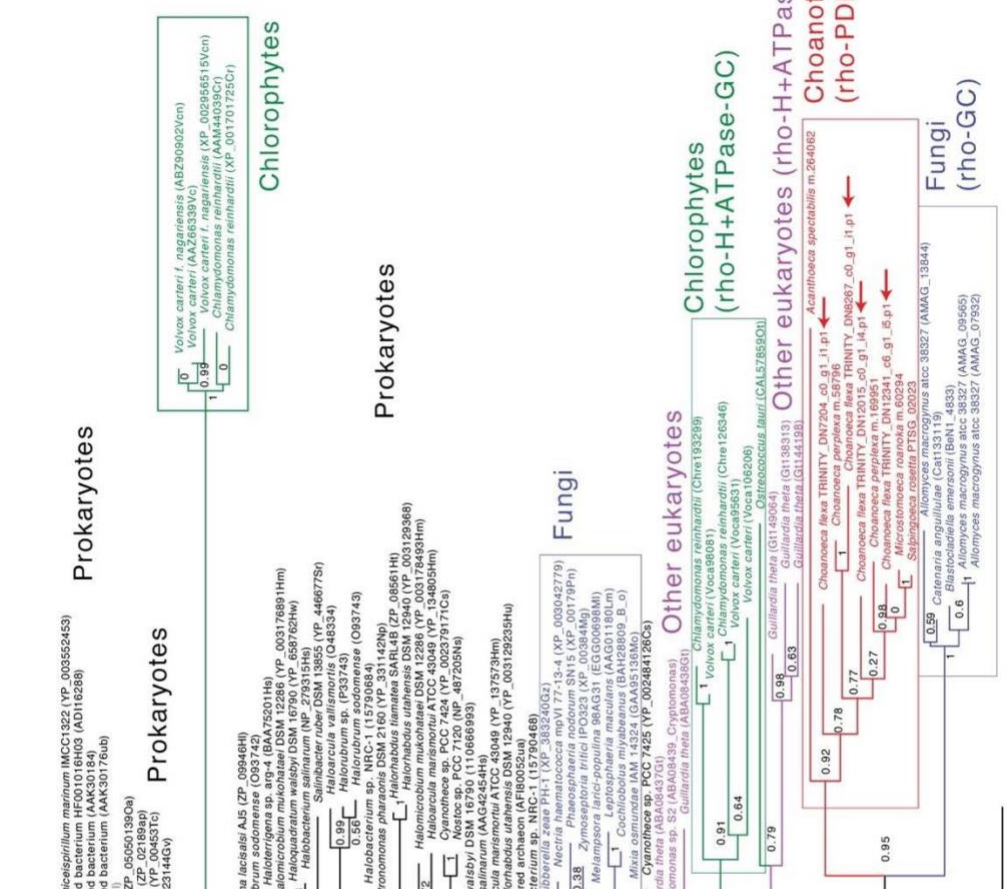


Figure 2S3 Neither *C. flexa* nor the bacterium *Pseudomonas oceani* encodes the complete retinal biosynthesis pathway. The *C. flexa* transcriptome, as well as the genomes of the bacterial species found in the polyxenic culture (Table 2S1), were searched for genes encoding enzymes in the retinal biosynthesis pathway using BLASTP. For each step in the pathway, multiple bacterial, plant ("p"), fungal ("f"), and/or animal ("a") genes were used as queries, and the highest returned score is shown. Notably, *C. flexa* does not encode a recognizable lycopene cyclase, the enzyme that synthesizes beta-carotene, the immediate precursor of retinal. Thus, *C. flexa* must receive retinal or beta-carotene from its food. *Pseudomonas oceani*, the bacterium in the monoxenic ChoPs culture, lacks any recognizable homologs of genes in this pathway, indicating that it likely cannot produce beta-carotene or retinal. The *Bordetella* sp. bacterium shown in Table 2S1 was not included in this analysis because the precise species could not be identified.


A
Prokaryotes
Prokaryotes
Chlorophytes
Prokaryotes
Fungi
Other eukaryotes
Chlorophytes
(rho-H+ATPase-GC)
Other eukaryotes (rho-H+ATPase)
Choanoflagellates
(rho-PDE)
Fungi
(rho-GC)
0.6

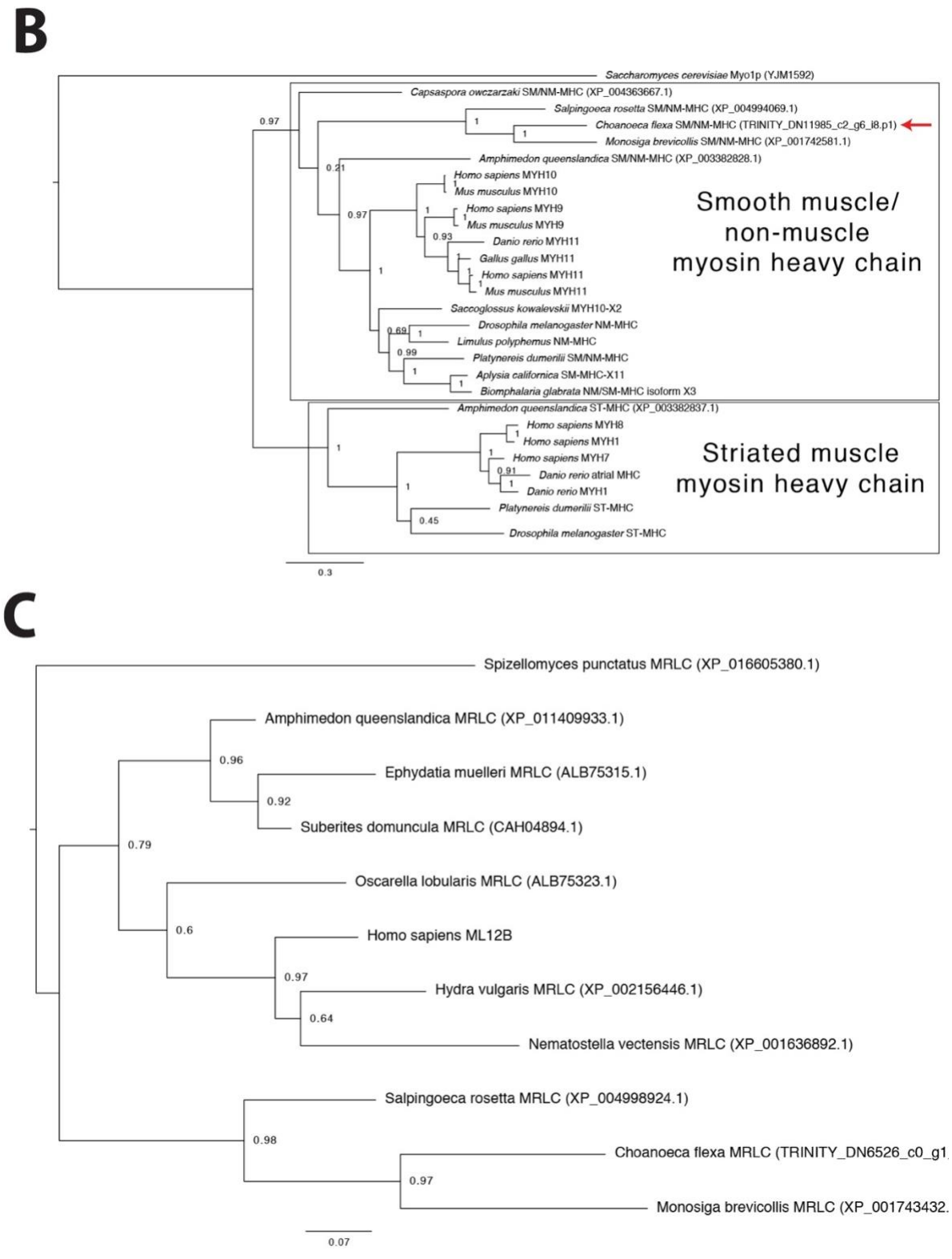


Figure 2S4 Phylogenetic trees of predicted proteins encoded by the *Choanoeca flexa* transcriptome. (A) Phylogenetic tree of microbial rhodopsin domains, based on sequences

compiled by Avelar et al. (2014) to which sequences from five choanoflagellates with microbial rhodopsins (*C. flexa* (red arrows) and sequences from Richter et al. (Richter et al., 2018) – see Fig. 2S2) were added. Choanoflagellate microbial rhodopsins are all fused to phosphodiesterase domain (see Fig. 2S4) and are sister to a clade of fungal microbial rhodospins fused to a guanylyl-cyclase, in agreement with the phylogenetic tree reconstructed by Avelar et al. (2014). (B) Phylogenetic tree of myosin heavy chains (MHC). The *C. flexa* transcriptome appears to encode a single predicted MHC homolog (red arrow) which, like in other studied choanoflagellates (Sebé-Pedrós et al., 2014; Steinmetz et al., 2012), belongs to the smooth muscle/non-muscle family. (C) Phylogenetic tree of myosin regulatory light chains (MRLC). The *C. flexa* transcriptome appears to encode a single predicted MRLC homolog.

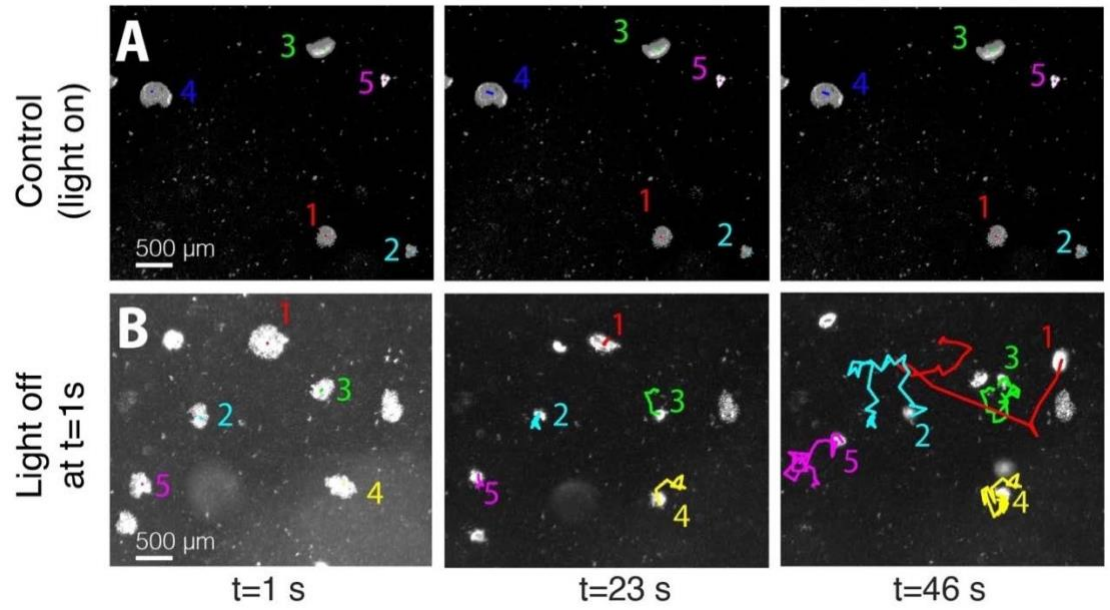


Figure 2S5 Sheets with flagella out are more motile than sheets with flagella in. (A) Five sheets observed at low magnification in a constant level of transmitted light (frames from Movie 2S7). Sheets remained flagella-in and nearly immotile. Colored curves represent tracks of individual sheets. (B) Five different sheets observed in transmitted light following a rapid (~1 sec) reduction in light intensity (frames from Movie 2S4). Following inversion, the sheets became actively motile. Colored lines represent tracks of individual sheets.

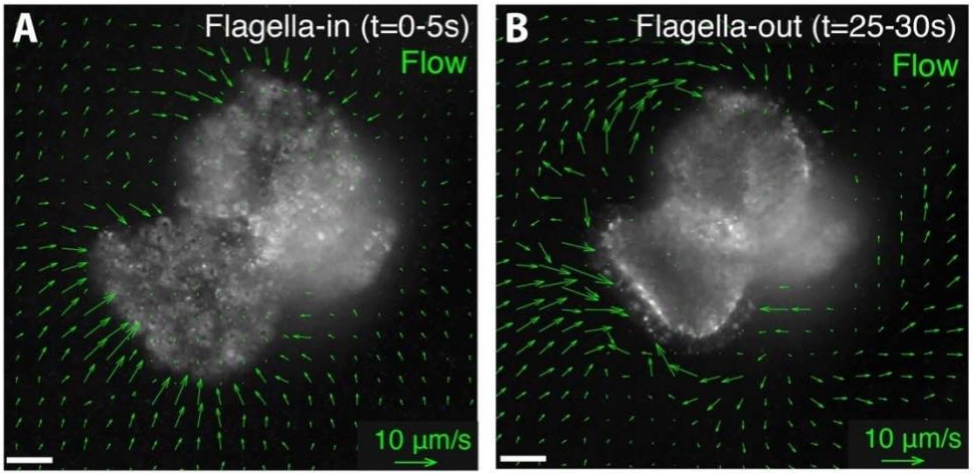


Figure 2S6 The flow field generated by flagellar beating is reoriented during sheet inversion. (A) Feeding flow generated by a sheet with flagella in (visualized by observation of 0.2 µm fluorescent beads in the water and computed by Particle Image Velocimetry; see Material and Methods). The flow is predominantly directed toward the center of the sheet, consistent with it carrying bacterial prey toward the cell bodies. (B) Flow generated by the same sheet after inversion. Note that the flow is predominantly directed parallel to border of the sheet, making it unlikely to support efficient feeding.

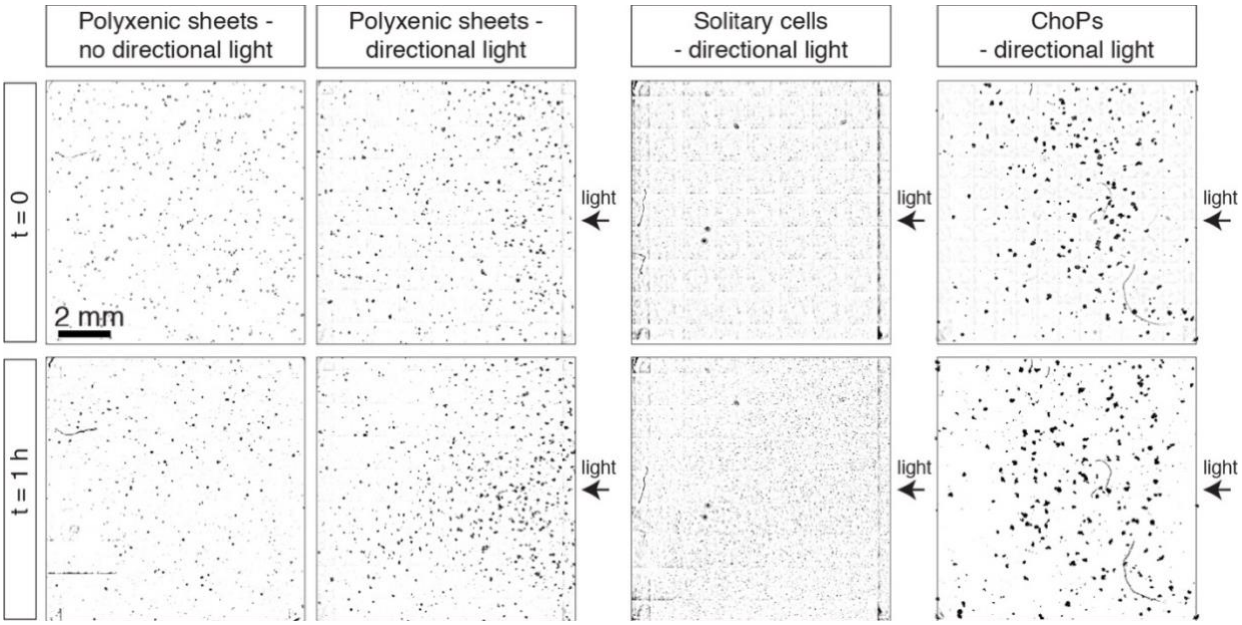


Figure 2S7 Sheet phototaxis requires retinal and multicellularity. Top row: sheets imaged at low magnification at t=0 by transmitted light. Bottom row: sheets imaged at t=1 hour by transmitted light. Polyxenic sheets (cultured with a diversity of co-isolated bacterial species) under directional illumination accumulated toward the side of the flask near the light source (on the right side of the image), while no directional accumulation was observed in ChoPs monoxenic sheets (without retinal and unable to perceive light), polyxenic sheets without directional illumination, and single cells.

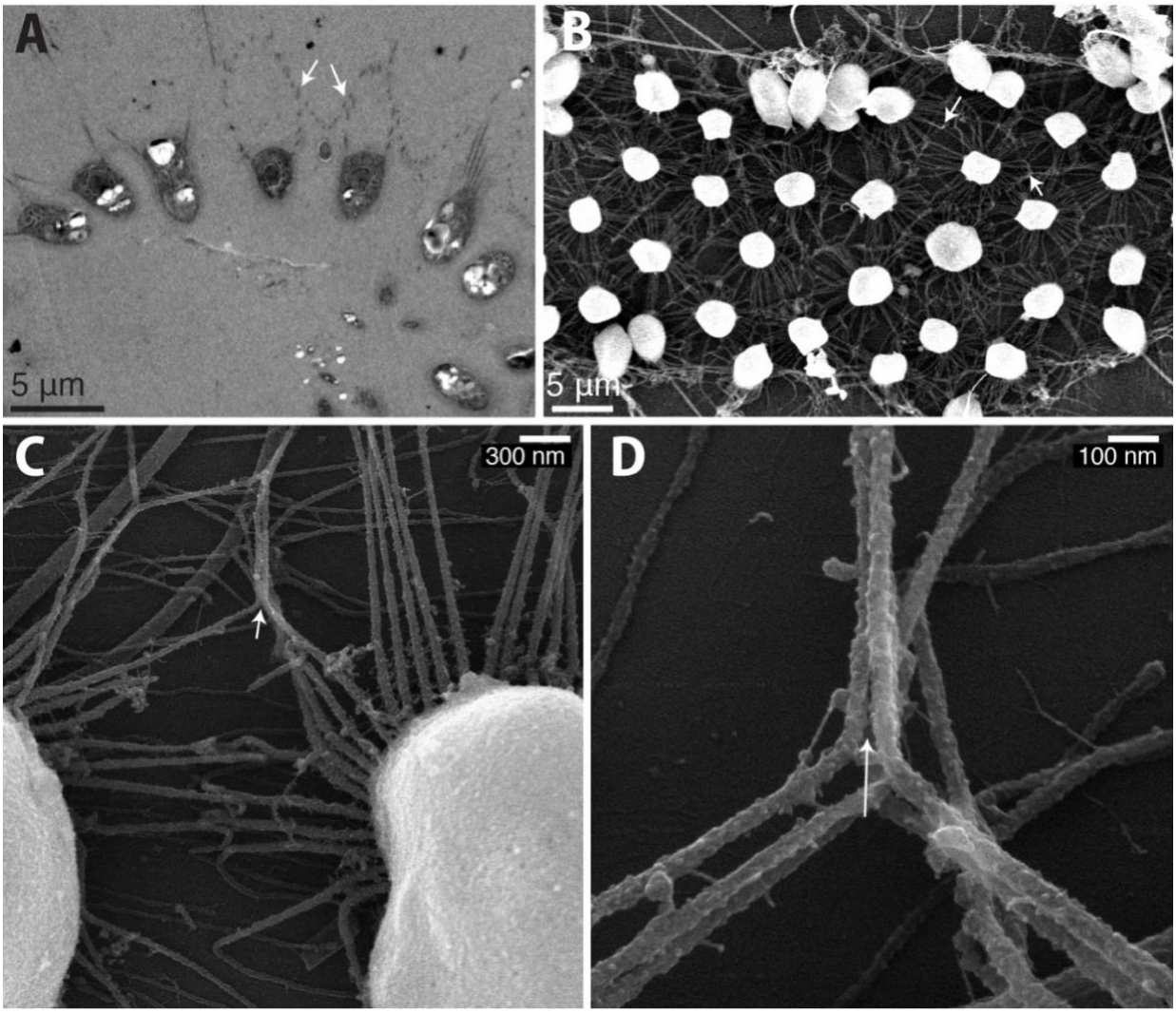


Figure 2S8 Neighboring cells within sheets are linked by their collars. (A) Cross-section of a sheet imaged by transmitted electron microscopy. Note the close proximity between neighboring collars (arrows), and the absence of visible basal extracellular matrix, basal filopodia, or intercellular bridges. (B) Sheet imaged by scanning electron microscopy (SEM). Collars are spread out on the EM disk with cell bodies sticking up. Note the direct contact between neighboring collars (arrows) and the lack of any other structure connecting neighboring cells. (C) Contact between microvilli (arrow) belonging to two neighboring cells imaged by SEM. (D) Same individual as (C) imaged at higher magnification. Note the close apposition of the microvilli belonging to two neighboring cells (arrow), which appear closer to each other than microvilli are within the collar of the same cell.

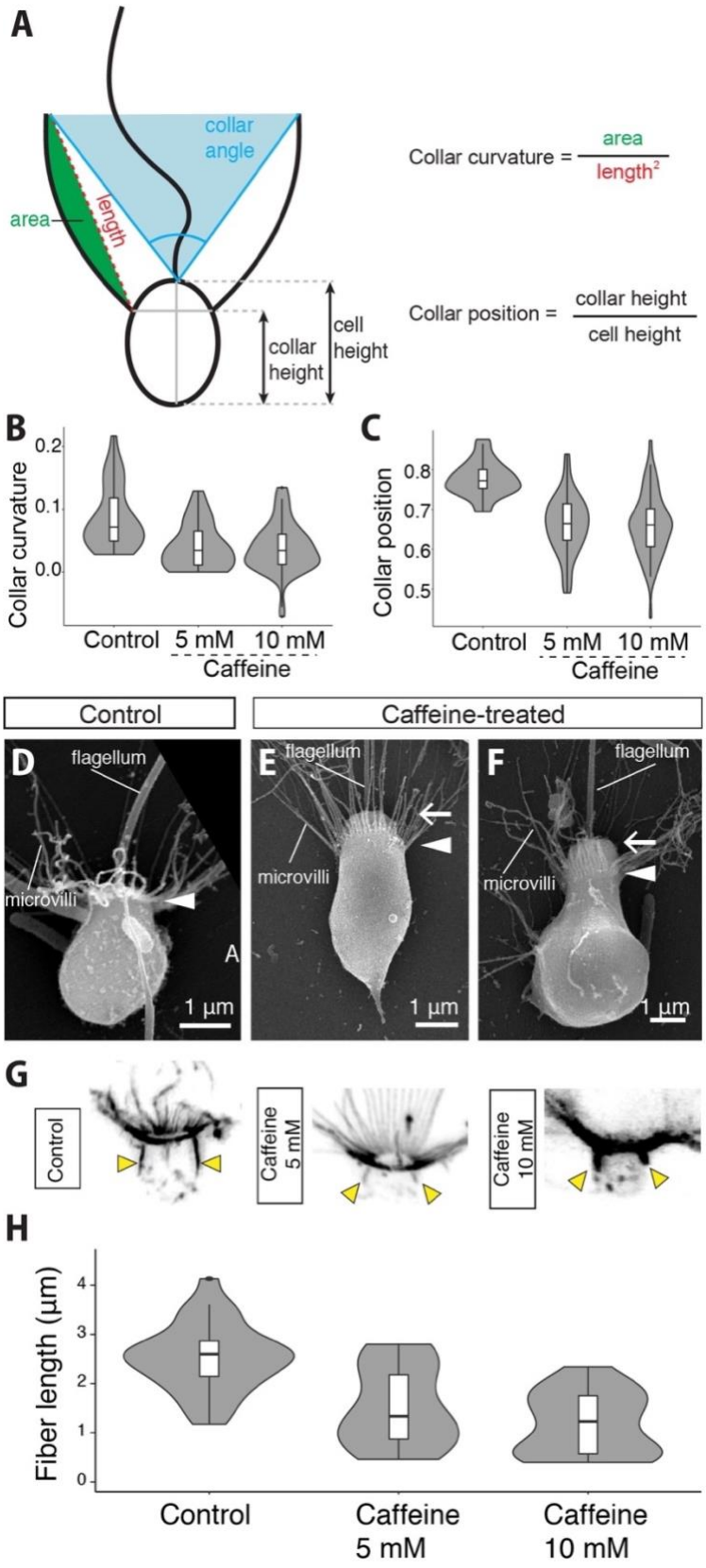

A
collar angle
area
length
collar height
cell height
Collar curvature = area / length²
Collar position = collar height / cell height
B
Collar curvature
0.2
0.1
0.0
Control
5 mM
10 mM
Caffeine
C
Collar position
0.8
0.7
0.6
0.5
Control
5 mM
10 mM
Caffeine
Control
Caffeine-treated
D
flagellum
microvilli
A
1 µm
E
flagellum
microvilli
1 µm
F
flagellum
microvilli
1 µm
G
Control
Caffeine 5 mM
Caffeine 10 mM
H
Fiber length (µm)
4
3
2
1
0
Control
Caffeine 5 mM
Caffeine 10 mM

Figure 2S9 Caffeine treatment results in cell-autonomous collar deformations. (A) Schematic summarizing the values measured and calculated to quantify collar shape. (B) Caffeine treatment leads to a decrease in collar curvature in dissociated cells. n = 30 untreated cells, n = 35 cells treated with 5 mM caffeine, and n = 59 cells treated with 10 mM caffeine. $p = 4.6*10^{-5}$ (control vs. 5 mM caffeine) and $p = 2.9*10^{-6}$ (control vs. 10 mM caffeine) by Dunnett's test for comparing several treatments with a control. (C) Caffeine treatment leads to sliding of the collar toward the basal pole in dissociated cells. n = 27 untreated cells, n = 35 cells treated with 5 mM caffeine, and n = 57 cells treated with 10 mM caffeine. $p = 1.6*10^{-8}$ (control vs. 5 mM caffeine) and $p = 5.7*10^{-5}$ (control vs. 10 mM caffeine) by Dunnett's test. (D to F) Untreated (D) and caffeine-treated (E and F) dissociated cells. In untreated cells, the base of the collar (arrowhead) appears close to the apical side of the cell (arrow; where the flagellum emerges), while it is significantly displaced toward the base in caffeine-treated cells. Note the "bottle cell"-like morphology of the cell in (F) (with a narrow apex and a bulbous base), which is characteristic of a fraction of caffeine-treated cells (or cells in intact sheets during inversion) and is considered diagnostic of apical constriction in animal cells (Kraus and Technau, 2006; Lee and Harland, 2007; Nakajima and Burke, 1996). (G and H) Longitudinal actin (yellow arrowheads) fibers shorten in response to caffeine treatment in dissociated cells. (G) Cells stained for F-actin by fluorescent phalloidin. Yellow arrowheads: longitudinal fibers. (H) Fiber length quantification. n = 28 control cells, n = 16 cells treated with 5 mM caffeine, and n = 20 cells treated with 10 mM caffeine. $p = 5.2*10^{-5}$ (control vs. 5 mM caffeine) and $p = 3.9*10^{-8}$ (control vs. 10 mM caffeine) by Dunnett's test.

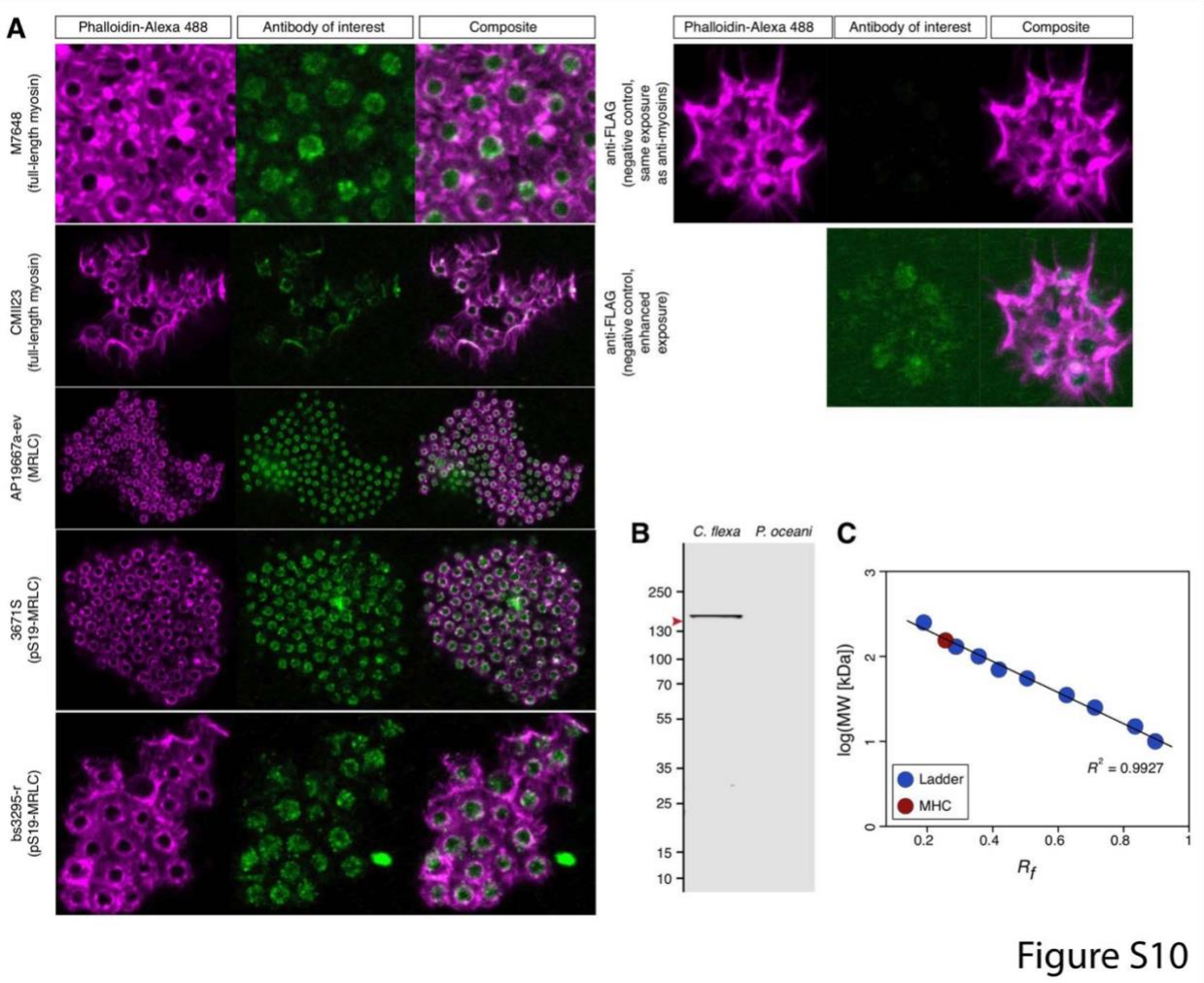


Figure 2S10 Five distinct anti-myosin II antibodies stain the apical actin ring in *Choanoeca flexa*. (A) Sheets immunostained by commercial anti-myosin II antibodies of interest (Sigma Aldrich M7648, Developmental Hybridoma Bank CMII23 against full-length myosin; Abgent AP19667a-ev against Myosin Regulatory Light Chain; and Cell Signalling Technology 3671S and Bioss USA bs-3295r against phosphorylated pS19-Myosin Regulatory Light Chain) and co-stained with rhodamine-phalloidin (targeting F-actin). All antibodies stain a domain nested in (and partly overlapping with) the apical actin ring, while a negative control antibody (Sigma Aldrich F1804 anti-FLAG) gives no detectable pattern under the same imaging conditions and rendering parameters. In sheets stained with the negative control antibody, only unspecific staining in the cell body becomes apparent after increasing the brightness of the corresponding channel in ImageJ. (B) An antibody specifically raised against the heavy chain against *C. flexa* myosin II recognizes a single specific band in a *C. flexa* protein extract by Western Blot, at the predicted size for myosin II heavy chain (154 kDa). Staining of a protein extract of the bacterium *Pseudomonas oceani* (co-cultured with the ChoPs line of *C. flexa*) gives no signal. (C) The predicted 154 kDa molecular weight matches the size of the band observed, as shown by a linear regression of molecular weight as a function of migration distance for both the protein ladder (blue dots) and the specifically stained Western Blot band (red dot).

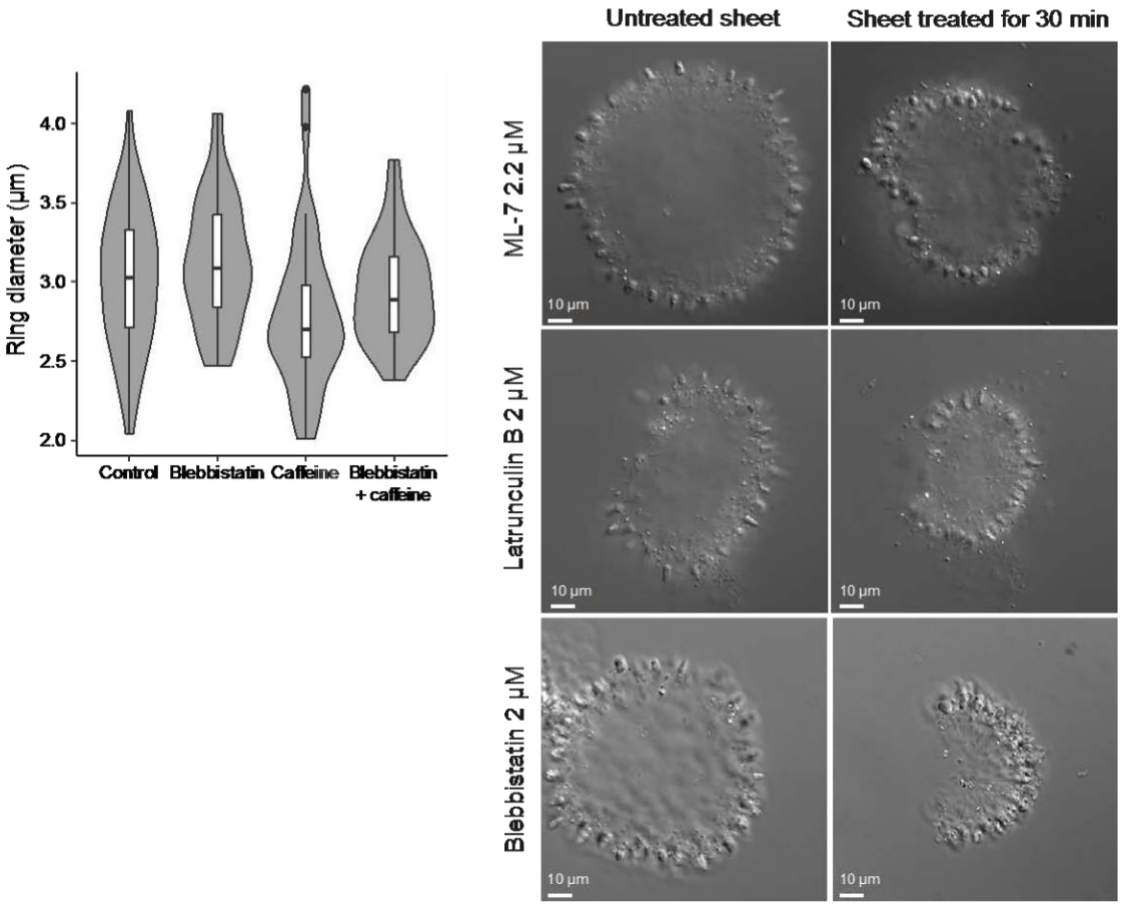


Figure 2S11 Actomyosin inhibition prevents caffeine-induced ring constriction and increases cell packing. (A) Ring diameter measured in dissociated cells, untreated, treated with 2 µM blebbistatin (which inhibits myosin II), treated with 5 mM caffeine, or treated with both 2 µM blebbistatin and 5 mM caffeine. Blebbistatin treatment prevents caffeine-induced ring constriction. n = 164 control cells, n = 106 blebbistatin-treated cells, n = 77 caffeine-treated cells, and n = 99 cells treated with both blebbistatin and caffeine, respectively. p = 0.0111 (caffeine vs. control), p = 0.0034 (caffeine vs. blebbistatin) and p = 0.0097 (caffeine vs. blebbistatin+caffeine) by Kruskal-Wallis test for multiple comparisons. Other comparisons show no significant difference: p = 1.00 (blebbistatin vs. control and blebbistatin vs. blebbistatin+caffeine) and p = 0.76 (blebbistatin vs. blebbistatin+caffeine). (B-D) Sheets treated with actomyosin inhibitors for 30 minutes show increased cell packing in the absence of inversion. Shown are colonies before (B, C and D) and after (B', C' and D') drug treatment.

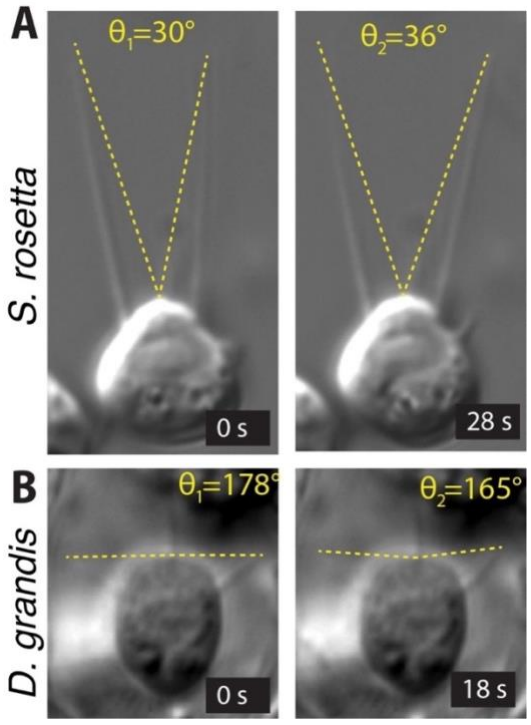


Figure 2S12 Spontaneous deformations of the collar in the choanoflagellates *Diaphanoeca grandis* and *Salpingoeca rosetta*. Time-lapse DIC imaging of live *S. rosetta* cells shows spontaneous reorientation of individual microvilli (Movie 2S11), notably following an increase in light intensity. In the loricate *D. grandis*, which is encased in a self-secreted silicon-based extracellular lodge, spontaneous and reversible changes in microvilli curvature are observed (Movie 2S12). Both types of collar deformation can be quantified as a change in collar angle (yellow doted lines).

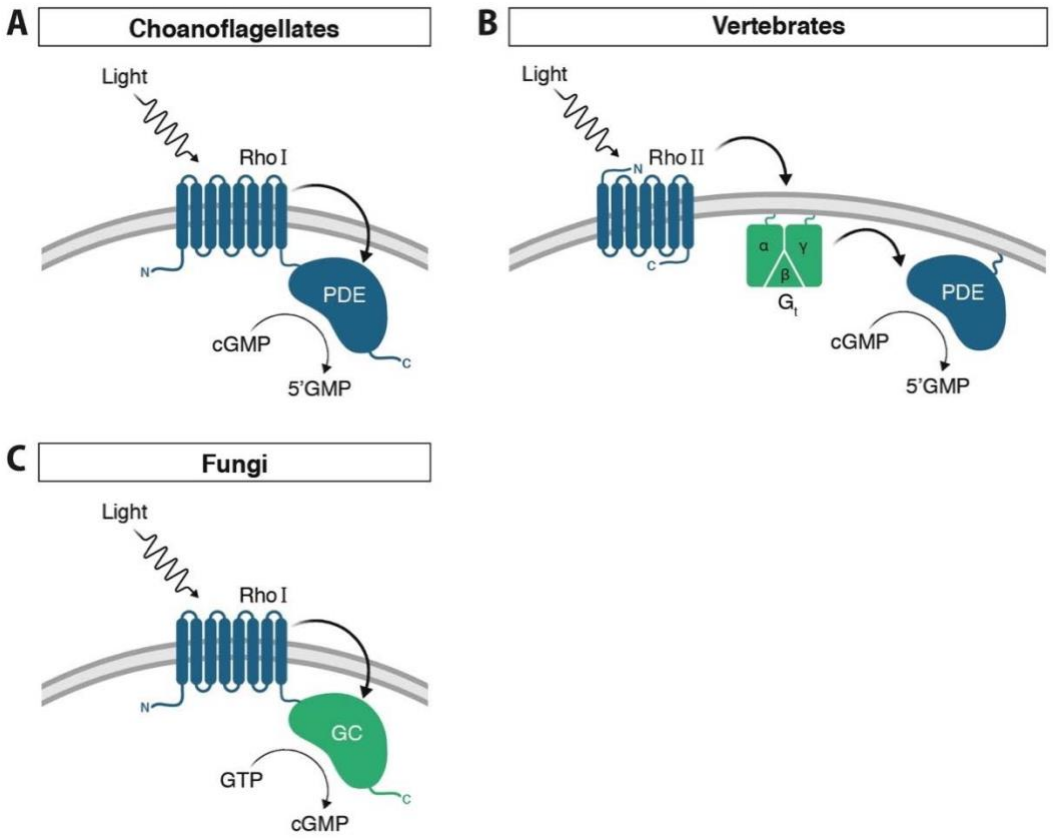


Figure 2S13 Phototransduction pathways in choanoflagellates, animals, and fungi. (A) Many choanoflagellates (Fig. 2S2) encode a fusion protein, RhoPDE, composed of a type I (bacterial) rhodopsin fused to a cyclic nucleotide phosphodiesterase (PDE). When illuminated, the rhodopsin activates the PDE domain, resulting in hydrolysis of cGMP (Watari et al., 2019; Yoshida et al., 2017). (B) A similar rhodopsin-cGMP pathway controls phototransduction in vertebrate photoreceptor cells. Light activates a type II (eukaryotic) rhodopsin in the disk membrane, which activates a cGMP-specific phosphodiesterase via a G-protein intermediary (Gt). The G-protein and the PDE are tethered to the membrane by lipid modifications (Arendt et al., 2015; Zhang and Cote, 2005). (C) Like choanoflagellates, some fungal zoospores use a type I rhodopsin fusion protein for phototransduction. However, this fungal protein, RhoGC, comprises a rhodopsin fused not to a PDE but to a guanylyl cyclase (GC), which catalyzes light-dependent synthesis of cGMP from GTP. Thus, illumination results in increased cellular cGMP levels (Avelar et al., 2014).

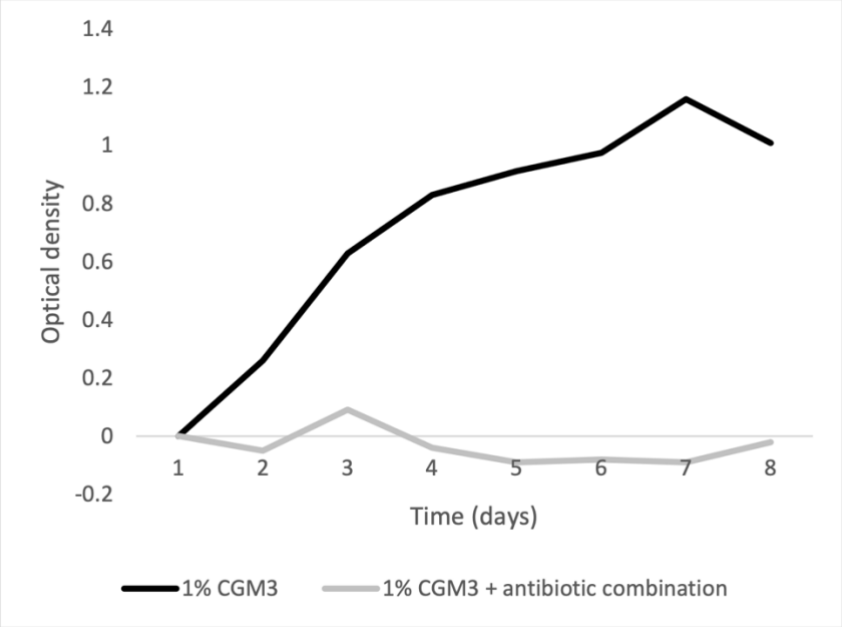


Figure 2S14 Growth of environmental bacteria co-isolated with *C. flexa*, monitored by optical density measurements, in the presence or in the absence of antibiotics.

Table 2S1. Bacterial composition of polyxenic sheet culture determined by 16S sequencing. The bacterial species present in the polyxenic *C. flexa* culture were identified by iTag sequencing of 16S rDNA (Degnan and Ochman, 2011). Shown are all species for which at least 10 reads were recovered. Due to the short length of the 16S amplicon (rightmost column), some reads were compatible with multiple bacterial species (e.g., first row). The 16S sequence of the *Bordetella* bacterium did not match any species in the NCBI database.

Phylum	Species	Reads	Representative 16S Sequence
Gammaproteobacteria	*Pseudoalteromonas* sp. *(distincta, hodoensis, marina, nigrifaciens,* or *translucida)*	147,406 (57.1%)	AGCGTTAATCGGAATTACTGGGCGTAAAGCG TACGCAGGCGGTTTGTTAAGCGAGATGTGAA AGCCCCGGGCTCAACCTGGGAACTGCATTTC GAACTGGCAAACTAGAGTGTGATAGAGGGTG GTAGAATTTCAGGTGTAGCGGTGAAATGCGT AGAGATCTGAAGGAATACCGATGGCGAAGGC AGCCACCTGGGTCAACACTGACGCTCATGTA CGAAAGCGTG
Gammaproteobacteria	*Alteromonas macleodii*	62,765 (24.3%)	AGCGTTAATCGGAATTACTGGGCGTAAAGCG CACGCAGGCGGTTTGTTAAGCTAGATGTGAA AGCCCCGGGCTCAACCTGGGATGGTCATTTA GAACTGGCAGACTAGAGTCTTGGAGAGGGGA GTGGAATTCCAGGTGTAGCGGTGAAATGCGT AGATATCTGGAGGAACATCAGTGGCGAAGGC GACTCCCTGGCCAAAGACTGACGCTCATGTG CGAAAGTGTG
Gammaproteobacteria	*Pseudomonas oceani*	21,464 (8.3%)	AGCGTTAATCGGAATTACTGGGCGTAAAGCG CGCGTAGGCGGCTTGATAAGATGGGTGTGAA ATCCCCGGGCTTAACCTGGGAACTGCATCCA TAACTGTCTGGCTAGAGTACAGTAGAGGGTG GTGGAATTTCCTGTGTAGCGGTGAAATGCGT AGATATAGGAAGGAACACCAGTGGCGAAGGC GACCACCTGGACTGATACTGACGCTGAGGTG CGAAAGCGTG
Gammaproteobacteria	*Pseudoalteromonas aliena*	1,6271 (6.3%)	AGCGTTAATCGGAATTACTGGGCGTAAAGCG TACGCAGGCGGTTTGTTAAGCGAGATGTGAA AGCCCCGGGCTCAACCTGGGAACTGCATTTC GAACTGGCAAGCTAGAGTGTGATAGAGGGTG GTAGAATTTCAGGTGTAGCGGTGAAATGCGT AGAGATCTGAAGGAATACCGATGGCGAAGGC AGCCACCTGGGTCAACACTGACGCTCATGTA CGAAAGCGTG
Bacteroidetes	*Muricauda* sp. *(aquimarina* or *lutimaris)*	5,419 (2.1%)	AGCGTTATCCGGAATCATTGGGTTTAAAGGG TCCGTAGGCGGGCCTGTAAGTCAGGGGTGAA AGTTTGTGGCTCAACCATAAAATTGCCTTTG ATACTGCAGGTCTTGAGTCATGGTGGGGTTG CCGGAACATGTGGTGTAGCGGTGAAATGCAT AGATATCACATAGAACACCGATCGCGAAGGC AGGTGACCAACCATGTACTGACGCTGATGGA CGAAAGCGTG
Gammaproteobacteria	*Pseudoalteromonas arabiensis*	4,567 (1.8%)	AGCGTTAATCGGAATTACTGGGCGTAAAGCG TACGCAGGCGGTTTGTTAAGCGAGATGTGAA AGCCCTGGGCTCAACCTGGGAACTGCATTTC GAACTGGCAAACTAGAGTGTGATAGAGGGTG GTAGAATTTCAGGTGTAGCGGTGAAATGCGT AGAGATCTGAAGGAATACCGATGGCGAAGGC AGCCACCTGGGTCAACACTGACGCTCATGTA CGAAAGCGTG

Gammaproteobacteria	*Bordetella* sp.	443 (0.171%)	AGCGTTAATCGGAATTACTGGGCGTAAAGCG TGCGCAGGCGGTTCGGAAAGAAAGATGTGAA ATCCCAGGGCTTAACCTTGGAACTGCATTTT TAACTGTCGAACTAGAGTGCGTCAGAGGGGG GTGGAATTCCGCGTGTAGCAGTGAAATGCGT AGAGATGCGGAGGAACACCGATGGCGAAGGC AGCCCCCTGGGATGACACTGACGCTCATGCA CGAAAGCGTG
Gammaproteobacteria	*Rhodanobacter* sp. *(lindaniclasticus, glycinis,* or *terrae)*	10 (0.004%)	AGCGTTAATCGGAATTACTGGGCGTAAAGGG TGCGTAGGCGGTTAGTTAAGTCTGTTGTGAA ATCCCCGGGCTCAACCTGGGAATGGCAATGG ATACTGGCTAGCTAGAGTGTGTCAGAGGATG GTGGAATTCCCGGTGTAGCGGTGAAATGCGT AGAGATCGGGAGGAACATCAGTGGCGAAGGC GGCCATCTGGGACAACACTGACGCTGAAGCA CGAAAGCGTG

Table 2S2 Primer sequences used for 18S cloning.

Primer sequences used for 18S cloning	Forward primer	Reverse primer
First PCR	CTCAARGAYTAAGCCATGCA	CCGCCCCAGYCAAACTCCC
Nested PCR	GAAACTGCGAATGGCTC	ACCTACGGAAACCTTGTTACG

Table 2S3 Commercial antibodies used for myosin immunostaining.

Provider	Antibody	Target
Developmental Hybridoma Bank	CMII23	Myosin heavy chain
Sigma Aldrich	M7648	Full-length myosin
Abgent	AP19667a-ev	Myosin regulatory light chain
Cell Signalling	3671S	pSer19-Myosin Regulatory Light Chain
Bioss Inc.	bs-3295R	pSer19-Myosin Regulatory Light Chain
Sigma Aldrich	F1804-50UG	FLAG (negative control)

Table 2S4 Drugs used in pharmacological assays.

Name (target)	Stock concentration	Working concentration
IBMX (PDE)	1 M in DMSO	1 mM
Caffeine (PDE)	77 mM in H2O	5 mM
Y-27632 (ROCK kinase)	14 mg/mL in H2O	43.7 μM
ML-7 (MRLC kinase)	10 mg/mL in DMSO	22 μM
All trans-retinal	88 mM in ethanol	88 μM
Latrunculin B (F-actin)	20 mM in DMSO	2-20 μM
Blebbistatin (myosin II)	17 mM in 90% DMSO	20 μM
8-Br-cAMP	100 mg/mL in H2O	10^{-5} to 10^{-3} M
8-Br-cGMP	50 mg/mL in H2O	10^{-5} to 10^{-3} M

Supplementary Movies

Supplementary movies 2S1-2S14 can be found online at
https://science.sciencemag.org/content/366/6463/326/suppl/DC1

Movie 2S1. Sheet inversion and relaxation observed in an environmental splash pool sample.

Movie 2S2. Sheet inversion (from flagella-in to flagella-out) observed in DIC transmitted light microscopy.

Movie 2S3. Sheet relaxation (from flagella-out to flagella-in) observed in DIC transmitted light microscopy.

Movie 2S4. Light-to-dark transitions induces sheet inversion followed by fast swimming. Sheet culture observed with a Zeiss AxioZoom transmitted light dissecting microscope.

Movie 2S5. Spontaneous sheet inversion, showing decrease in projected area. Time lapse movie of a sheet observed in DIC transmitted light microscopy.

Movie 2S6. Light-to-dark transitions reliably induce sheet inversion, which can be quantified as a decrease in projected area. Left: time lapse movie of a sheet observed in phase contrast microscopy. Right: projected area. The light-to-dark transition is visible as a white frame at t=52 seconds, and corresponds to a 75-fold decrease in the intensity of incident light. Camera exposure time was increased correspondingly to allow continued imaging.

Movie 2S7. Sheets in constant light display have low mobility. Sheet culture observed with a Zeiss AxioZoom transmitted light dissecting microscope.

Movie 2S8. Spontaneous collar contractions in the thecate form of *Choaneca flexa*. Note that the thecate form lacks a flagellum, as in *C. perplexa* (Leadbeater, 1977), the sister-species of *C. flexa*.

Movie 2S9. Spontaneous collar contractions in *Salpingoeca urceolata*. Cells are attached to the substrate by a stalked, cup-shaped extracellular lodge called a theca. The most spectacular contraction event is visible shortly after t=610 seconds and corresponds to a rapid decrease in collar angle, concomitant with retraction of the cell body inside the theca. The cell then reverts to its resting shape in the following 20 seconds.

Movie 2S10. Spontaneous collar contractions in *Monosiga brevicollis*. Bacteria co-cultured with *M. brevicollis* are visible around the cell body.

Movie 2S11. Spontaneous reorientation of microvilli in *Salpingoeca rosetta*.

Movie 2S12. Spontaneous changes in collar curvature in *Diaphanoeca grandis*. The cell is encased within an extracellular lodge called a lorica, composed of self-secreted silicon strips.

Movie 2S13. Inversion observed of a very small sheet (24 cells) observed in DIC transmitted light microscopy.

Movie 2S14. Inversion observed of a very large sheet (>565 cells) observed in DIC transmitted light microscopy.

Chapter 3

Widespread distribution of collagens and collagen-associated domains in eukaryotes

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

Linden, T.A., King, N. (2021) Widespread distribution of collagens and collagen-associated domains in eukaryotes. *bioRxiv* [https://doi.org/10.1101/2021.10.08.463732].

Abstract

The origin of collagen, the dominant structural component of metazoan extracellular matrix, has long been cited as a critical step in the evolution of metazoan multicellularity. While collagens were once thought to be found only in metazoans, scattered reports of collagen domains in Fungi, and more recently in close relatives of metazoans, have called into question whether collagens are truly unique to metazoans. Here, we take advantage of recently sequenced genomes and transcriptomes of diverse holozoans (the clade encompassing metazoans and their close relatives), as well as publicly available proteomes from diverse non-holozoan eukaryotes, to conduct a systematic search for collagen domains across eukaryotic diversity. We find that collagen domains are ubiquitous in choanoflagellates, the sister group of metazoans, and widespread across many other major eukaryotic taxa. Many predicted collagens in non-metazoans are comparable to metazoan collagens in length and proline content. Moreover, most are present in species that also encode putative prolyl 4-hydroxylase domains, suggesting that, like metazoan collagens, they may be stabilized through the hydroxylation of prolines. Fibrillar collagen and collagen IV appear potentially unique to metazoans, and we posit that their ability to assemble into superstructures may have contributed to the origin of metazoan multicellularity.

Introduction

In metazoans, the extracellular matrix (ECM) is indispensable to development, form, and function. Collagen is the major structural protein of metazoan ECM (Frantz et al., 2010), and is the most abundant protein in metazoan bodies by mass, making up one third of the dry weight of an adult mammal (Shoulders and Raines, 2009) and over 70% of the weight of some sponges (Ehrlich et al., 2018). The diagnostic feature of collagens is the repeated triplet motif Gly-X-Y, in which X and Y can be any amino acid (Box 1). The repetition of the collagen repeat motif Gly-X-Y leads to the formation of a stable, rod-like triple helix, which in turn supplies tensile strength to metazoan tissues (Buehler, 2006; Shoulders and Raines, 2009; Wenger et al., 2007).

For most of the twentieth century, collagen was thought to be unique to metazoans and absent from other lineages with complex multicellularity (e.g., plants and fungi). Thus, because of its apparent phylogenetic restriction and its importance to metazoan ECM, collagen was repeatedly hypothesized to be one of the key innovations that allowed metazoans and their characteristic tissues to evolve (Erwin, 1993; Towe, 1970). The hypothesized connection

between metazoan origins and the evolution of collagen was brought into question with the sequencing of genomes from three close relatives of metazoans—two choanoflagellates and one filasterean—each of which encode collagen domains (Fairclough et al., 2013; Grau-Bové et al., 2017; Hynes, 2012; King et al., 2008). These joined reports of putative collagens in Dikarya Fungi (Celerin et al., 1996; de Bruin et al., 2002; Tuntevski et al., 2013; Wang and Leger, 2006), the malaria parasite *Plasmodium falciparum* (Rasmussen et al., 2003), and the amoebozoan *Dictyostelium discoideum* (Fidler et al., 2017) to create a growing catalog of putative collagens in non-metazoan eukaryotes. In addition, collagen domains have been detected in several species of bacteria (Bachert et al., 2015; Charalambous et al., 1988; Ghosh et al., 2012; Kananavičiūtė et al., 2020; Lukomski et al., 2017; Mohs et al., 2007; Rasmussen et al., 2003; Yu et al., 2014) and viruses (Bamford and Bamford, 1990; La Scola et al., 2008; Luther et al., 2011; Medveczky et al., 1993; Raoult et al., 2004; Tidona and Darai, 1997; van Hulten et al., 2001; Yau et al., 2011; Zhang et al., 2004), although due to their evolutionary distance, these were typically hypothesized to have evolved independently from metazoan collagens (Rasmussen et al., 2003) or through horizontal gene transfer from metazoans (Luther et al., 2011; Rasmussen et al., 2003). Nonetheless, the growing number of reports has pointed to the possibility that collagens might be more ancient and/or more widespread than previously realized.

Therefore, to better understand the phylogenetic distribution and ancestry of collagen domains across the tree of eukaryotes, we analyzed recently sequenced genomes and transcriptomes of close relatives of metazoans, the choanoflagellates and other non-metazoan holozoans (Table 3S1), as well as those of diverse other eukaryotes representing many major clades. We found that collagen domains are common in the close relatives of metazoans, including choanoflagellates, many of which encode extended collagen domains comparable in length and proline content to metazoan collagens. Furthermore, we found that collagen domains are widely distributed across many eukaryotes that are only distantly related to metazoans, including Fungi, Archaeplastida, and members of the SAR clade. In contrast, canonical metazoan collagens such as fibrillar collagens and collagen IV are potentially restricted to metazoans and may help explain unique features of metazoan biology.

Box 1. Collagen terminology.

The criteria for what constitutes "a collagen" are not well defined (Garrone, 1999; Gay and Miller, 1983; Ricard-Blum, 2011). The major defining characteristic of a collagen is the collagen triple helix, but some authors restrict the term "a collagen" to mean a protein that plays a structural role in the ECM (Garrone, 1999; Gay and Miller, 1983), while others define "a collagen" as any protein containing a collagen triple helix, including intracellular proteins of the immune system (Casals et al., 2019; Fraser and Tenner, 2008) or transmembrane signaling proteins (Maertens et al., 2007). Here, we adopt the latter usage, for two reasons: first, many of the proteins we discuss have not been functionally characterized, and whether a given protein plays a structural role in the ECM is difficult to predict from sequence alone. Second, this usage acknowledges the possibility that collagens could evolve dynamically between structural and non-structural roles or between intracellular and extracellular localizations over evolutionary time. For brevity and simplicity, we use the term "collagen domain" to refer to any stretch of collagen triple helix repeats, i.e., $(Gly\text{-}X\text{-}Y)_n$.

Results

Collagen domains and conserved collagen-associated domains predate metazoan origins

To start, we focused our search on close relatives of metazoans, hypothesizing that they were most likely to illuminate the proximal ancestry of metazoan collagens (Fig. 3.1). From the 29 genomes and transcriptomes analyzed (Table 3S1), we found that all 22 choanoflagellates, three of four filastereans (*Ministeria vibrans, Pigoraptor vietnamica,* and *Pigoraptor chileana)* and one of two ichthyosporeans (*Sphaeroforma arctica*) encoded proteins with collagen domains (domains annotated by InterProScan as IPR008160). Like metazoan collagens, many of these proteins were predicted to carry transmembrane regions and/or N-terminal signal sequences (Fig. 3S1), suggesting that they are secreted into the ECM. Furthermore, all 29 species were found to encode putative prolyl 4-hydroxylase domains (IPR006620), the class of enzyme responsible for the post-translational modification of proline to 4-hydroxyproline, a step thought to be required for collagen triple helix formation in metazoans (Berg and Prockop, 1973; Canty and Kadler, 2005; Juva et al., 1966; Kao et al., 1979; Walmsley et al., 1999). (The strictness of this requirement has been challenged by observations of triple helix formation without hydroxyproline in some contexts, such as hymenopteran cocoons (Sutherland et al., 2013), recombinant non-metazoan cells (Olsen et al., 2001; Perret et al., 2001; Ruggiero et al., 2000), and bacteria (Yu et al., 2014).) Thus, 26 of the 29 non-metazoan holozoans that we analyzed appear to have the genetic machinery required to produce collagen triple helices that may be modified with 4-hydroxyproline.

At least two classes of collagens may have been present in the Urmetazoan (the last common ancestor of all metazoans) based on their broad phylogenetic distribution within metazoans. The first is fibrillar collagen, which is characterized by a long, uninterrupted collagen domain and a C-terminal COLF1 domain (Exposito et al., 2010). The second is collagen IV, which is characterized by a triple helix with many interruptions and a C-terminal NC1 domain (Boute et al., 1997; Exposito et al., 2002; Fidler et al., 2018, 2017). Included in this second class is the collagen IV variant spongin, which is distinguished by a truncated collagen domain and the lack of a conserved HSQ motif in its NC1 domain (Aouacheria et al., 2006; Fidler et al., 2017). To determine whether these key metazoan collagens may have predated the Urmetazoan, we searched our data set for proteins with collagen domains and COLF1 or NC1 domains. Prior investigations found that *S. rosetta* and *M. brevicollis* encode COLF1 domains homologous to the C-terminal domains of metazoan fibrillar collagens, but that these domains are not present in the same proteins as collagen domains (Hynes, 2012; King et al., 2008). We found that this pattern holds true across choanoflagellate diversity: COLF1 was detected in 12 of 22 choanoflagellate proteomes (Fig. 3.1), but none of the choanoflagellate COLF1 domains were present in proteins with collagen domains. COLF1 was not detected in any other holozoans. Thus, the COLF1 domain likely evolved on the Choanozoan stem lineage, but canonical fibrillar collagens appear to have evolved, presumably through domain shuffling, after the split between the choanoflagellate and metazoan lineages.

Next, we analyzed whether any non-metazoan holozoans encoded NC1 domains. Consistent with previous findings in *S. rosetta* and *M. brevicollis* (Fidler et al., 2017), we did not detect NC1 domains in any of the 22 choanoflagellate proteomes. Outside choanozoans, we detected an NC1 domain only in one protein encoded by the filasterean *Ministeria vibrans* (Fig.

3.1; Fig. 3S2C), as previously reported by Grau-Bové et al. (2017). Like metazoan collagen IV, this *M. vibrans* protein (hereafter "MvCN", for *Ministeria vibrans* Collagen + NC1) contains collagen domains and a C-terminal NC1 domain (Fig. 3S2C). This raises the question of whether MvCN might be homologous to collagen IV, and thus that collagen IV might be an ancient protein that predates the last common ancestor of filastereans and metazoans. To investigate this possibility more closely, we examined each of the domains of MvCN individually. By aligning the MvCN NC1 domain with metazoan collagen IVs, we found that it contains the conserved HSQ motif found in the NC1 domains of metazoan collagen IVs but not in those of spongins (Fig. 3S3B) (Fidler et al., 2017). Next, we examined the MvCN collagenous region. In contrast with many metazoan collagen IVs, the MvCN collagenous region is much shorter: it contains a total of 81 Gly-X-Y repeats, as opposed to the hundreds of Gly-X-Y repeats typical of metazoan collagen IVs: e.g., 443 repeats in *Mus musculus* (UniProt ID P02463), 264 repeats in the mollusk *Lottia gigantea* (UniProt ID V4A1B9), or 473 repeats in the cnidarian *Nematostella vectensis* (UniProt ID V9GW22). Finally, we examined its non-collagenous domain. Unlike metazoan collagen IV, MvCN contains a long, N-terminal non-collagenous domain of unknown function. Through BLAST search, we found that this domain shares sequence similarity with uncharacterized bacterial proteins (Fig. 3S3A), indicating that MvCN may be a eukaryotic-bacterial fusion protein. Thus, the NC1 domain appears to predate the divergence between filastereans and choanozoans, but whether MvCN shares a common ancestry with metazoan collagen IV remains unclear. If MvCN and collagen IV are homologous, this would mean that a collagen IV homolog was present in the last common ancestor of filastereans and choanozoans. Alternately, it is possible that an NC1 domain was present in a protein without collagen domains in a stem holozoan and became fused to collagen domains independently in the filasterean and metazoan lineages. Regardless, it appears that large collagen IVs—i.e., those containing hundreds of collagen triple helix repeats—are restricted to, and conserved within, metazoans. Functional experiments to test whether the large size of metazoan collagen IV is important to its capacity as a basement membrane scaffold, and/or whether MvCN is capable of forming collagen IV-like networks, could shed light on the evolution of collagen IV and the metazoan basement membrane.

Choanoflagellates encode collagens with diverse non-collagenous domains

Because choanoflagellates are the closest living relatives of metazoans and provide a unique window into the ancestry of metazoan genes, we examined the protein domain architecture of their collagens in more detail. The non-collagenous protein domain found most frequently in choanoflagellate collagens was the von Willebrand Factor A (VWFA) domain, a domain commonly found in metazoan ECM proteins such as collagen VI and FACIT collagens (Fig. 3S2D) (Ricard-Blum, 2011). VWFA domains often mediate protein-protein interactions, facilitating the binding of ECM proteins to each other (Hynes and Naba, 2012).

Interestingly, two choanoflagellate species encode collagens with domains that are also present in metazoan collagenous immune proteins. The choanoflagellate *Salpingoeca kvevrii* encodes a collagen with an SRCR domain, a combination found in some metazoan scavenger receptors that play a role in phagocytosis and innate immunity (Fig. 3S2D) (Gowen et al., 2001; Kodama et al., 1996; Neubauer et al., 2016; Yamada et al., 1998). *Microstomoeca roanoka* encodes a collagenous protein with a gamma fibrinogen domain, making it similar to ficolins, a

group of collagens that function in innate immune system activation in chordates (Endo et al., 2015; Matsushita, 2010). Because both species nest deeply within choanoflagellates, we posit that these choanoflagellate proteins evolved independently from metazoan scavenger receptors and ficolins. Nonetheless, the shared combination of protein domains raises the possibility that these collagens might play analogous roles in bacterial recognition, phagocytosis, or immunity in choanoflagellates in a manner similar to the metazoan defense collagens.

Because the choanoflagellate clade contains as much genetic diversity as its sister group, the Metazoa (Richter et al., 2018), we were interested in whether our data set could reveal patterns of collagen evolution within the choanoflagellate lineage. The diversity of collagen repertoires across choanoflagellates and the high degree of variability in collagen number and domain structure between closely related species suggested that collagens are rapidly evolving within choanoflagellates. For example, *Choanoeca flexa* was found to encode six collagens, whereas its closely related sister species *C. perplexa* encodes 16 collagens, and their next-closest relative, *M. brevicollis,* encodes only two (Fig. 3S1); many of these proteins have unique domain structures without clear orthologs in the other species. By contrast, some collagens appear to be conserved across choanoflagellate evolution. Mirroring the conserved collagens of metazoans (collagen IV and fibrillar collagens), we identified a possibly ancient collagen that is distributed widely within choanoflagellates. This protein, which we refer to as "ChCL" (for Choanoflagellate Collagenous Lectin) has a characteristic architecture of a ~20-amino acid N-terminal transmembrane domain followed by an extracellular region with a ~50-amino acid collagen domain, a ~700-amino acid non-collagenous domain, and a ~200-amino acid C-terminal Concanavalin A-like lectin/glucanase domain (IPR013320) (Fig. 3S4A). The non-collagenous domain of ChCL has sequence similarity with non-collagenous metazoan proteins of unknown function (Fig. 3S4B), while the C-terminal domain closely resembles genes of bacterial origin (such as MSX41286.1, an Actinobacterium protein of unknown function, or WP_007139689.1, a putative cell wall-associated protein in a Flavobacterium) (Fig. 3S4C). Based on its distribution and on the most recent understanding of choanoflagellate phylogenetic relationships, ChCL appears to have been present in the last common ancestor of all choanoflagellates (Fig. 3.1).

Collagen domains are widespread across eukaryotes

The finding that collagenous proteins are nearly ubiquitous in the closest living relatives of metazoans inspired us to broaden our search to include more distant lineages. To investigate the distribution of collagen domains across eukaryotic diversity, we took advantage of the proteomes available in the UniProtKB database (UniProt Consortium, 2021), which includes representatives of 288 families of non-metazoan eukaryotes across many, but not all, of the major eukaryotic groups (Fig. 3.2A). We detected proteins with collagen domains in nearly all major eukaryotic groups analyzed (Fig. 3.2B). For example, out of 169 fungal families in our analysis, 73 families encode at least one protein with a collagen domain (IPR008160) (Fig. 3.2C). Families encoding collagens were concentrated within Ascomycota, but we also detected collagens sparsely distributed in Basidiomycota, Zoopagomycota, Mucoromycota, Cryptomycota, and Microsporidia (Fig. 3S5B). Within Archaeplastida, 17 out of 71 families encode collagens (Fig. 3.2D). Nine of these 17 families are within land plants (Embryophyta), revealing that collagens are present in at least two lineages with complex multicellularity. In the SAR clade, 19 out of 31 families were found to encode collagens (Fig. 3.2E). The only major

eukaryotic group in which we detected no collagen domains was Haptista (Fig. 3.2F), which is represented by only two proteomes in our data set.

The low information-density of the collagen domain (Box 1) makes it difficult to infer whether the collagen domains found in diverse eukaryotic clades are homologous with metazoan collagens or whether they evolved convergently. However, we can ask if non-metazoan collagen domains resemble metazoan collagens in length and proline content, and whether the collagen-encoding species also encode a putative prolyl 4-hydroxylase. We found that most non-metazoan eukaryotic proteomes in our data set encode at least one protein with a putative prolyl 4-hydroxylase domain (Fig. 3S6); specifically, 84% of the 2,011 proteomes that encode collagens also encode a putative prolyl 4-hydroxylase. Whether an enzyme is capable of hydroxylating proline in a collagen-like substrate is difficult to predict from sequence alone, but such enzymatic activity outside metazoans would not be unprecedented: endogenous enzymes with collagen prolyl 4-hydroxylase activity have been reported in plants (Hieta and Myllyharju, 2002) and Fungi (de Bruin et al., 2002). Next, we compared the collagen repeat length across all collagens detected in metazoans (n = 57,104 collagens), choanoflagellates (n = 258 collagens), Archaeplastida (n = 75 collagens), Fungi (n = 508 collagens), and SAR (n = 176 collagens) (Fig. 3.3A; Fig. 3S7A). Metazoan collagens are the longest, with a median collagen repeat length of 192 amino acids (i.e., a median of 64 Gly-X-Y repeats per protein); choanoflagellate collagens have the second-longest, with a median collagen repeat length of 145.5 amino acids; and SAR collagens have the shortest, with a median of 85.5 amino acids. We then quantified proline content across the same data set (Fig. 3.3B; Fig. 3S7B) and found that metazoan collagen domains contain the most proline at the X and Y amino acid positions, with a median of 29.9% proline. Choanoflagellate collagens were the second-most proline-rich at 15.6%, while SAR collagens were the least proline-rich at a median of 0%. While the median metazoan collagen may have longer collagen domains and more proline than the median non-metazoan collagen, it is notable that the distributions of collagen repeat length and proline content overlap significantly across all these eukaryotic groups (Fig. 3.3). Together, these data suggest that some non-metazoans likely express collagen triple helices with similar structural qualities to metazoan collagens, which could therefore play similar roles in their biology.

Discussion

Extracellular proline-rich glycoproteins have long been known to exist in non-metazoan eukaryotes (Davies et al., 1997; Hallmann, 2006; Lamport, 1977; Perfect et al., 1998; Showalter, 1993; Sumper and Hallmann, 1998). However, collagen domains have rarely been reported in distant eukaryotic relatives of Metazoa. The furthest examples, to our knowledge, have been reports of collagen repeats in the genomes of the apicomplexan (SAR) parasite *Plasmodium falciparum* (Rasmussen et al., 2003) and the amoebozoan *Dictyostelium discoideum* (Fidler et al., 2017), and of cell-surface proteins with collagen repeats in a handful of species of Dikarya fungi (Celerin et al., 1996; de Bruin et al., 2002; Tuntevski et al., 2013; Wang and Leger, 2006). Our findings suggest that collagen domains may be a near-ubiquitous feature of holozoans (Fig. 3.1, Fig. 3S5C) and may also be widespread across eukaryotic diversity (Fig. 3.2). The collagen domain is characterized by a repeated triplet motif (Box 1) whose simplicity makes it a reasonable candidate for repeated convergent evolution (Chen et al., 1997; Stern, 2013). With the exception of metazoans (Fig. 3S5A), choanoflagellates (Fig. 3S5C), and perhaps ascomycete

Fungi (Fig. 3S5B), the distribution of collagen domains across all of the eukaryotic groups in our data set is notably scattered: many pairs of sister families differ from each other in whether they encode collagen domains (Fig. 3.2C-F). Such a pattern could indicate either rampant convergent evolution or rampant loss. Due to the inherent difficulty of aligning highly repetitive domains, it remains unclear how many times collagen domains may have evolved in the history of eukaryotes. Even if the collagen domains outside Metazoa evolved independently from those within Metazoa or Holozoa, it is nonetheless possible that proteins with collagen-like properties might be important to the biology of many organisms across eukaryotic diversity, not just metazoans.

The evolution of collagen has often been cited as a key prerequisite for metazoan complex multicellularity (Erwin, 1993; Fidler et al., 2018; Reinhard et al., 2016; Towe, 1970). Here, we found that while collagen domains are encoded by diverse eukaryotes, several features distinguish metazoan collagen repertoires from those of non-metazoans. Metazoan collagen domains tend to be slightly longer and more proline-rich than those of any other eukaryotic clade in our data set (Fig. 3.3). Furthermore, fibrillar collagen, a major component of metazoan connective tissues, is restricted to metazoans in our data set. Collagen IV, the scaffolding protein of metazoan basement membranes, has a murkier evolutionary history due to the presence of a potential homolog in *Ministeria vibrans;* however, large collagen IVs—i.e., those with hundreds of collagen triple helix repeats—appear to be restricted to, and conserved within, metazoans. The C-terminal domains of fibrillar collagens and collagen IV allow them to form highly oligomeric, cross-linked superstructures—fibrils and networks, respectively—that are key to their roles as the structural scaffold of metazoan ECM (Ricard-Blum, 2011). In addition to proline hydroxylation, metazoan collagens also undergo diverse, functionally important post-translational modifications, such as lysine hydroxylation and cross-linking (Eyre and Wu, 2005; Fidler et al., 2014; Rodriguez-Pascual and Slatter, 2016), proteolytic cleavage (Canty and Kadler, 2005; Ricard-Blum, 2011), and glycosylation (Hennet, 2019), and it remains to be seen whether non-metazoan collagens might undergo similar modifications. Thus, while collagen domains are not unique to Metazoa, it may be the case that innovations related to collagen, such as the origin of fibrillar collagens/large collagen IVs and their associated superstructures, increased modification with 4-hydroxyproline, and/or innovations in collagen-interacting and collagen-modifying proteins not addressed by our data set, facilitated the evolution of complex multicellularity during metazoan origins.

The ubiquity of collagen domains in choanoflagellates raises the question of what role collagens might be playing in single-celled and colonial relatives of metazoans. In a possible parallel to fibrillar collagen and collagen IV in metazoans, we detected an ancient collagenous lectin, ChCL, that is widespread across choanoflagellates and was likely present in the last common ancestor of all choanoflagellates. Choanoflagellates secrete extracellular matrices that are dynamically regulated across multiple life history stages and play functional roles in substrate adhesion (Dayel and King, 2014; Leadbeater, 2008), mating regulation (Woznica et al., 2017), and multicellular colony formation (Dayel et al., 2011; Larson et al., 2020; Levin et al., 2014). Furthermore, choanoflagellates interact closely with bacteria as predators (Dayel and King, 2014), as signaling partners (Alegado et al., 2012; Ireland et al., 2020; Woznica et al., 2017, 2016), and possibly as symbionts (Hake et al., 2021). We found that some choanoflagellates encode collagens with domains characteristic of metazoan defense collagens, such as scavenger receptors and ficolins, and that the choanoflagellate collagen ChCL contains a domain similar to putative bacterial cell wall proteins, suggestive of a potential role for these

collagens in mediating choanoflagellate-bacteria interactions. In general, while non-ECM collagens—i.e., the defense collagens (Casals et al., 2019; Fraser and Tenner, 2008), transmembrane collagens (Franzke et al., 2005; Loria et al., 2004; Maertens et al., 2007), and soluble collagenous signaling proteins (Leclère et al., 2020; Mikkola and Thesleff, 2003; Pyagay et al., 2005)—are relatively well characterized in vertebrates and in some invertebrate model organisms, little is known about non-ECM collagens in basal metazoans. This leaves a gap in our understanding of any ancient non-ECM collagens that may have been present in early metazoan evolution. Future work to characterize the function and localization of collagens in basal metazoans and close metazoan relatives, including choanoflagellates, may shed light on the early roles of collagens both within and outside of the ECM.

Materials and Methods

Identification of collagen domains and collagen-associated domains in non-metazoan holozoans

To catalogue putative prolyl 4-hydroxylase and collagen domains in holozoans, we obtained the predicted proteomes of 22 choanoflagellates, four filastereans, two ichthyosporeans, and one pluriformean (Table 3S1). We performed InterProScan annotation of each proteome using the AgBase InterProScan Singularity container (running InterProScan version 5.41-78.0; available at https://hub.docker.com/r/agbase/interproscan) on the University of California Berkeley Savio computational cluster to obtain Gene3D, Panther, Pfam, Phobius, PrositePatterns, SMART, and SuperFamily annotations of every protein. The Phobius data were used for transmembrane region predictions and signal sequence predictions. Any protein region annotated as IPR008160 ("Collagen triple helix repeat"; https://www.ebi.ac.uk/interpro/entry/InterPro/IPR008160/) was considered a collagen domain, and any region annotated as IPR006620 ("Prolyl 4-hydroxylase, alpha subunit"; https://www.ebi.ac.uk/interpro/entry/InterPro/IPR006620/) was considered a putative prolyl 4-hydroxylase domain. Domains annotated as IPR001442 ("Collagen IV, non-collagenous"; https://www.ebi.ac.uk/interpro/entry/InterPro/IPR001442/) or IPR036954 ("Collagen IV, non-collagenous domain superfamily"; https://www.ebi.ac.uk/interpro/entry/InterPro/IPR036954/) were considered NC1 domains, domains annotated as IPR000885 ("Fibrillar collagen, C-terminal"; https://www.ebi.ac.uk/interpro/entry/InterPro/IPR000885/) were considered COLF1 domains.

Identification of collagen domains and prolyl 4-hydroxylases in metazoans and non-holozoan eukaryotes

To characterize the distribution of putative prolyl 4-hydroxylase domains and collagen domains across non-holozoan eukaryotes and metazoans, we took advantage of the publicly available proteomes in the UniProtKB database. We first assembled a list of all eukaryotic proteomes available in the UniProtKB database as of June 30th, 2021 (UniProt release 2021_03; https://www.uniprot.org/) by searching the UniProtKB Proteomes database (https://www.uniprot.org/proteomes/) with the query

taxonomy:"Eukaryota [2759]"

and downloading the results. We filtered out low-quality proteomes using the following criteria: first, we discarded any proteomes with a protein count less than 200; second, we discarded proteomes classified by UniProt's Complete Proteome Detector (CPD) as "Outlier (low value)";

and third, we discarded proteomes with BUSCO scores lower than 50% (or, for proteomes with no BUSCO score provided, we discarded those with a protein count less than 20,000). These criteria were chosen in order to minimize inclusion of low-quality proteomes while retaining at least one proteome from each major eukaryotic group.

After filtering the proteomes, the "Taxonomic lineage" provided by UniProt for each proteome was then used as the basis for the topologies of all phylogenetic trees presented in this paper, including assignment of each organism to a taxonomic family. At this step, the following modifications were made: (1) We rearranged the base of the eukaryotic phylogeny to match the most up-to-date eukaryotic tree of life based on Burki et al. (2020), and the base of the metazoan phylogeny to reflect known relationships between phyla. (2) We discarded proteomes from organisms labeled "fungal sp. no. X" (where X is a species number) due to unclear phylogenetic placement. (3) For 40 species whose UniProt taxonomic lineage did not include a family name, a family name and taxonomic lineage were assigned from the Global Biodiversity Information Facility (GBIF) database (https://www.gbif.org/); in the 11 out of 40 cases where the GBIF database did not list a family name due to phylogenetic uncertainty, the organism and its proteome was discarded.

Having assembled our eukaryotic phylogeny and list of proteomes for analysis, we next searched for collagen domains in metazoans and non-holozoan eukaryotes by searching the UniProtKB protein database (https://www.uniprot.org/uniprot/) with the following queries:

1. NOT taxonomy:"Opisthokonta [33154]" IPR008160 taxonomy:"Eukaryota [2759]"
2. taxonomy:"Fungi [4751]" IPR008160
3. taxonomy:"Metazoa [33208]" IPR008160

We downloaded all hits for these three queries, then discarded any hits whose UniProt Organism ID was not present in our list of filtered proteomes. To catalogue putative prolyl 4-hydroxylase domains in metazoans and non-holozoan eukaryotes, we repeated the above procedure replacing "IPR008160" with "IPR006620".

Quantification of collagen repeat length and proline content

To compare the characteristics of collagens across eukaryotic diversity, we analyzed the collagen repeat length and proline content of collagens in different eukaryotic clades. Collagen repeat length was assessed by quantifying, for each protein, the total number of triplet repeats matching the pattern $(Gly\text{-}X\text{-}Y)_n$, where $n \geq 2$ and X and Y can be any amino acid. The resulting value was either expressed in units of repeats or multiplied by 3 when expressed in units of amino acids. Proline content was assessed by dividing the total number of prolines found in the X and Y positions of those triplet repeats by twice the total number of repeats. For example, a protein with the sequence MGPPGNNGPNNNNPGNPNNNNGPNGPN* has a "collagen repeat length" of 5 repeats or 15 amino acids, and its "proportion of proline in X and Y positions" is 50%.

Analysis of MvCN domains

We investigated the non-collagenous domains of *Ministeria vibrans* MvCN through a combination of BLAST search and amino acid alignment. First, we conducted a search of the NCBI BLAST database (https://blast.ncbi.nlm.nih.gov/Blast.cgi) using the MvCN N-terminal non-collagenous domain (amino acids 1-731) as a query. This resulted in three hits, NBV83445.1, NTW51976.1, and WP_161718411.1, all of which are bacterial proteins. Alignment of the MvCN N-terminal domain against these hits (Fig. 3S3) was performed using

the Geneious alignment algorithm in Geneious Prime 2021.1.1 (https://www.geneious.com). We aligned the MvCN NC1 domain (amino acids 938-1122) against the NC1 domains of collagen IV from the ctenophore *Mnemiopsis leidyi* (protein ML18198a-PA) and the sponge *Oscarella carmela* (protein Ocar_m.306941) using the same alignment procedure (Fig. 3S3).

Analysis of ChCL domains

To investigate the origins of the conserved choanoflagellate collagen ChCL, we first performed a search of the NCBI BLAST database using the entire amino acid sequence of the representative *Salpingoeca urceolata* ChCL protein (Surc_m.480486) as a query. This yielded as hits many metazoan proteins that aligned to the middle non-collagenous domain of the protein (amino acids ~200-900). We aligned *S. urceolata* ChCL to the top hit, XP_013383323.1, as described above (Fig. 3S4). We then performed a second search with only the C-terminal domain of *S. urceolata* ChCL (amino acids 920-1136) as a query, which resulted in predominantly bacterial hits, and we aligned the *S. urceolata* ChCL C-terminal domain to the top hit, MSX41286.1, as described above (Fig. 3S4).

FIGURES

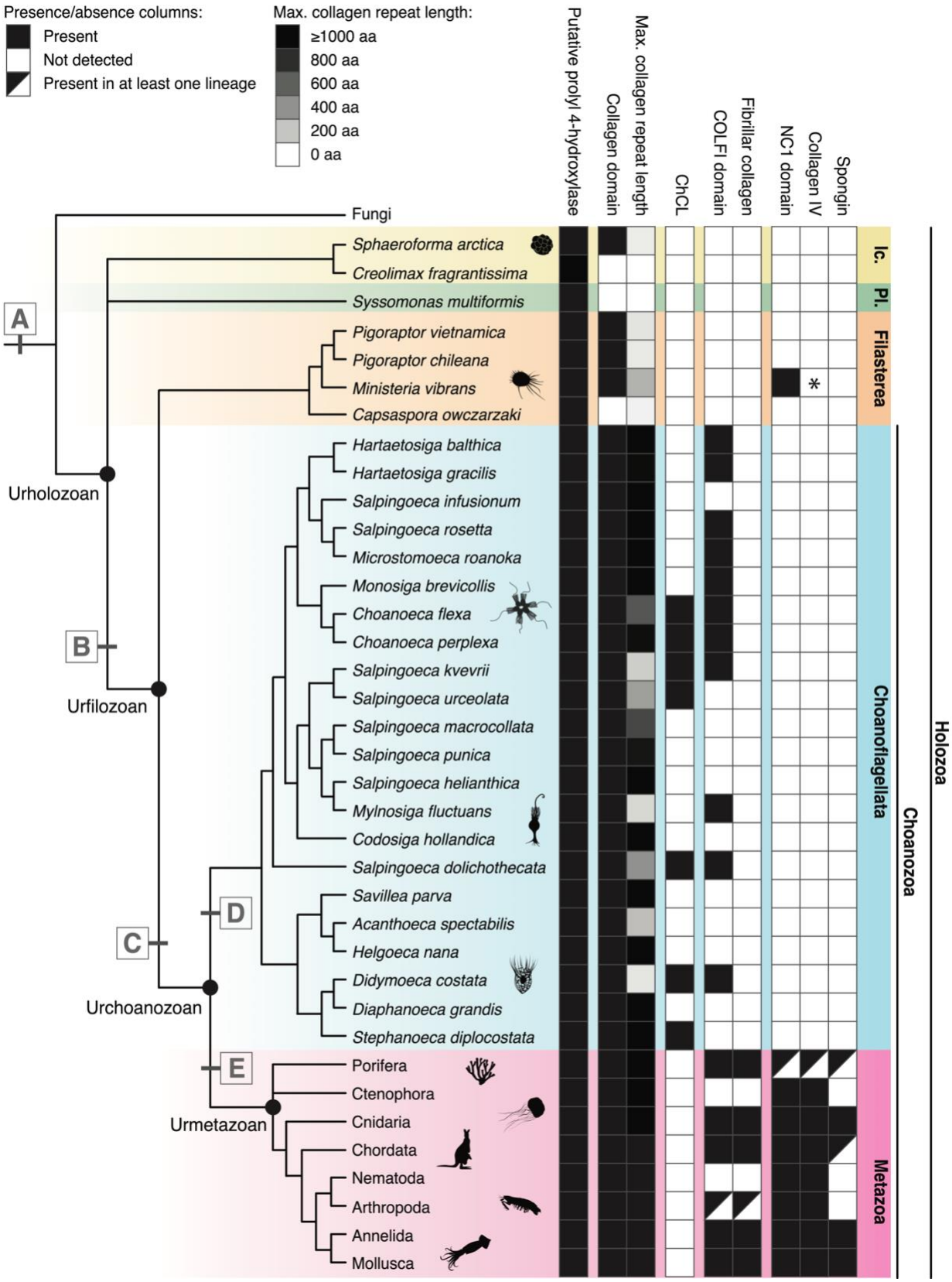
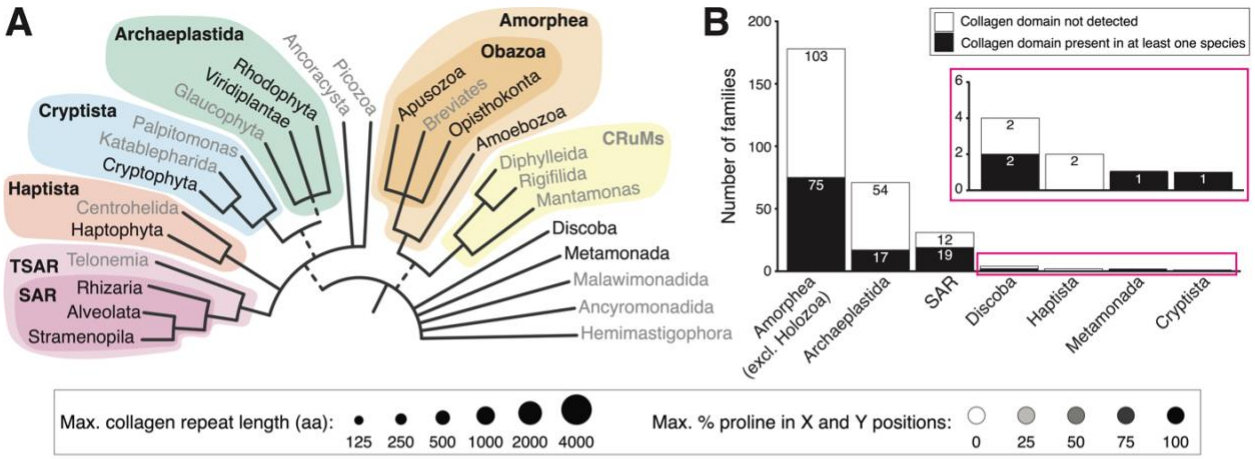
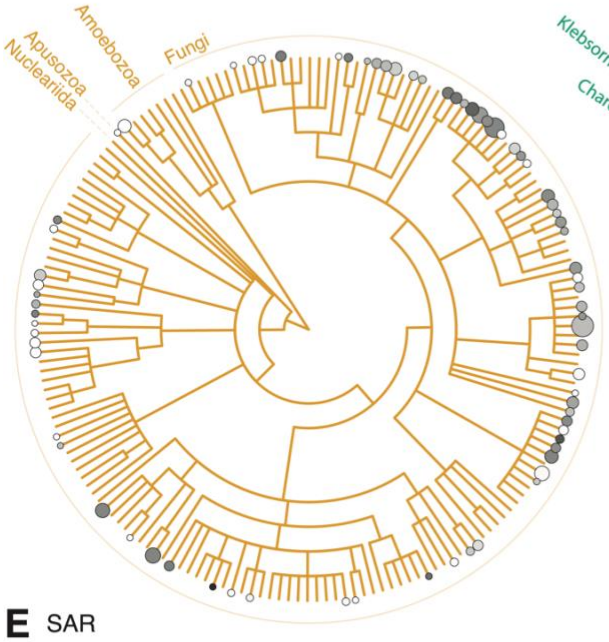
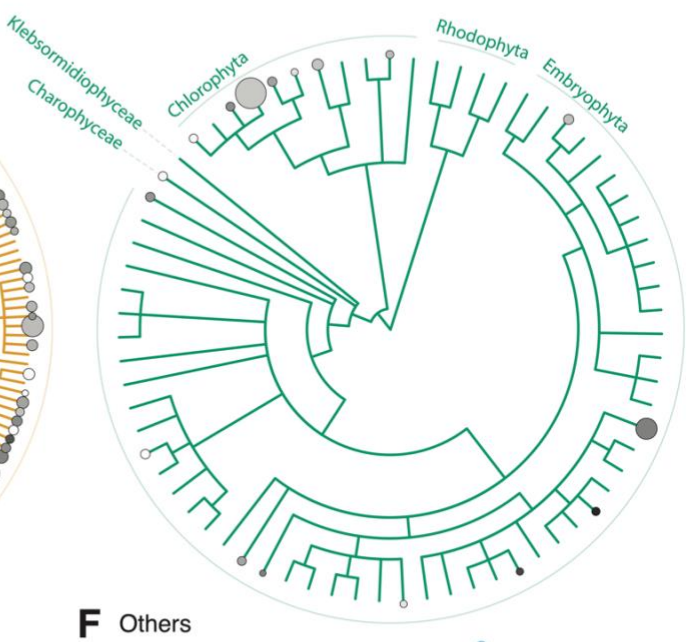
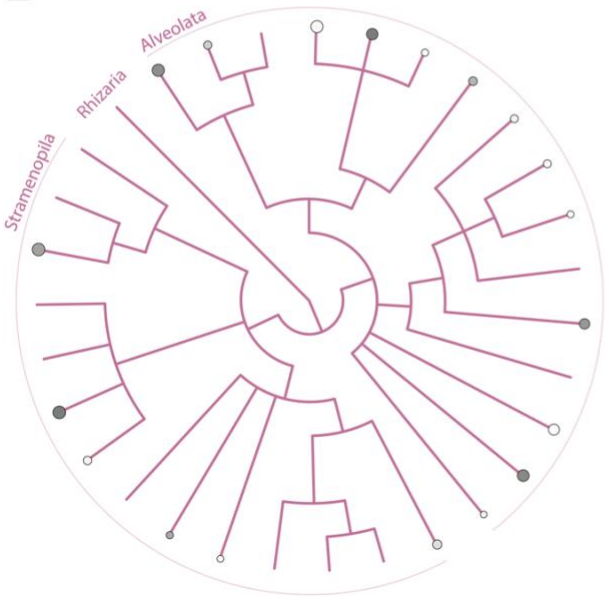

Presence/absence columns:
Present
Not detected
Present in at least one lineage
Max. collagen repeat length:
≥1000 aa
800 aa
600 aa
400 aa
200 aa
0 aa
Putative prolyl 4-hydroxylase
Collagen domain
Max. collagen repeat length
ChCL
COLFI domain
Fibrillar collagen
NC1 domain
Collagen IV
Spongin
Ic.
Pl.
Filasterea
Choanoflagellata
Choanozoa
Metazoa
Holozoa
Fungi
Sphaeroforma arctica
Creolimax fragrantissima
Syssomonas multiformis
Pigoraptor vietnamica
Pigoraptor chileana
Ministeria vibrans
Capsaspora owczarzaki
Hartaetosiga balthica
Hartaetosiga gracilis
Salpingoeca infusionum
Salpingoeca rosetta
Microstomoeca roanoka
Monosiga brevicollis
Choanoeca flexa
Choanoeca perplexa
Salpingoeca kvevrii
Salpingoeca urceolata
Salpingoeca macrocollata
Salpingoeca punica
Salpingoeca helianthica
Mylnosiga fluctuans
Codosiga hollandica
Salpingoeca dolichothecata
Savillea parva
Acanthoeca spectabilis
Helgoeca nana
Didymoeca costata
Diaphanoeca grandis
Stephanoeca diplocostata
Porifera
Ctenophora
Cnidaria
Chordata
Nematoda
Arthropoda
Annelida
Mollusca
A
B
C
D
E
Urholozoan
Urfilozoan
Urchoanozoan
Urmetazoan

Figure 3.1 Collagen domains and collagen-associated domains are conserved in close relatives of Metazoa. We detected collagen domains (IPR008160) and putative prolyl 4-hydroxylase domains (IPR006620) in the predicted proteomes of all metazoans and choanoflagellates and in most other holozoans analyzed, and found that the distributions of other collagen-associated domains such as COLF1 (IPR000885) and NC1 (IPR001442/IPR036954) are also consistent with pre-metazoan origins. Holozoans include choanoflagellates, filastereans, pluriformeans ("Pl.") and ichthyosporeans ("Ic."). Collagen length in choanoflagellates rivals that of metazoans, as shown here in the "Max. collagen repeat length" column, in which the maximum collagen repeat length of any protein detected in the species' predicted proteome is indicated by degree of shading (see key and Methods). (A – E) Mapping the distribution of collagen domains and key collagen-associated domains helps resolve the pre-metazoan evolution of collagens. (A) Putative prolyl 4-hydroxylase domains, which in metazoans perform a post-translational modification step important for collagen triple helix formation, were detected in all 29 non-metazoan holozoans. This domain is also present in many non-holozoan eukaryotes (Fig. 3S6), and thus might predate the divergence between Fungi and Holozoa. (B) The NC1 domain characteristic of metazoan collagen IV was detected in a collagen domain-containing protein (MvCN) of the filasterean *Ministeria vibrans* as previously reported (Grau-Bové et al., 2017), but not in any other holozoans, including choanoflagellates. Thus, the NC1 domain was likely present in the common ancestor of filastereans and choanozoans. (C) COLF1 domains, which are diagnostic of metazoan fibrillar collagens, are broadly distributed across choanoflagellates and metazoans, indicating that they predate the Urchoanozoan. However, in choanoflagellates, the COLF1 domain was not detected in any proteins that also contained collagen domains. (D) ChCL, a choanoflagellate-specific transmembrane protein consisting of a collagen domain and a lectin/glucanase domain, is broadly distributed across the choanoflagellate phylogeny, suggesting that it was likely present in the last common ancestor of choanoflagellates. (E) Fibrillar collagens (those containing COLF1 and collagen domains) appear to be restricted to metazoans. Likewise, canonical collagen IVs (proteins with >100 Gly-X-Y repeats and a C-terminal NC1 domain with a conserved HSQ motif) and spongins (truncated collagen IV variants lacking the conserved HSQ motif) are also restricted to metazoans. *: The MvCN protein is potentially homologous to metazoan collagen IV; however, their relationship is difficult to determine for reasons discussed in the main text. Figure format is adapted with permission from Figure 1 of Brunet et al. (2017). Consensus phylogeny of choanoflagellate relationships is based on Richter et al. (2018).


A
Archaeplastida
Rhodophyta
Viridiplantae
Glaucophyta
Ancoracysta
Picozoa
Amorphea
Obazoa
Apusozoa
Breviates
Opisthokonta
Amoebozoa
CRuMs
Diphylleida
Rigifilida
Mantamonas
Cryptista
Palpitomonas
Katablepharida
Cryptophyta
Haptista
Centrohelida
Haptophyta
TSAR
Telonemia
SAR
Rhizaria
Alveolata
Stramenopila
Discoba
Metamonada
Malawimonadida
Ancyromonadida
Hemimastigophora
B
Collagen domain not detected
Collagen domain present in at least one species
Number of families
200
150
100
50
0
103
75
54
17
12
19
6
4
2
0
2
2
2
1
1
Amorphea (excl. Holozoa)
Archaeplastida
SAR
Discoba
Haptista
Metamonada
Cryptista
Max. collagen repeat length (aa):
125
250
500
1000
2000
4000
Max. % proline in X and Y positions:
0
25
50
75
100


C Amorphea (excl. Holozoa)
Apusozoa
Nucleariida
Amoebozoa
Fungi


D Archaeplastida
Klebsormidiophyceae
Charophyceae
Chlorophyta
Rhodophyta
Embryophyta


E SAR
Stramenopila
Rhizaria
Alveolata

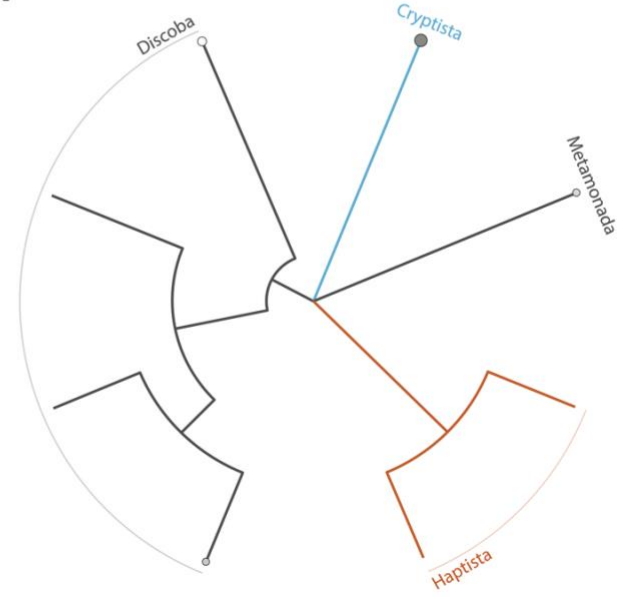

F Others
Discoba
Cryptista
Metamonada
Haptista

Figure 3.2 Collagen domains are detected throughout eukaryotic diversity. Having found that collagens are widespread in holozoans, we broadened our search to investigate whether collagen domains are found in other eukaryotic taxa. Our analysis included 2,598 eukaryotic proteomes present in the UniProtKB database that met quality criteria based on BUSCO score and protein count (see Methods). (A) The current consensus eukaryotic tree. The UniProtKB database contains proteomes from many branches of the eukaryotic tree of life, but not all. Clades shown in black are represented by at least one proteome in UniProtKB, while clades in gray are unrepresented. Dashed lines represent uncertainty about the monophyly of the indicated groups. (B) For each major eukaryotic group, the number of analyzed families in which a collagen domain was detected (black bar) or not detected (white bar) is shown (y-axis). For the Amorphea supergroup, which contains the Holozoa, we only show data for non-holozoans. Collagen domains were detected in at least one family from all eukaryotic groups analyzed, with the exception of Haptista. Even when omitting Holozoa, proteomes from Amorphea (particularly Fungi; see panel C) are heavily overrepresented in the UniProtKB database. Meanwhile, several major eukaryotic groups are represented by <10 families each (inset). (C to F) Collagen domains are broadly distributed not just across, but within, all major eukaryotic groups. In these trees, each stem represents one taxonomic family for which which at least one proteome meeting our quality criteria is present in the UniProtKB database. The presence of a circle indicates that at least one species in that family encodes a collagen domain (IPR008160). The size of the circle represents the maximum collagen repeat length of any collagen in the family, while the shading of the circle represents the maximum proline content of any collagen in the family (see Methods). Note that the maximum collagen repeat length and maximum proline content may be from different proteins. (C) Distribution of collagen domains across Amorphea families. Holozoans were excluded in order to highlight that collagen domains are broadly distributed even in distant relatives of metazoans. (D) Distribution of collagen domains across Archaeplastida families. The chlorophyte *Coccomyxa subellipsoidea* encodes the longest collagen domain detected in our entire data set, with a collagen repeat length of >4000 amino acids. (E) Distribution of collagen domains across SAR families. Collagen domains were detected in Stramenopiles and Alveolates, but not Rhizarians. (F) Distribution of collagen domains across the families of the other major groups of eukaryotes: Haptista, Cryptista, Metamonada, and Discoba. Haptista, which is represented by only two families in our analysis, is the only eukaryotic group for which no collagen domains were detected. Panel A was adapted with permission from Figure 1 of Burki et al. (2020).

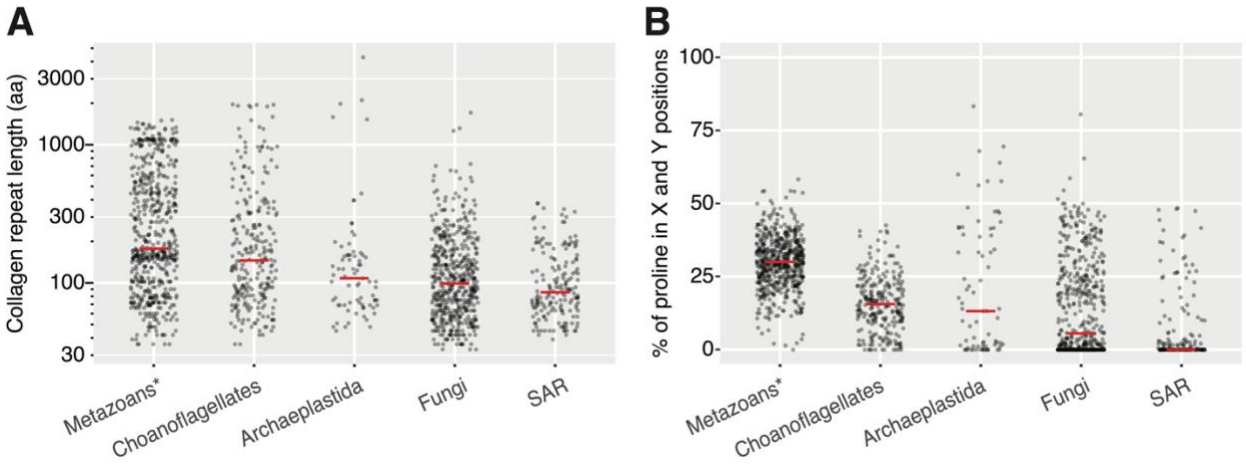
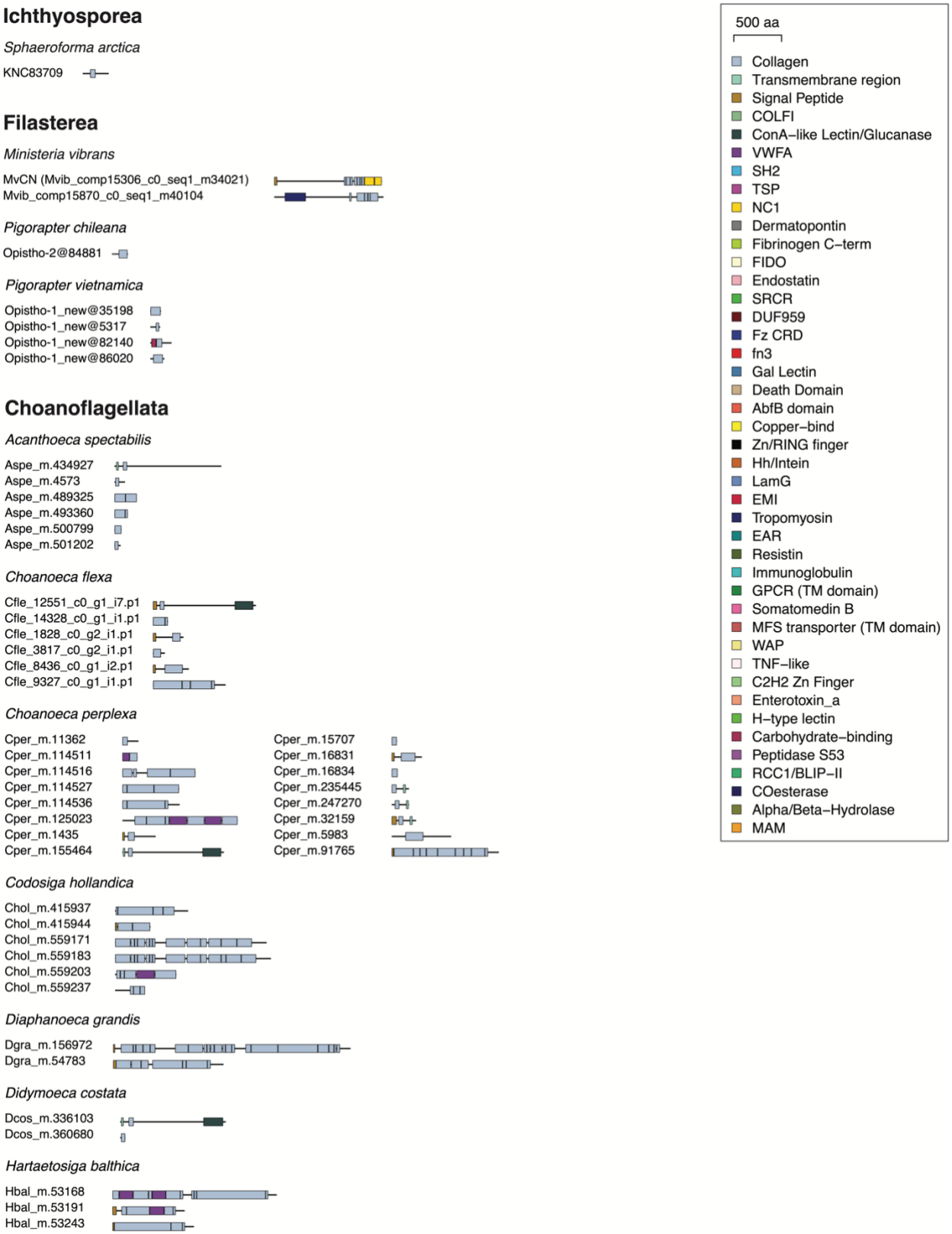
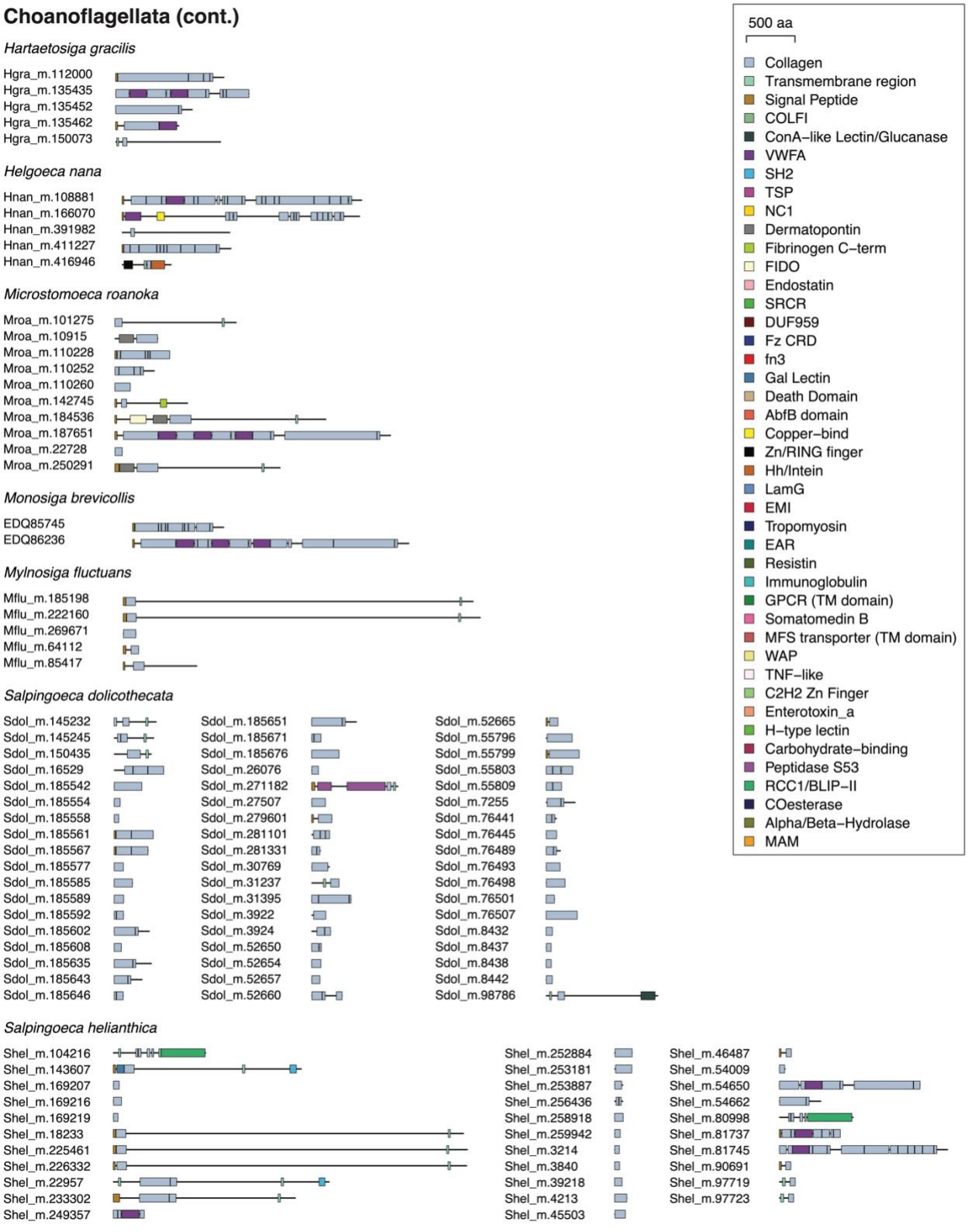


Figure 3.3 Metazoans encode collagens with more collagen repeats and higher proline content than other eukaryotes. To compare the characteristics of collagen domains across eukaryotic taxa, we examined collagen repeat length and proline content (see Methods) across all collagens detected in all analyzed species of choanoflagellates (n = 258 collagens), metazoans (n = 57,104 collagens), Archaeplastida (n = 75 collagens); fungi (n = 508 collagens); and SAR (n = 176 collagens). (A) Metazoans have the highest total collagen repeat length per protein (media$n$ = 192 amino acids), choanoflagellates the second-highest (media$n$ = 145.5 amino acids), and SAR the lowest (media$n$ = 85.5 amino acids). Note that the y-axis is on a log-scale. (B) Metazoan collagens have the highest proline content (media$n$ = 29.9%), choanoflagellates have the second-highest (media$n$ = 15.6%), and SAR the lowest (media$n$ = 0%). Thus, metazoan collagens are characterized by a greater number of repeats and higher proline content compared to those of other eukaryotic taxa; however, it is notable that the distributions of these characteristics across all five taxa are highly overlapping. Red bars represent medians. *: to more clearly visualize the distribution of the metazoan collagen data, a randomly selected 1% of metazoan collagens are shown here. For comparison of collagen repeat length and proline content across all eukaryotic groups analyzed, see Figure S7.

SUPPLEMENTARY MATERIALS



Choanoflagellata (cont.)



Choanoflagellata (cont.)

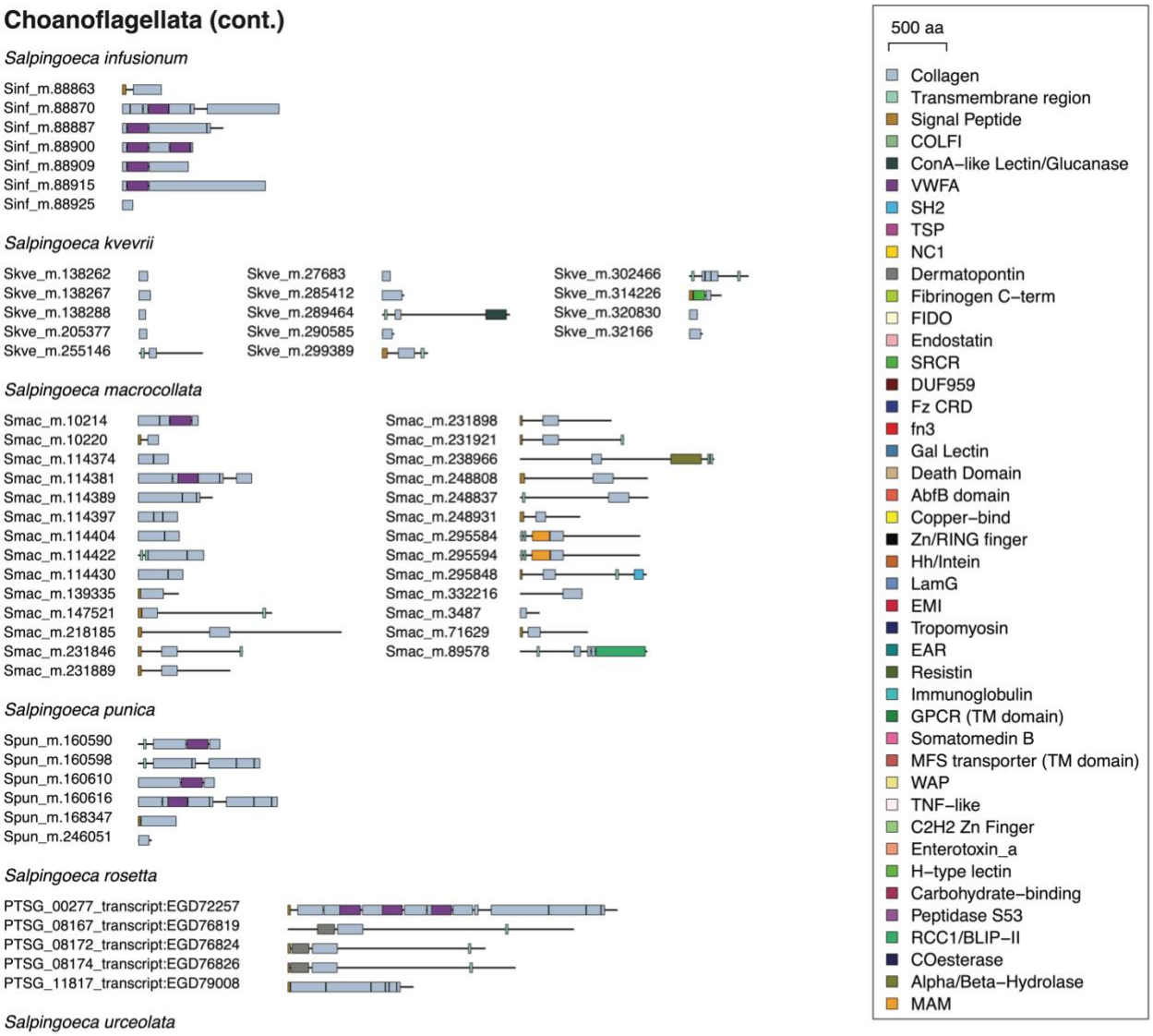


Figure 3S1 Collagen repertoires of 29 non-metazoan holozoans. Schematic representations of all collagens proteins detected in 22 choanoflagellates, 4 filastereans, 2 ichthyosporeans, and 1 pluriformean using InterProScan annotation. Protein diagrams were generated in R using the make_prot_fig script (Linden, 2021; https://github.com/tesslinden/interpro-scripts).

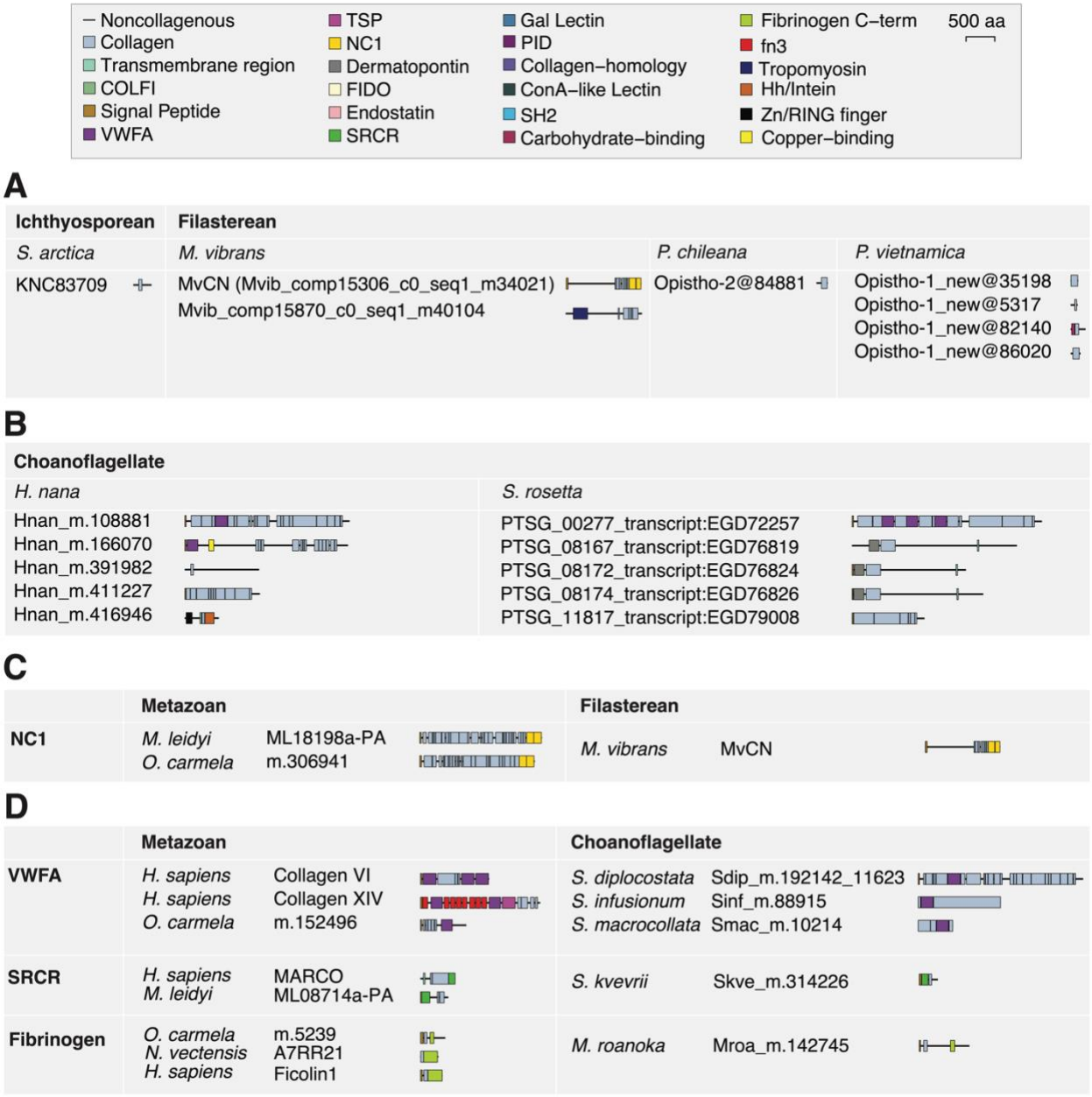
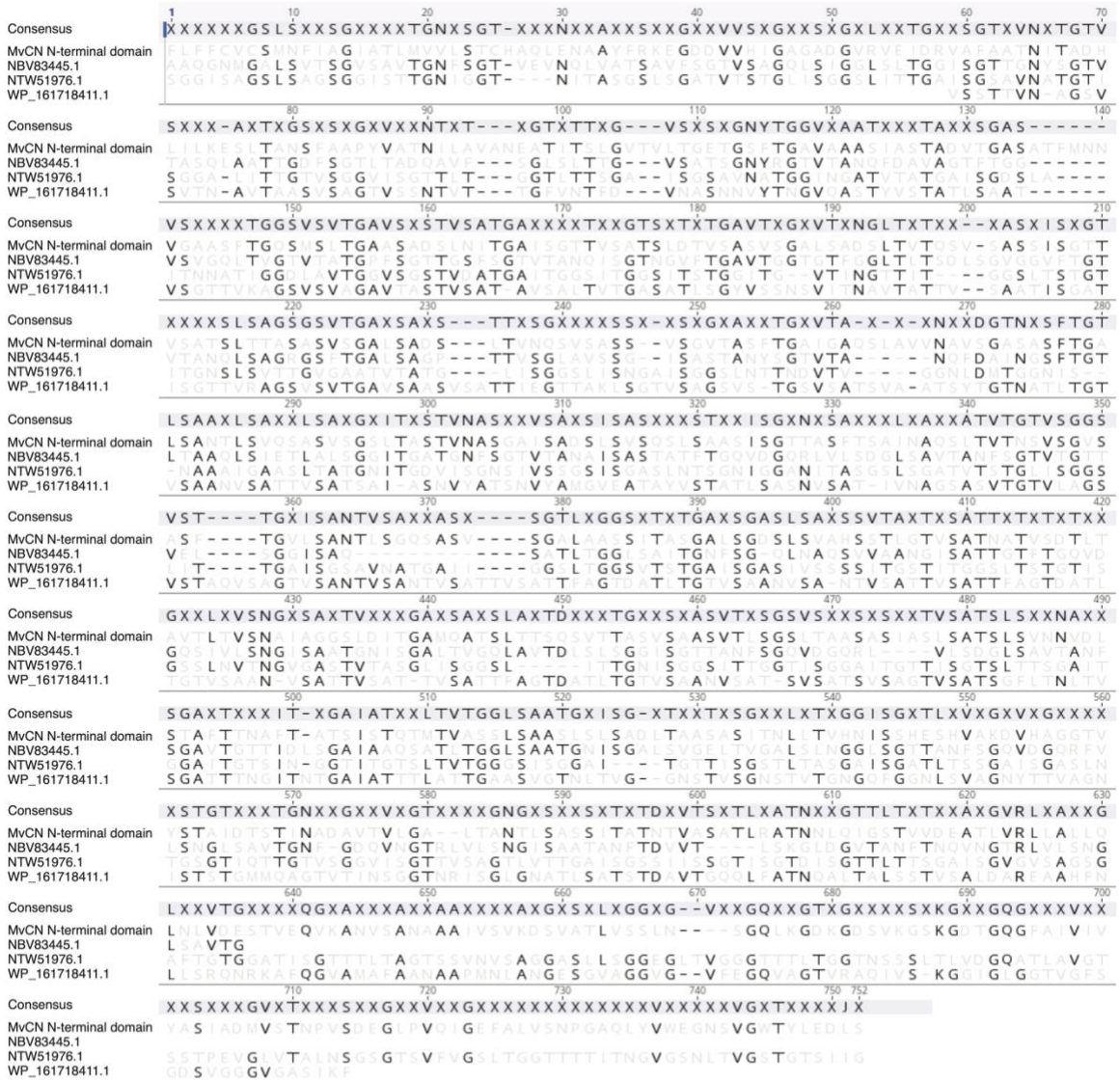


Figure 3S2 Some holozoan collagens contain known metazoan collagen-associated domains. (A) Schematic representations of the complete collagen repertoires of the ichthyosporean *Sphaeroforma arctica* and the filastereans *Ministeria vibrans, Pigoraptor chileana,* and *Pigoraptor vietnamica.* (B) Schematic representations of the complete collagen repertoires of the representative loricate choanoflagellate *Helgoeca nana* and the representative craspedid choanoflagellate *Salpingoeca rosetta*. (C) The filasterean *Ministeria vibrans* encodes a protein (MvCN) with collagen domains and an NC1 domain, shown here next to two examples of metazoan collagen IV. (D) Many choanoflagellates encode collagens with VWFA domains, a common collagen-associated domain in metazoans. Note that not all choanoflagellate VWFA domain-containing collagens are shown here. The choanoflagellate *Salpingoeca kvevrii* encodes a collagen with an SRCR domain, similar to metazoan scavenger receptors. The choanoflagellate

Microstomoeca roanoka encodes a collagen with a Fibrinogen domain, similar to metazoan ficolins. *Mnemiopsis leidyi* protein sequences were obtained from Ensembl Metazoa (https://metazoa.ensembl.org/Mnemiopsis_leidyi/). *Oscarella carmela* protein sequences were obtained from Compagen (http://www.compagen.org/datasets.html). *Homo sapiens* and *Nematostella vectensis* protein sequences were obtained from UniProtKB (https://www.uniprot.org/uniprot/). Protein diagrams were generated in R using the make_prot_fig script (Linden, 2021; https://github.com/tesslinden/interpro-scripts).

A


Consensus
MvCN N-terminal domain
NBV83445.1
NTW51976.1
WP_161718411.1

B

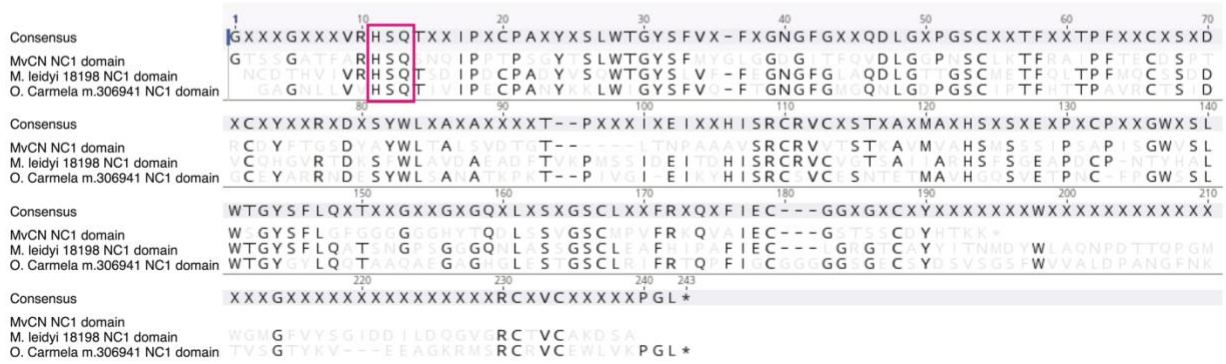

Consensus
MvCN NC1 domain
M. leidyi 18198 NC1 domain
O. Carmela m.306941 NC1 domain

Figure 3S3 The *Ministeria vibrans* collagen MvCN is a bacterial fusion protein. (A) Alignment of the N-terminal non-collagenous domain of MvCN (amino acids 1-731) against its top three BLAST hits, all of which are bacterial proteins. (B) Alignment of the MvCN NC1 domain (amino acids 938-1,122) against the NC1 domains of collagen IV in two basal metazoans, the ctenophore *Mnemiopsis leidyi* (protein ML18198a-PA) and the sponge *Oscarella carmela* (protein m.306941). The pink box highlights the conserved HSQ motif that is found in the NC1 domains of collagen IVs but not those of spongins (Fidler et al., 2017). Alignments were performed using the Geneious alignment algorithm in Geneious Prime 2021.1.1 (https://www.geneious.com).

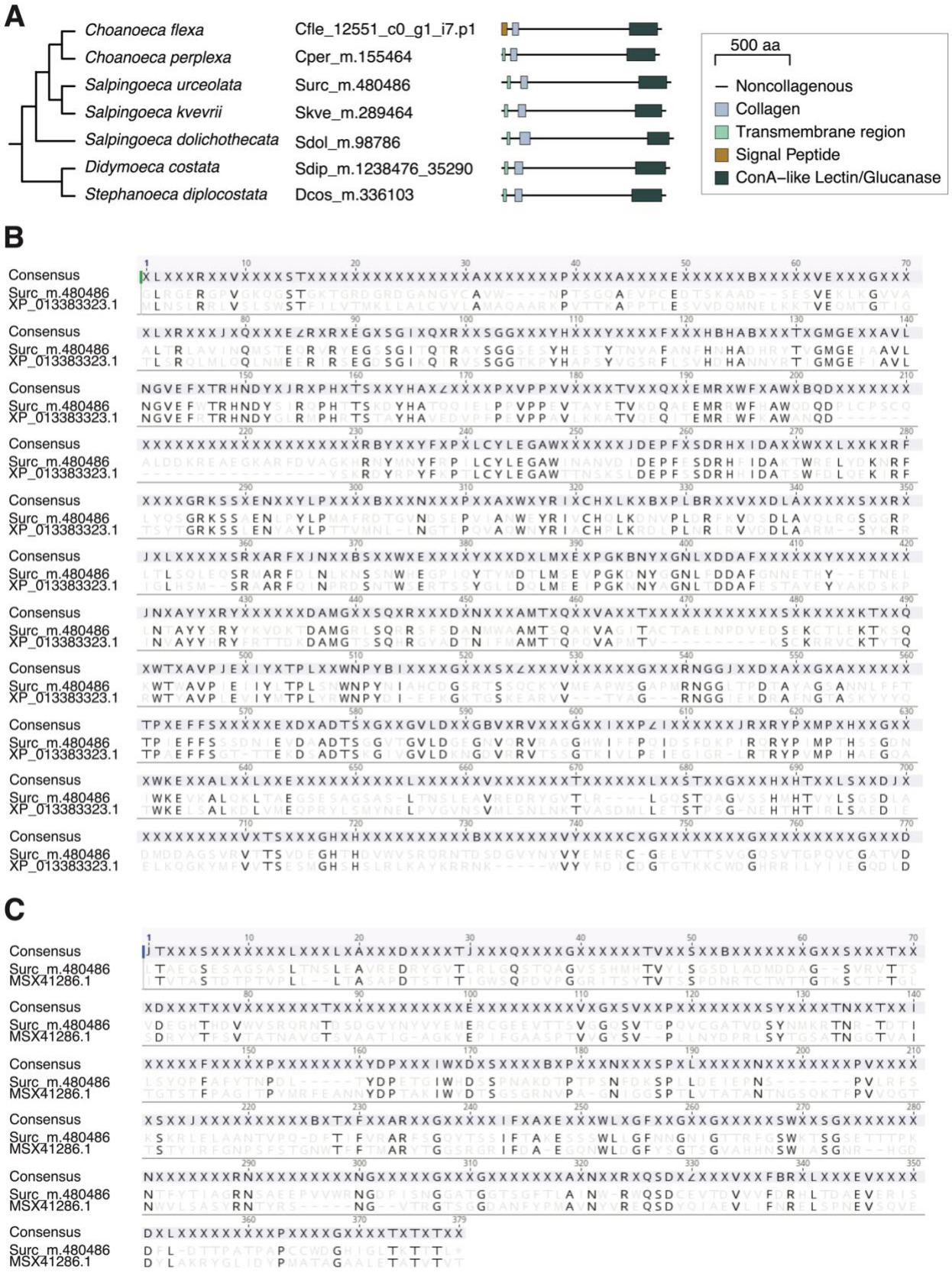

A
Choanoeca flexa
Choanoeca perplexa
Salpingoeca urceolata
Salpingoeca kvevrii
Salpingoeca dolichothecata
Didymoeca costata
Stephanoeca diplocostata
Cfle_12551_c0_g1_i7.p1
Cper_m.155464
Surc_m.480486
Skve_m.289464
Sdol_m.98786
Sdip_m.1238476_35290
Dcos_m.336103
500 aa
Noncollagenous
Collagen
Transmembrane region
Signal Peptide
ConA–like Lectin/Glucanase
B
Consensus
Surc_m.480486
XP_013383323.1
C
Consensus
Surc_m.480486
MSX41286.1

Figure 3S4 The choanoflagellate-specific collagen ChCL is a bacterial fusion protein. (A) Schematic representations of all seven ChCL proteins detected across 22 species of choanoflagellates, most of which are predicted to have an N-terminal transmembrane region, and all of which have an extracellular region consisting of a collagen domain, a non-collagenous domain, and a C-terminal Concanavalin A-like lectin/glucanase domain. These species are broadly distributed across the choanoflagellate phylogeny (see Figure 1), suggesting that ChCL may have been present in the most recent common ancestor of choanoflagellates. Protein diagrams were generated in R using the make_prot_fig script (Linden, 2021; https://github.com/tesslinden/interpro-scripts). (B) Alignment of the non-collagenous region (amino acids 144-899) of a representative ChCL from *Salpingoeca urceolata* against its top BLAST hit, a non-collagenous metazoan protein from *Lingula anatina*. (C) Alignment of the C-terminal domain (amino acids 776-1,135) of *S. urceolata* ChCL against its top BLAST hit, a protein from an Actinobacterium. Alignments were performed using the Geneious alignment algorithm in Geneious Prime 2021.1.1 (https://www.geneious.com).

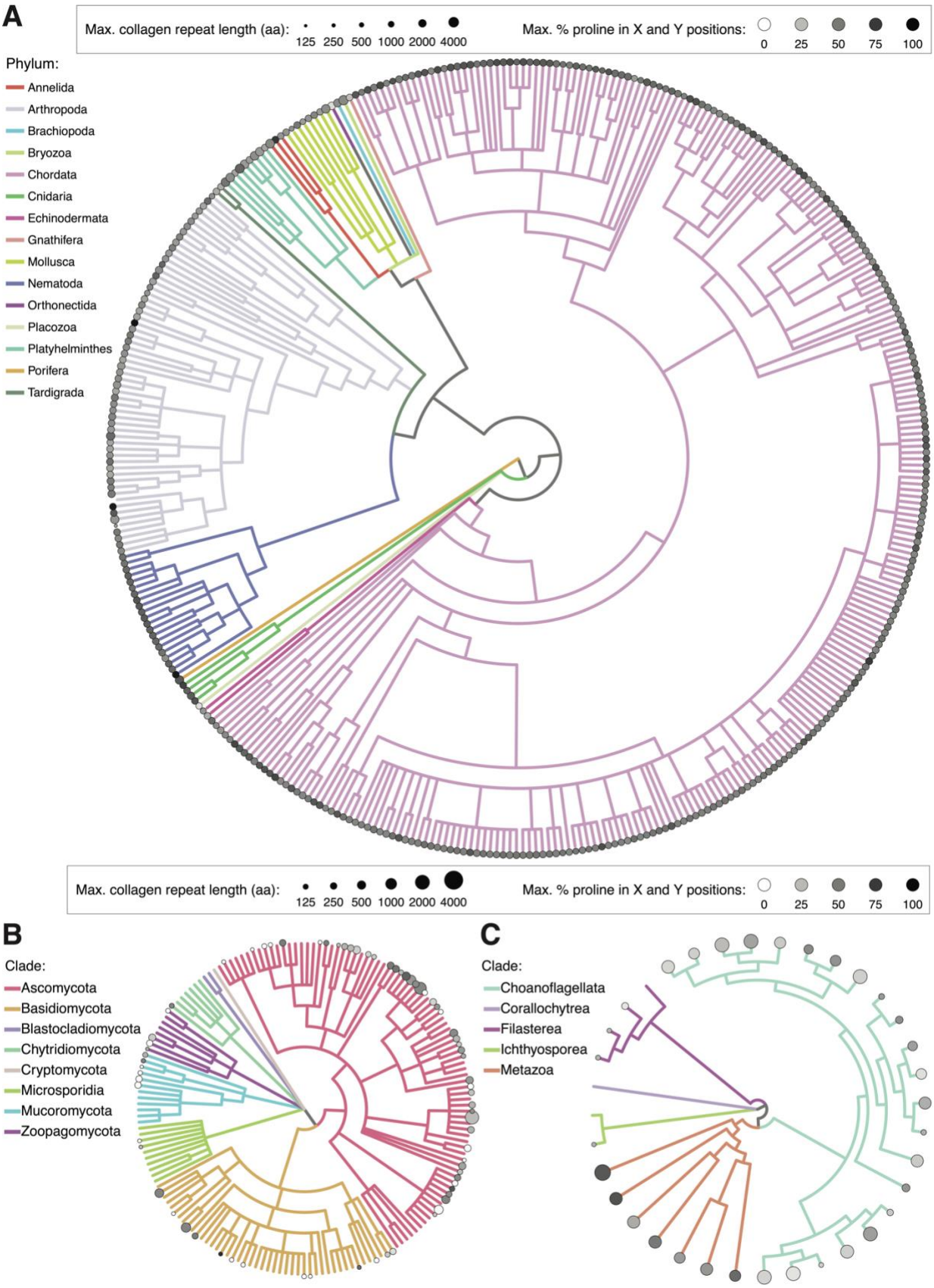

A
Max. collagen repeat length (aa):
125 250 500 1000 2000 4000
Max. % proline in X and Y positions:
0 25 50 75 100
Phylum:
Annelida
Arthropoda
Brachiopoda
Bryozoa
Chordata
Cnidaria
Echinodermata
Gnathifera
Mollusca
Nematoda
Orthonectida
Placozoa
Platyhelminthes
Porifera
Tardigrada
Max. collagen repeat length (aa):
125 250 500 1000 2000 4000
Max. % proline in X and Y positions:
0 25 50 75 100
B
Clade:
Ascomycota
Basidiomycota
Blastocladiomycota
Chytridiomycota
Cryptomycota
Microsporidia
Mucoromycota
Zoopagomycota
C
Clade:
Choanoflagellata
Corallochytrea
Filasterea
Ichthyosporea
Metazoa

Figure 3S5 Collagen domains across metazoan, fungal, and holozoan diversity. (A,B) In these trees, each stem represents one taxonomic family for which at least one proteome meeting our quality criteria (see Methods) is present in the UniProtKB database. The presence of a circle indicates that at least one species in that family encodes a collagen domain (IPR008160). The size of the circle represents the maximum collagen repeat length of any collagen in the family, while the shading of the circle represents the maximum proline content of any collagen in the family (see Methods). Note that the maximum collagen repeat length and maximum proline content may be from different proteins, and that the length scale is different between the top and bottom panels, as indicated in the key. (A) Collagen domains across metazoan families, color-coded by phylum. (B) Collagen domains across fungal families, color-coded by phylum. (C) In this tree of holozoan species, each stem represents a single species whose proteome was investigated for collagen domains. The presence of a circle indicates that the species encodes a collagen domain. The size of the circle represents the maximum collagen repeat length of any collagen in the species, while the shading of the circle represents the maximum proline content of any collagen in the species; again, note that the maximum collagen repeat length and maximum proline content may be from different proteins. The seven metazoan species shown here (*Daphnia pulex, Branchiostoma floridae, Nematostella vectensis, Strongylocentrotus purpuratus, Amphimedon queenslandica, Capitella teleta,* and *Lottia gigantea)* were subsetted from our UniProtKB analysis and were selected to represent seven maximally diverse phyla while also maximizing ancestral gene family retention based on the analysis of Richter et al. (2018).

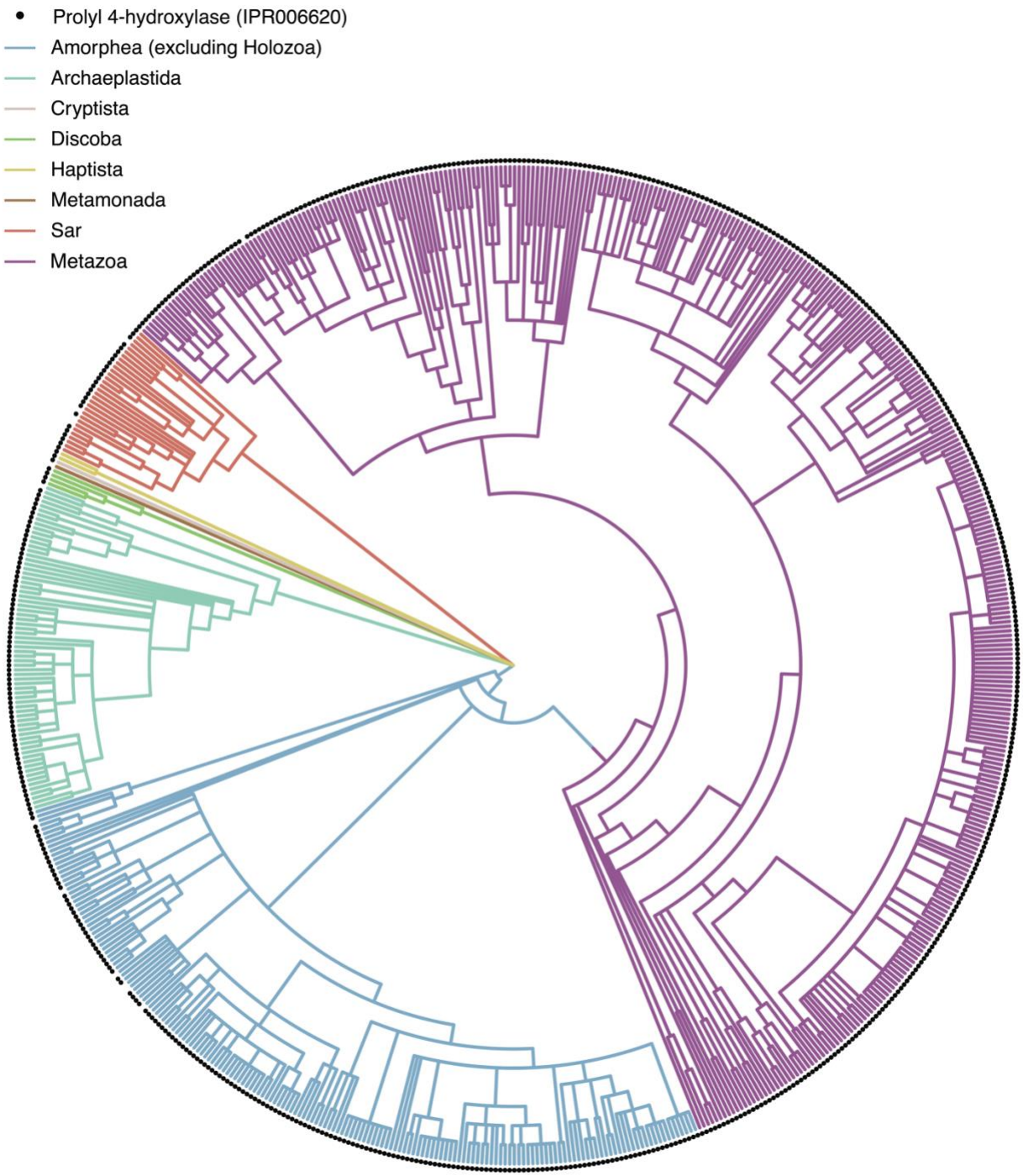


Figure 3S6 Distribution of putative prolyl 4-hydroxylase domains across eukaryotes. We searched for putative prolyl 4-hydroxylase domains (IPR00620) across all eukaryotic proteomes in the UniProtKB database. Each branch of this phylogenetic tree represents a single taxonomic family represented by at least one proteome in UniProtKB that met our quality cutoffs (see Methods). The presence of a black circle indicates that at least one proteome within that family encodes a prolyl 4-hydroxylase domain. The topology of the tree is identical to the trees shown in Fig. 3.2C-F and Fig. 3S5.

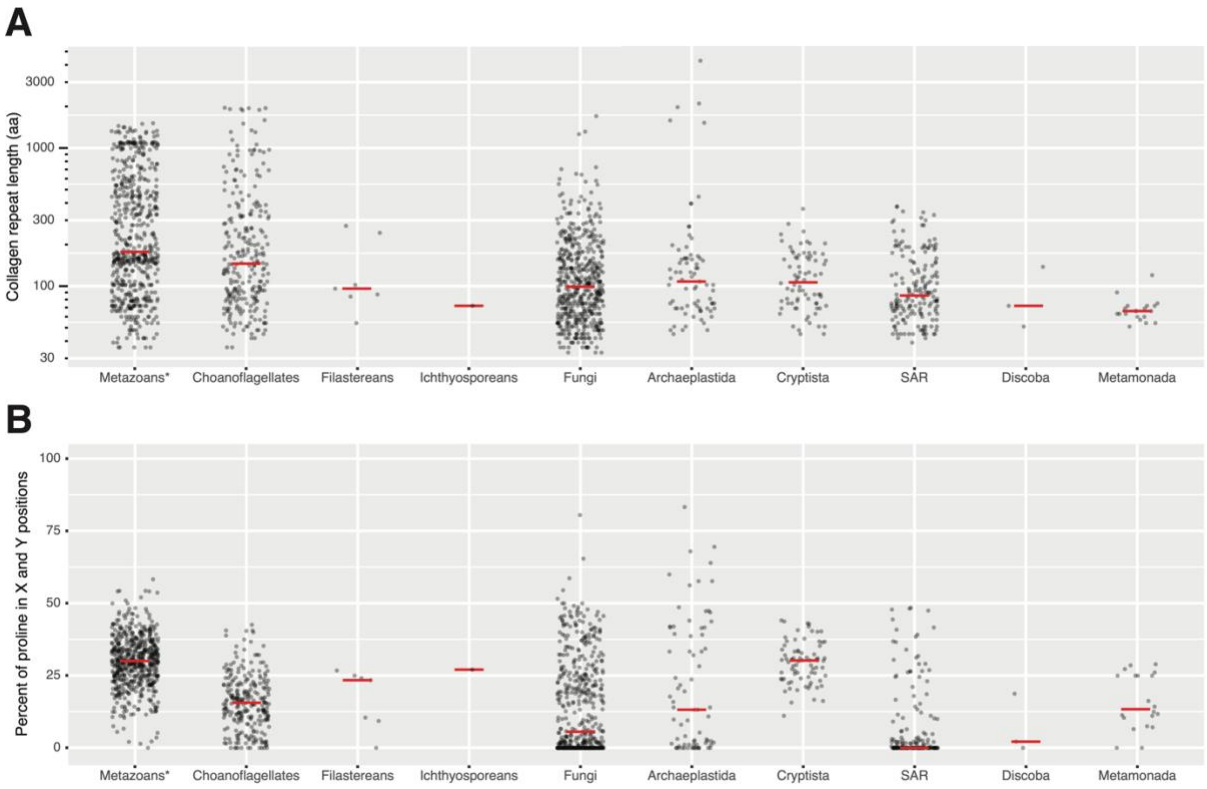


Figure 3S7 Collagen domain length and proline content across eukaryotic diversity. We examined (A) collagen repeat length and (B) proline content (see Methods) across all collagens detected in metazoans (n = 57,104 collagens), choanoflagellates (n = 258 collagens); filastereans (n = 7 collagens); ichthyosporeans (n = 1 collagen); fungi (n = 508 collagens); Archaeplastida (n = 75 collagens); Cryptista (n = 76 collagens); SAR (n = 176 collagens); Discoba (n = 3 collagens); and Metamonada (n = 20 collagens). Note that the y-axis for panel A is on a log-scale. Red bars represent medians. *: to more clearly visualize the distribution of the metazoan collagen data, a randomly selected 1% of metazoan collagens are shown here.

Table 3S1 Holozoan proteome sources.

Species	Basis for proteome prediction‡	Source
Acanthoeca spectabilis	T	(Richter et al., 2018) (https://figshare.com/articles/dataset/Data_from_The_ancestral_animal_genetic_toolkit_revealed_by_diverse_choanoflagellate_Ts/5686984?file=9955288)
Capsaspora owczarzaki	G	Ensembl Protist (https://protists.ensembl.org/Capsaspora_owczarzaki_atcc_30864_gca_000151315)
Choanoeca flexa	T	(Brunet et al., 2019) (https://figshare.com/articles/dataset/Choanoeca_flexa_nonredundant_predicted_proteins_fasta/8216291?file=15312674)
Choanoeca perplexa	T	(Richter et al., 2018) (https://figshare.com/articles/dataset/Data_from_The_ancestral_animal_genetic_toolkit_revealed_by_diverse_choanoflagellate_transcriptomes/5686984?file=9955288)
Codosiga hollandica	T	(Richter et al., 2018) (https://figshare.com/articles/dataset/Data_from_The_ancestral_animal_genetic_toolkit_revealed_by_diverse_choanoflagellate_transcriptomes/5686984?file=9955288)
Creolimax fragrantissima	G	(de Mendoza et al., 2015) (https://figshare.com/articles/dataset/Creolimax_fragrantissima_genome_data/1403592?file=3328016)
Didymoeca costata	T	(Richter et al., 2018) (https://figshare.com/articles/dataset/Data_from_The_ancestral_animal_genetic_toolkit_revealed_by_diverse_choanoflagellate_transcriptomes/5686984?file=9955288)
Hartaetosiga balthica	T	(Richter et al., 2018) (https://figshare.com/articles/dataset/Data_from_The_ancestral_animal_genetic_toolkit_revealed_by_diverse_choanoflagellate_transcriptomes/5686984?file=9955288)
Hartaetosiga gracilis	T	(Richter et al., 2018) (https://figshare.com/articles/dataset/Data_from_The_ancestral_animal_genetic_toolkit_revealed_by_diverse_choanoflagellate_transcriptomes/5686984?file=9955288)

Helgoeca nana	T	(Richter et al., 2018) (https://figshare.com/articles/dataset/Data_from_The_ancestral_animal_genetic_toolkit_revealed_by_diverse_choanoflagellate_transcriptomes/5686984?file=9955288)
Microstomoeca roanoka	T	(Richter et al., 2018) (https://figshare.com/articles/dataset/Data_from_The_ancestral_animal_genetic_toolkit_revealed_by_diverse_choanoflagellate_transcriptomes/5686984?file=9955288)
Ministeria vibrans	T	(Grau-Bové et al., 2017)
Monosiga brevicollis	G	Ensembl Protist (https://protists.ensembl.org/Monosiga_brevicollis_mx1_gca_000002865)
Mylnosiga fluctuans	T	(Richter et al., 2018) (https://figshare.com/articles/dataset/Data_from_The_ancestral_animal_genetic_toolkit_revealed_by_diverse_choanoflagellate_transcriptomes/5686984?file=9955288)
Pigoraptor chileana	T	(Hehenberger et al., 2017) (https://datadryad.org/stash/dataset/doi:10.5061/dryad.26bv4)
Pigoraptor vietnamica	T	(Hehenberger et al., 2017) (https://datadryad.org/stash/dataset/doi:10.5061/dryad.26bv4)
Salpingoeca dolichothecata	T	(Richter et al., 2018) (https://figshare.com/articles/dataset/Data_from_The_ancestral_animal_genetic_toolkit_revealed_by_diverse_choanoflagellate_transcriptomes/5686984?file=9955288)
Salpingoeca helianthica	T	(Richter et al., 2018) (https://figshare.com/articles/dataset/Data_from_The_ancestral_animal_genetic_toolkit_revealed_by_diverse_choanoflagellate_transcriptomes/5686984?file=9955288)
Salpingoeca infusionum	T	(Richter et al., 2018) (https://figshare.com/articles/dataset/Data_from_The_ancestral_animal_genetic_toolkit_revealed_by_diverse_choanoflagellate_transcriptomes/5686984?file=9955288)
Salpingoeca kvevrii	T	(Richter et al., 2018) (https://figshare.com/articles/dataset/Data_from_The_ancestral_animal_genetic_toolkit_revealed_by_diverse_choanoflagellate_transcriptomes/5686984?file=9955288)

Salpingoeca macrocollata	T	(Richter et al., 2018) (https://figshare.com/articles/dataset/Data_from_The_ancestral_animal_genetic_toolkit_revealed_by_diverse_choanoflagellate_transcriptomes/5686984?file=9955288)
Salpingoeca punica	T	(Richter et al., 2018) (https://figshare.com/articles/dataset/Data_from_The_ancestral_animal_genetic_toolkit_revealed_by_diverse_choanoflagellate_transcriptomes/5686984?file=9955288)
Salpingoeca rosetta	G	Ensembl Protist (https://protists.ensembl.org/Salpingoeca_rosetta_gca_000188695)
Salpingoeca urceolata	T	(Richter et al., 2018) (https://figshare.com/articles/dataset/Data_from_The_ancestral_animal_genetic_toolkit_revealed_by_diverse_choanoflagellate_transcriptomes/5686984?file=9955288)
Savillea parva	T	(Richter et al., 2018) (https://figshare.com/articles/dataset/Data_from_The_ancestral_animal_genetic_toolkit_revealed_by_diverse_choanoflagellate_transcriptomes/5686984?file=9955288)
Sphaeroforma arctica	G	Ensembl Protist (https://protists.ensembl.org/Sphaeroforma_arctica_jp610_gca_001186125/)
Stephanoeca diplocostata	T	(Richter et al., 2018) (https://figshare.com/articles/dataset/Data_from_The_ancestral_animal_genetic_toolkit_revealed_by_diverse_choanoflagellate_transcriptomes/5686984?file=9955288) (Note: Australian and French proteomes were concatenated into one proteome for this study)
Syssomonas multiformis	T	(Hehenberger et al., 2017) (https://datadryad.org/stash/dataset/doi:10.5061/dryad.26bv4)

‡T = transcriptome; G = genome

Supplementary Data Files

Supplementary Data Files 1-4 and Supplementary Data Key can be found online at https://www.biorxiv.org/content/10.1101/2021.10.08.463732v1.supplementary-material.

References

Abedin, M., King, N., 2008. The premetazoan ancestry of cadherins. Science 319, 946–948. https://doi.org/10.1126/science.1151084

Adams, E.D.M., Goss, G.G., Leys, S.P., 2010. Freshwater sponges have functional, sealing epithelia with high transepithelial resistance and negative transepithelial potential. PloS One 5, e15040. https://doi.org/10.1371/journal.pone.0015040

Alegado, R.A., Brown, L.W., Cao, S., Dermenjian, R.K., Zuzow, R., Fairclough, S.R., Clardy, J., King, N., 2012. A bacterial sulfonolipid triggers multicellular development in the closest living relatives of animals. eLife 2012. https://doi.org/10.7554/eLife.00013

Anisimova, M., Gascuel, O., 2006. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. Syst. Biol. 55, 539–552. https://doi.org/10.1080/10635150600755453

Aouacheria, A., Geourjon, C., Aghajari, N., Navratil, V., Deléage, G., Lethias, C., Exposito, J.-Y., 2006. Insights into early extracellular matrix evolution: spongin short chain collagen-related proteins are homologous to basement membrane type IV collagens and form a novel family widely distributed in invertebrates. Mol. Biol. Evol. 23, 2288–2302. https://doi.org/10.1093/molbev/msl100

Apprill, A., McNally, S., Parsons, R., Weber, L., 2015. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. Aquat. Microb. Ecol. 75, 129–137.

Arendt, D., 2003. Evolution of eyes and photoreceptor cell types. Int. J. Dev. Biol. 47, 563–571.

Arendt, D., Benito-Gutierrez, E., Brunet, T., Marlow, H., 2015. Gastric pouches and the mucociliary sole: Setting the stage for nervous system evolution. Philos. Trans. R. Soc. B Biol. Sci. 370. https://doi.org/10.1098/rstb.2015.0286

Arendt, D., Musser, J.M., Baker, C.V.H., Bergman, A., Cepko, C., Erwin, D.H., Pavlicev, M., Schlosser, G., Widder, S., Laubichler, M.D., Wagner, G.P., 2016. The origin and evolution of cell types. Nat. Rev. Genet. 17, 744–757. https://doi.org/10.1038/nrg.2016.127

Armon, S., Bull, M.S., Aranda-Diaz, A., Prakash, M., 2018. Ultrafast epithelial contractions provide insights into contraction speed limits and tissue integrity. Proc. Natl. Acad. Sci. 115, E10333–E10341. https://doi.org/10.1073/pnas.1802934115

Aumailley, M., 2013. The laminin family. Cell Adhes. Migr. 7, 48–55. https://doi.org/10.4161/cam.22826

Avelar, G.M., Schumacher, R.I., Zaini, P.A., Leonard, G., Richards, T.A., Gomes, S.L., 2014. A rhodopsin-guanylyl cyclase gene fusion functions in visual perception in a fungus. Curr. Biol. CB 24, 1234–1240. https://doi.org/10.1016/j.cub.2014.04.009

Bachert, B.A., Choi, S.J., Snyder, A.K., Rio, R.V.M., Durney, B.C., Holland, L.A., Amemiya, K., Welkos, S.L., Bozue, J.A., Cote, C.K., Berisio, R., Lukomski, S., 2015. A Unique Set of the Burkholderia Collagen-Like Proteins Provides Insight into Pathogenesis, Genome Evolution and Niche Adaptation, and Infection Detection. PLOS ONE 10, e0137578. https://doi.org/10.1371/journal.pone.0137578

Bailles, A., Collinet, C., Philippe, J.-M., Lenne, P.-F., Munro, E., Lecuit, T., 2019. Transcriptional initiation and mechanically driven propagation of a tissue morphogenetic wave during axis elongation. bioRxiv 430512. https://doi.org/10.1101/430512

Bamford, D.H., Bamford, J.K.H., 1990. Collagenous proteins multiply. Nature 344, 497–497. https://doi.org/10.1038/344497b0
Beljan, S., Herak Bosnar, M., Ćetković, H., 2020. Rho Family of Ras-Like GTPases in Early-Branching Animals. Cells 9, 2279. https://doi.org/10.3390/cells9102279
Berg, R.A., Prockop, D.J., 1973. The thermal transition of a non-hydroxylated form of collagen. Evidence for a role for hydroxyproline in stabilizing the triple-helix of collagen. Biochem. Biophys. Res. Commun. 52, 115–120. https://doi.org/10.1016/0006-291x(73)90961-3
Booth, D.S., Szmidt-Middleton, H., King, N., 2018. Transfection of choanoflagellates illuminates their cell biology and the ancestry of animal septins. Mol. Biol. Cell. https://doi.org/10.1091/mbc.e18-08-0514
Boswell-Smith, V., Spina, D., Page, C.P., 2006. Phosphodiesterase inhibitors. Br. J. Pharmacol. 147 Suppl, S252-7. https://doi.org/10.1038/sj.bjp.0706495
Boute, N., Exposito, J.Y., Boury-Esnault, N., Vacelet, J., Noro, N., Miyazaki, K., Yoshizato, K., Garrone, R., 1997. Type IV collagen in sponges, the missing link in basement membrane ubiquity. Biol. Cell 88, 37–44. https://doi.org/10.1016/S0248-4900(97)86829-3
Brunet, T., 2019. Choanoeca flexa transcriptome and predicted nonredundant proteome. https://doi.org/10.6084/m9.figshare.8216291.v2
Brunet, T., Fischer, A.H.L., Steinmetz, P.R.H., Lauri, A., Bertucci, P., Arendt, D., 2016. The evolutionary origin of bilaterian smooth and striated myocytes. eLife. https://doi.org/10.7554/eLife.19607
Brunet, T., King, N., 2017. The Origin of Animal Multicellularity and Cell Differentiation. Dev. Cell 43, 124–140. https://doi.org/10.1016/j.devcel.2017.09.016
Brunet, T., Larson, B.T., Linden, T.A., Vermeij, M.J.A., McDonald, K., King, N., 2019. Light-regulated collective contractility in a multicellular choanoflagellate. Science 366, 326. https://doi.org/10.1126/science.aay2346
Buehler, M.J., 2006. Nature designs tough collagen: explaining the nanostructure of collagen fibrils. Proc. Natl. Acad. Sci. U. S. A. 103, 12285–12290. https://doi.org/10.1073/pnas.0603216103
Burki, F., Roger, A.J., Brown, M.W., Simpson, A.G.B., 2020. The New Tree of Eukaryotes. Trends Ecol. Evol. 35, 43–55. https://doi.org/10.1016/j.tree.2019.08.008
Canty, E.G., Kadler, K.E., 2005. Procollagen trafficking, processing and fibrillogenesis. J. Cell Sci. 118, 1341–1353. https://doi.org/10.1242/jcs.01731
Carr, M., Richter, D.J., Fozouni, P., Smith, T.J., Jeuck, A., Leadbeater, B.S.C., Nitsche, F., 2017. A six-gene phylogeny provides new insights into choanoflagellate evolution. Mol. Phylogenet. Evol. 107, 166–178. https://doi.org/10.1016/j.ympev.2016.10.011
Casals, C., García-Fojeda, B., Minutti, C.M., 2019. Soluble defense collagens: Sweeping up immune threats. Mol. Immunol. 112, 291–304.
Cavalier-Smith, T., 1997. Amoeboflagellates and mitochondrial cristae in eukaryote evolution: megasystematics of the new protozoan subkingdoms Eozoa and Neozoa. Arch. Für Protistenkd. 147, 237–258.
Celerin, M., Ray, J.M., Schisler, N.J., Day, A.W., Stetler-Stevenson, W.G., Laudenbach, D.E., 1996. Fungal fimbriae are composed of collagen. EMBO J. 15, 4445–4453.
Charalambous, B.M., Keen, J.N., McPherson, M.J., 1988. Collagen-like sequences stabilize homotrimers of a bacterial hydrolase. EMBO J. 7, 2903–2909. https://doi.org/10.1002/j.1460-2075.1988.tb03148.x

Chaves, I., Pokorny, R., Byrdin, M., Hoang, N., Ritz, T., Brettel, K., Essen, L.-O., van der Horst, G.T.J., Batschauer, A., Ahmad, M., 2011. The Cryptochromes: Blue Light Photoreceptors in Plants and Animals. Annu. Rev. Plant Biol. https://doi.org/10.1146/annurev-arplant-042110-103759

Chen, L., DeVries, A.L., Cheng, C.-H.C., 1997. Convergent evolution of antifreeze glycoproteins in Antarctic notothenioid fish and Arctic cod. Proc. Natl. Acad. Sci. 94, 3817–3822. https://doi.org/10.1073/pnas.94.8.3817

Coulon, A., Chow, C.C., Singer, R.H., Larson, D.R., 2013. Eukaryotic transcriptional dynamics: from single molecules to cell populations. Nat. Rev. Genet. 14, 572.

Cunningham, J.A., Liu, A.G., Bengtson, S., Donoghue, P.C.J., 2017. The origin of animals: Can molecular clocks and the fossil record be reconciled? BioEssays News Rev. Mol. Cell. Dev. Biol. 39, 1–12. https://doi.org/10.1002/bies.201600120

Darren Gilmour, Martina Rembold, Maria Leptin, 2017. From morphogen to morphogenesis and back. Nature. https://doi.org/10.1038/nature21348

Davies, H.A., Findlay, K., Daniels, M.J., Dow, J.M., 1997. A novel proline-rich glycoprotein associated with the extracellular matrix of vascular bundles of Brassica petioles. Planta 202, 28–35.

Dayel, M.J., Alegado, R.A., Fairclough, S.R., Levin, T.C., Nichols, S.A., McDonald, K., King, N., 2011. Cell differentiation and morphogenesis in the colony-forming choanoflagellate Salpingoeca rosetta. Dev. Biol. 357, 73–82. https://doi.org/10.1016/j.ydbio.2011.06.003

Dayel, M.J., King, N., 2014. Prey capture and phagocytosis in the choanoflagellate Salpingoeca rosetta. PLoS ONE 9. https://doi.org/10.1371/journal.pone.0095577

Dayraud, C., Alié, A., Jager, M., Chang, P., Le Guyader, H., Manuel, M., Quéinnec, E., 2012. Independent specialisation of myosin II paralogues in muscle vs. non-muscle functions during early animal evolution: A ctenophore perspective, BMC Evolutionary Biology. https://doi.org/10.1186/1471-2148-12-107

de Bruin, E.C., Werten, M.W.T., Laane, C., de Wolf, F.A., 2002. Endogenous prolyl 4-hydroxylation in Hansenula polymorpha and its use for the production of hydroxylated recombinant gelatin. FEMS Yeast Res. 1, 291–298. https://doi.org/10.1111/j.1567-1364.2002.tb00047.x

de Mendoza, A., Suga, H., Permanyer, J., Irimia, M., Ruiz-Trillo, I., 2015. Complex transcriptional regulation and independent evolution of fungal-like traits in a relative of animals. eLife 4, e08904. https://doi.org/10.7554/eLife.08904

Degnan, P.H., Ochman, H., 2011. Illumina-based analysis of microbial community diversity. Isme J. 6, 183.

Dickinson, D.J., Nelson, W.J., Weis, W.I., 2012a. An epithelial tissue in Dictyostelium challenges the traditional origin of metazoan multicellularity. BioEssays News Rev. Mol. Cell. Dev. Biol. 34, 833–840. https://doi.org/10.1002/bies.201100187

Dickinson, D.J., Nelson, W.J., Weis, W.I., 2011. A polarized epithelium organized by beta- and alpha-catenin predates cadherin and metazoan origins. Science 331, 1336–1339. https://doi.org/10.1126/science.1199633

Dickinson, D.J., Robinson, D.N., Nelson, W.J., Weis, W.I., 2012b. α-catenin and IQGAP regulate myosin localization to control epithelial tube morphogenesis in Dictyostelium. Dev. Cell 23, 533–546. https://doi.org/10.1016/j.devcel.2012.06.008

Domínguez-Giménez, P., Brown, N.H., Martín-Bermudo, M.D., 2007. Integrin-ECM interactions regulate the changes in cell shape driving the morphogenesis of the

Drosophila wing epithelium. J. Cell Sci. 120, 1061–1071. https://doi.org/10.1242/jcs.03404
Dudin, O., Ondracka, A., Grau-Bové, X., Haraldsen, A.A.B., Toyoda, A., Suga, H., Bråte, J., Ruiz-Trillo, I., 2019. A unicellular relative of animals generates an epithelium-like cell layer by actomyosin-dependent cellularization. https://doi.org/10.1101/563726
Dziadek, M., 1995. Role of laminin-nidogen complexes in basement membrane formation during embryonic development. Experientia 51, 901–913. https://doi.org/10.1007/BF01921740
Ehrlich, H., Wysokowski, M., \.Zó\ltowska-Aksamitowska, S., Petrenko, I., Jesionowski, T., 2018. Collagens of Poriferan Origin, Marine Drugs. https://doi.org/10.3390/md16030079
Ellis, W.N., 1930. Recent researches on the Choanoflagellata (Craspedomonadines)(fresh-water and marine) with description of new genera and species. Ann Soc Roy Zool Belg 60, 49–88.
Endo, Y., Matsushita, M., Fujita, T., 2015. Chapter Two - New Insights into the Role of Ficolins in the Lectin Pathway of Innate Immunity, in: Jeon, K.W. (Ed.), International Review of Cell and Molecular Biology. Academic Press, pp. 49–110. https://doi.org/10.1016/bs.ircmb.2015.01.003
Engvall, E., Wewer, U.M., 1996. Domains of laminin. J. Cell. Biochem. 61, 493–501. https://doi.org/10.1002/(SICI)1097-4644(19960616)61:4%3C493::AID-JCB2%3E3.0.CO;2-J
Ernst, O.P., Lodowski, D.T., Elstner, M., Hegemann, P., Brown, L.S., Kandori, H., 2014. Microbial and animal rhodopsins: structures, functions, and molecular mechanisms. Chem. Rev. 114, 126–163. https://doi.org/10.1021/cr4003769
Erwin, D.H., 1993. The origin of metazoan development: a palaeobiological perspective. Biol. J. Linn. Soc. 50, 255–274. https://doi.org/10.1111/j.1095-8312.1993.tb00931.x
Exposito, J.-Y., Cluzel, C., Garrone, R., Lethias, C., 2002. Evolution of collagens. Anat. Rec. 268, 302–316. https://doi.org/10.1002/ar.10162
Exposito, J.Y., Valcourt, U., Cluzel, C., Lethias, C., 2010. The fibrillar collagen family, International Journal of Molecular Sciences. https://doi.org/10.3390/ijms11020407
Eyre, D.R., Wu, J.-J., 2005. Collagen Cross-Links, in: Brinckmann, J., Notbohm, H., Müller, P.K. (Eds.), Collagen: Primer in Structure, Processing and Assembly, Topics in Current Chemistry. Springer, Berlin, Heidelberg, pp. 207–229. https://doi.org/10.1007/b103828
Fahey, B., Degnan, B.M., 2010. Origin of animal epithelia: insights from the sponge genome. Evol. Dev. 12, 601–617. https://doi.org/10.1111/j.1525-142X.2010.00445.x
Fairclough, S.R., Chen, Z., Kramer, E., Zeng, Q., Young, S., Robertson, H.M., Begovic, E., Richter, D.J., Russ, C., Westbrook, M.J., Manning, G., Lang, B.F., Haas, B., Nusbaum, C., King, N., 2013. Premetazoan genome evolution and the regulation of cell differentiation in the choanoflagellate Salpingoeca rosetta. Genome Biol. 14, 1–15. https://doi.org/10.1186/gb-2013-14-2-r15
Fidler, A.L., Boudko, S.P., Rokas, A., Hudson, B.G., 2018. The triple helix of collagens–an ancient protein structure that enabled animal multicellularity and tissue evolution. J. Cell Sci. 131.
Fidler, A.L., Darris, C.E., Chetyrkin, S.V., Pedchenko, V.K., Boudko, S.P., Brown, K.L., Gray Jerome, W., Hudson, J.K., Rokas, A., Hudson, B.G., 2017. Collagen iv and basement membrane at the evolutionary dawn of metazoan tissues. eLife 6. https://doi.org/10.7554/eLife.24176.001

Fidler, A.L., Vanacore, R.M., Chetyrkin, S.V., Pedchenko, V.K., Bhave, G., Yin, V.P., Stothers, C.L., Rose, K.L., McDonald, W.H., Clark, T.A., Borza, D.-B., Steele, R.E., Ivy, M.T., Aspirnauts, T., Hudson, J.K., Hudson, B.G., 2014. A unique covalent bond in basement membrane is a primordial innovation for tissue evolution. Proc. Natl. Acad. Sci. 111, 331–336. https://doi.org/10.1073/pnas.1318499111

Frantz, C., Stewart, K.M., Weaver, V.M., 2010. The extracellular matrix at a glance. J. Cell Sci. 123, 4195–4200. https://doi.org/10.1242/jcs.023820

Franzen, W., 1988. Oogenesis and larval development of Scypha ciliata (Porifera, Calcarea). Zoomorphology. https://doi.org/10.1007/BF00312218

Franzke, C.-W., Bruckner, P., Bruckner-Tuderman, L., 2005. Collagenous transmembrane proteins: recent insights into biology and pathology. J. Biol. Chem. 280, 4005–4008. https://doi.org/10.1074/jbc.R400034200

Fraser, D.A., Tenner, A.J., 2008. Directing an appropriate immune response: the role of defense collagens and other soluble pattern recognition molecules. Curr. Drug Targets 9, 113–122. https://doi.org/10.2174/138945008783502476

García-Valdés, E., Gomila, M., Mulet, M., Lalucat, J., 2018. Draft Genome Sequence of Pseudomonas oceani DSM 100277 T , a Deep-Sea Bacterium. Genome Announc. https://doi.org/10.1128/genomea.00254-18

Garrone, R., 1999. Collagen, a common thread in extracellular matrix evolution. Proc. Indian Acad. Sci.-Chem. Sci. 111, 51–56.

Gay, S., Miller, E.J., 1983. What is collagen, what is not. Ultrastruct. Pathol. 4, 365–377.

Ghosh, N., McKillop, T.J., Jowitt, T.A., Howard, M., Davies, H., Holmes, D.F., Roberts, I.S., Bella, J., 2012. Collagen-Like Proteins in Pathogenic E. coli Strains. PLOS ONE 7, e37872. https://doi.org/10.1371/journal.pone.0037872

Gibson, M.C., Perrimon, N., 2003. Apicobasal polarization: epithelial form and function. Curr. Opin. Cell Biol. 15, 747–752. https://doi.org/10.1016/j.ceb.2003.10.008

Gilbert, S.F., 2013. Developmental Biology. Sinauer Associates Incorporated.

Gouy, M., Guindon, S., Gascuel, O., 2010. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol. Biol. Evol. 27, 221–224. https://doi.org/10.1093/molbev/msp259

Gowen, B.B., Borg, T.K., Ghaffar, A., Mayer, E.P., 2001. The collagenous domain of class A scavenger receptors is involved in macrophage adhesion to collagens. J. Leukoc. Biol. 69, 575–582. https://doi.org/10.1189/jlb.69.4.575

Grau-Bové, X., Torruella, G., Donachie, S., Suga, H., Leonard, G., Richards, T.A., Ruiz-Trillo, I., 2017. Dynamics of genomic innovation in the unicellular ancestry of animals. eLife 6, e26036. https://doi.org/10.7554/eLife.26036

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., MacManes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., LeDuc, R.D., Friedman, N., Regev, A., 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8, 1494–1512. https://doi.org/10.1038/nprot.2013.084

Hake, K.H., West, P.T., McDonald, K., Laundon, D., Feng, C., Burkhardt, P., Richter, D., Banfield, J.F., King, N., 2021. Colonial choanoflagellate isolated from Mono Lake harbors a microbiome. https://doi.org/10.1101/2021.03.30.437421

Hallmann, A., 2006. The pherophorins: Common, versatile building blocks in the evolution of extracellular matrix architecture in Volvocales. Plant J. 45, 292–307. https://doi.org/10.1111/j.1365-313X.2005.02627.x
Hehenberger, E., Tikhonenkov, D.V., Kolisko, M., Campo, J. del, Esaulov, A.S., Mylnikov, A.P., Keeling, P.J., 2017. Novel Predators Reshape Holozoan Phylogeny and Reveal the Presence of a Two-Component Signaling System in the Ancestor of Animals. Curr. Biol. 27, 2043-2050.e6. https://doi.org/10.1016/j.cub.2017.06.006
Heisenberg, C.P., Bellaïche, Y., 2013. Forces in tissue morphogenesis and patterning, Cell. https://doi.org/10.1016/j.cell.2013.05.008
Hennet, T., 2019. Collagen glycosylation. Curr. Opin. Struct. Biol. 56, 131–138. https://doi.org/10.1016/j.sbi.2019.01.015
Hieta, R., Myllyharju, J., 2002. Cloning and characterization of a low molecular weight prolyl 4-hydroxylase from Arabidopsis thaliana. Effective hydroxylation of proline-rich, collagen-like, and hypoxia-inducible transcription factor alpha-like peptides. J. Biol. Chem. 277, 23965–23971. https://doi.org/10.1074/jbc.M201865200
Höhn, S., Honerkamp-Smith, A.R., Haas, P.A., Trong, P.K., Goldstein, R.E., 2015. Dynamics of a Volvox embryo turning itself inside out. Phys. Rev. Lett. https://doi.org/10.1103/PhysRevLett.114.178101
Hynes, R.O., 2012. The evolution of metazoan extracellular matrix. J. Cell Biol. 196, 671–679. https://doi.org/10.1083/jcb.201109041
Hynes, R.O., 2002. Integrins: bidirectional, allosteric signaling machines. Cell 110, 673–687. https://doi.org/10.1016/s0092-8674(02)00971-6
Hynes, R.O., Naba, A., 2012. Overview of the matrisome-An inventory of extracellular matrix constituents and functions. Cold Spring Harb. Perspect. Biol. 4. https://doi.org/10.1101/cshperspect.a004903
Ireland, E.V., Woznica, A., King, N., 2020. Synergistic Cues from Diverse Bacteria Enhance Multicellular Development in a Choanoflagellate. Appl. Environ. Microbiol. 86, e02920-19. https://doi.org/10.1128/AEM.02920-19
James-Clark, H., 1867. On the Spongiae Ciliatae as Infusoria Flagellata: or observations on the structure, animality and relationship of Leucosolenia botryoides Bowerbank. Mem. Boston Soc. Nat. Hist. 1, 305–340.
Jékely, G., 2009. Evolution of phototaxis. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 364, 2795–2808. https://doi.org/10.1098/rstb.2009.0072
Juva, K., Prockop, D.J., Cooper, G.W., Lash, J.W., 1966. Hydroxylation of Proline and the Intracellular Accumulation of a Polypeptide Precursor of Collagen. Science 152, 92–94.
Kananavičiūtė, R., Kvederavičiūtė, K., Dabkevičienė, D., Mackevičius, G., Kuisienė, N., 2020. Collagen-like sequences encoded by extremophilic and extremotolerant bacteria. Genomics 112, 2271–2281. https://doi.org/10.1016/j.ygeno.2019.12.023
Kao, W.W., Prockop, D.J., Berg, R.A., 1979. Kinetics for the secretion of nonhelical procollagen by freshly isolated tendon cells. J. Biol. Chem. 254, 2234–2243.
Kent, W.S., 1878. Observations upon Prof. Ernst Haeckel's 'Physemaria' and on the affinity of the sponges. Ann. Mag. Nat. Hist. 5, 1–17.
King, N., Hittinger, C.T., Carroll, S.B., 2003. Evolution of key cell signaling and adhesion protein families predates animal origins. Science 301, 361–363. https://doi.org/10.1126/science.1083853

King, N., Rokas, A., 2017. Embracing Uncertainty in Reconstructing Early Animal Evolution. Curr. Biol. CB 27, R1081–R1088. https://doi.org/10.1016/j.cub.2017.08.054
King, N., Westbrook, M.J., Young, S.L., Kuo, A., Abedin, M., Chapman, J., Fairclough, S., Hellsten, U., Isogai, Y., Letunic, I., Marr, M., Pincus, D., Putnam, N., Rokas, A., Wright, K.J., Zuzow, R., Dirks, W., Good, M., Goodstein, D., Lemons, D., Li, W., Lyons, J.B., Morris, A., Nichols, S., Richter, D.J., Salamov, A., Bork, P., Lim, W.A., Manning, G., Miller, W.T., McGinnis, W., Shapiro, H., Tjian, R., Grigoriev, I.V., Rokhsar, D., 2008. The genome of the choanoflagellate Monosiga brevicollis and the origin of metazoans. Nature 451, 783–788. https://doi.org/10.1038/nature06617
King, N., Young, S.L., Abedin, M., Carr, M., Leadbeater, B.S.C., 2009a. Starting and maintaining Monosiga brevicollis cultures. Cold Spring Harb. Protoc. 2009, pdb.prot5148. https://doi.org/10.1101/pdb.prot5148
King, N., Young, S.L., Abedin, M., Carr, M., Leadbeater, B.S.C., 2009b. Long-term frozen storage of choanoflagellate cultures. Cold Spring Harb. Protoc. 2009, pdb.prot5149. https://doi.org/10.1101/pdb.prot5149
Kirkegaard, J.B., Bouillant, A., Marron, A.O., Leptos, K.C., Goldstein, R.E., 2016. Aerotaxis in the closest relatives of animals. eLife 5. https://doi.org/10.7554/eLife.18109
Knoll, A.H., 2011. The Multiple Origins of Complex Multicellularity. Annu. Rev. Earth Planet. Sci. 39, 217–239. https://doi.org/10.1146/annurev.earth.031208.100209
Knox, S.M., Whitelock, J.M., 2006. Perlecan: how does one molecule do so many things? Cell. Mol. Life Sci. CMLS 63, 2435–2445. https://doi.org/10.1007/s00018-006-6162-z
Kodama, T., Doi, T., Suzuki, H., Takahashi, K., Wada, Y., Gordon, S., 1996. Collagenous macrophage scavenger receptors. Curr. Opin. Lipidol. 7, 287–291. https://doi.org/10.1097/00041433-199610000-00005
Kovács, M., Tóth, J., Hetényi, C., Málnási-Csizmadia, A., Seller, J.R., 2004. Mechanism of blebbistatin inhibition of myosin II. J. Biol. Chem. https://doi.org/10.1074/jbc.M405319200
Koyanagi, M., Takano, K., Tsukamoto, H., Ohtsu, K., Tokunaga, F., Terakita, A., 2008. Jellyfish vision starts with cAMP signaling mediated by opsin-G(s) cascade. Proc. Natl. Acad. Sci. U. S. A. 105, 15576–15580. https://doi.org/10.1073/pnas.0806215105
Kraus, Y., Technau, U., 2006. Gastrulation in the sea anemone Nematostella vectensis occurs by invagination and immigration: An ultrastructural study. Dev. Genes Evol. https://doi.org/10.1007/s00427-005-0038-3
La Scola, B., Desnues, C., Pagnier, I., Robert, C., Barrassi, L., Fournous, G., Merchat, M., Suzan-Monti, M., Forterre, P., Koonin, E., Raoult, D., 2008. The virophage as a unique parasite of the giant mimivirus. Nature 455, 100–104. https://doi.org/10.1038/nature07218
Lamarche, L.B., Kumar, R.P., Trieu, M.M., Devine, E.L., Cohen-Abeles, L.E., Theobald, D.L., Oprian, D.D., 2017. Purification and Characterization of RhoPDE, a Retinylidene/Phosphodiesterase Fusion Protein and Potential Optogenetic Tool from the Choanoflagellate Salpingoeca rosetta. Biochemistry 56, 5812–5822. https://doi.org/10.1021/acs.biochem.7b00519
Lamport, D.T.A., 1977. Structure, Biosynthesis and Significance of Cell Wall Glycoproteins BT - The Structure, Biosynthesis, and Degradation of Wood, in: Loewus, F.A., Runeckles, V.C. (Eds.), . Springer US, Boston, MA, pp. 79–115. https://doi.org/10.1007/978-1-4615-8873-3_3

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. Bioinforma. Oxf. Engl. 23, 2947–2948. https://doi.org/10.1093/bioinformatics/btm404

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

Laundon, D., Larson, B.T., McDonald, K., King, N., Burkhardt, P., 2019. The architecture of cell differentiation in choanoflagellates and sponge choanocytes. PLoS Biol. 17, e3000226. https://doi.org/10.1371/journal.pbio.3000226

Leadbeater, B.S.C., 2015. The Choanoflagellates, The Choanoflagellates. Cambridge University Press. https://doi.org/10.1017/cbo9781139051125

Leadbeater, B.S.C., 2008. Choanoflagellate evolution: the morphological perspective. Protistology 5, 256–267.

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

Leadbeater, B.S.C., 1977. Observations on the life-history and ultrastructure of the marine choanoflagellate choanoeca perplex a ellis. J. Mar. Biol. Assoc. U. K. https://doi.org/10.1017/S0025315400021767

Leclère, L., Nir, T.S., Bazarsky, M., Braitbard, M., Schneidman-Duhovny, D., Gat, U., 2020. Dynamic Evolution of the Cthrc1 Genes, a Newly Defined Collagen-Like Family. Genome Biol. Evol. 12, 3957–3970. https://doi.org/10.1093/gbe/evaa020

Lee, J.L., Streuli, C.H., 2014. Integrins and epithelial cell polarity. J. Cell Sci. jcs.146142. https://doi.org/10.1242/jcs.146142

Lee, J.Y., Harland, R.M., 2007. Actomyosin contractility and microtubules drive apical constriction in Xenopus bottle cells. Dev. Biol. https://doi.org/10.1016/j.ydbio.2007.08.010

Leptin, M., 2005. Gastrulation Movements: the Logic and the Nuts and Bolts. Dev. Cell 8, 305–320. https://doi.org/10.1016/j.devcel.2005.02.007

Leptin, M., 1999. Gastrulation in Drosophila: the logic and the cellular mechanisms. EMBO J. 18, 3187–3192. https://doi.org/10.1093/emboj/18.12.3187

Levin, T.C., Greaney, A.J., Wetzel, L., King, N., 2014. The Rosetteless gene controls development in the choanoflagellate S. rosetta. eLife 3. https://doi.org/10.7554/eLife.04070

Leys, S.P., Eerkes-Medrano, D., 2005a. Gastrulation in Calcareous Sponges: In Search of Haeckel's Gastraea. Integr. Comp. Biol. 45, 342–351. https://doi.org/10.1093/icb/45.2.342

Leys, S.P., Eerkes-Medrano, D., 2005b. Gastrulation in calcareous sponges: In search of Haeckel's Gastraea, in: Integrative and Comparative Biology. https://doi.org/10.1093/icb/45.2.342

Leys, S.P., Nichols, S.A., Adams, E.D.M., 2009. Epithelia and integration in sponges. Integr. Comp. Biol. 49, 167–177. https://doi.org/10.1093/icb/icp038

Leys, S.P., Riesgo, A., 2012. Epithelia, an evolutionary novelty of metazoans. J. Exp. Zoolog. B Mol. Dev. Evol. 318, 438–447. https://doi.org/10.1002/jez.b.21442

Li, R., Neundorf, I., Nitsche, F., 2018. First Efficient Transfection in Choanoflagellates using Cell-Penetrating Peptides. https://doi.org/10.1101/260190
Li, W., Godzik, A., 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinforma. Oxf. Engl. 22, 1658–1659. https://doi.org/10.1093/bioinformatics/btl158
Loria, P.M., Hodgkin, J., Hobert, O., 2004. A Conserved Postsynaptic Transmembrane Protein Affecting Neuromuscular Signaling in Caenorhabditis elegans. J. Neurosci. 24, 2191–2201. https://doi.org/10.1523/JNEUROSCI.5462-03.2004
Lukomski, S., Bachert, B.A., Squeglia, F., Berisio, R., 2017. Collagen-like proteins of pathogenic streptococci. Mol. Microbiol. 103, 919–930. https://doi.org/10.1111/mmi.13604
Luther, K.B., Hülsmeier, A.J., Schegg, B., Deuber, S.A., Raoult, D., Hennet, T., 2011. Mimivirus collagen is modified by bifunctional lysyl hydroxylase and glycosyltransferase enzyme. J. Biol. Chem. 286, 43701–43709. https://doi.org/10.1074/jbc.M111.309096
Mackie, G.O., 1970. Neuroid Conduction and the Evolution of Conducting Tissues. Q. Rev. Biol. https://doi.org/10.1086/406645
Maertens, B., Hopkins, D., Franzke, C.-W., Keene, D.R., Bruckner-Tuderman, L., Greenspan, D.S., Koch, M., 2007. Cleavage and Oligomerization of Gliomedin, a Transmembrane Collagen Required for Node of Ranvier Formation*. J. Biol. Chem. 282, 10647–10659. https://doi.org/10.1074/jbc.M611339200
Magie, C.R., Daly, M., Martindale, M.Q., 2007. Gastrulation in the cnidarian Nematostella vectensis occurs via invagination not ingression. Dev. Biol. https://doi.org/10.1016/j.ydbio.2007.02.044
Marieb, E., Hoehn, K., 2015. Pregnancy and human development. Hum. Anat. Physiol. 10th Ed Pearson Educ. Ltd. 1116–7.
Martin, A.C., Goldstein, B., 2014. Apical constriction: themes and variations on a cellular mechanism driving morphogenesis. Development. https://doi.org/10.1242/dev.102228
Mateos, C.S.-C., Valencia-Expósito, A., Palacios, I.M., Martín-Bermudo, M.D., 2020. Integrins regulate epithelial cell shape by controlling the architecture and mechanical properties of basal actomyosin networks. PLOS Genet. 16, e1008717. https://doi.org/10.1371/journal.pgen.1008717
Matsushita, M., 2010. Ficolins: Complement-Activating Lectins Involved in Innate Immunity. J. Innate Immun. 2, 24–32. https://doi.org/10.1159/000228160
McDonald, K.L., 2014. Rapid embedding methods into epoxy and LR White resins for morphological and immunological analysis of cryofixed biological specimens. Microsc. Microanal. Off. J. Microsc. Soc. Am. Microbeam Anal. Soc. Microsc. Soc. Can. 20, 152–163. https://doi.org/10.1017/S1431927613013846
Medveczky, M.M., Geck, P., Vassallo, R., Medveczky, P.G., 1993. Expression of the collagen-like putative oncoprotein ofHerpesvirus saimiri in transformed T cells. Virus Genes 7, 349–365. https://doi.org/10.1007/BF01703391
Merle, T., Farge, E., 2018. Trans-scale mechanotransductive cascade of biochemical and biomechanical patterning in embryonic development: the light side of the force. Curr. Opin. Cell Biol. 55, 111–118. https://doi.org/10.1016/j.ceb.2018.07.003
Mikkola, M.L., Thesleff, I., 2003. Ectodysplasin signaling in development. Cytokine Growth Factor Rev. 14, 211–224. https://doi.org/10.1016/S1359-6101(03)00020-0

Miño, G.L., Koehl, M.A.R., King, N., Stocker, R., 2017. Finding patches in a heterogeneous aquatic environment: pH-taxis by the dispersal stage of choanoflagellates. Limnol. Oceanogr. Lett. 2, 37–46. https://doi.org/10.1002/lol2.10035

Mohs, A., Silva, T., Yoshida, T., Amin, R., Lukomski, S., Inouye, M., Brodsky, B., 2007. Mechanism of Stabilization of a Bacterial Collagen Triple Helix in the Absence of Hydroxyproline *. J. Biol. Chem. 282, 29757–29765. https://doi.org/10.1074/jbc.M703991200

Nakajima, Y., Burke, R.D., 1996. The initial phase of gastrulation in sea urchins is accompanied by the formation of bottle cells. Dev. Biol. https://doi.org/10.1006/dbio.1996.0273

Neubauer, E.F., Poole, A.Z., Weis, V.M., Davy, S.K., 2016. The scavenger receptor repertoire in six cnidarian species and its putative role in cnidarian-dinoflagellate symbiosis. PeerJ 4, e2692. https://doi.org/10.7717/peerj.2692

Nichols, S.A., Roberts, B.W., Richter, D.J., Fairclough, S.R., King, N., 2012. Origin of metazoan cadherin diversity and the antiquity of the classical cadherin/β-catenin complex. Proc. Natl. Acad. Sci. U. S. A. 109, 13046–13051. https://doi.org/10.1073/pnas.1120685109

Nickel, M., Scheer, C., Hammel, J.U., Herzen, J., Beckmann, F., 2011. The contractile sponge epithelium sensu lato--body contraction of the demosponge Tethya wilhelma is mediated by the pinacoderm. J. Exp. Biol. 214, 1692–1698. https://doi.org/10.1242/jeb.049148

Nielsen, C., 2012. Animal Evolution. Interrelationships of the living phyla, 3rd ed. Oxford University Press.

Olsen, D.R., Leigh, S.D., Chang, R., McMullin, H., Ong, W., Tai, E., Chisholm, G., Birk, D.E., Berg, R.A., Hitzeman, R.A., Toman, P.D., 2001. Production of Human Type I Collagen in Yeast Reveals Unexpected New Insights into the Molecular Assembly of Collagen Trimers *. J. Biol. Chem. 276, 24038–24043. https://doi.org/10.1074/jbc.M101613200

Pang, K., Martindale, M.Q., 2008. Developmental expression of homeobox genes in the ctenophore Mnemiopsis leidyi. Dev. Genes Evol. https://doi.org/10.1007/s00427-008-0222-3

Parada, A.E., Needham, D.M., Fuhrman, J.A., 2016. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. Environ. Microbiol. 18, 1403–1414. https://doi.org/10.1111/1462-2920.13023

Parfrey, L.W., Lahr, D.J.G., 2013. Multicellularity arose several times in the evolution of eukaryotes (Response to DOI 10.1002/bies.201100187). BioEssays 35, 339–347. https://doi.org/10.1002/bies.201200143

Paulsson, M., 1992. Basement membrane proteins: structure, assembly, and cellular interactions. Crit. Rev. Biochem. Mol. Biol. 27, 93–127.

Pérez-Ortín, J.E., Alepuz, P.M., Moreno, J., 2007. Genomics and gene transcription kinetics in yeast. TRENDS Genet. 23, 250–257.

Perfect, S.E., O'Connell, R.J., Green, E.F., Doering-Saad, C., Green, J.R., 1998. Expression cloning of a fungal proline-rich glycoprotein specific to the biotrophic interface formed in the Colletotrichum–bean interaction. Plant J. 15, 273–279. https://doi.org/10.1046/j.1365-313X.1998.00196.x

Perret, S., Merle, C., Bernocco, S., Berland, P., Garrone, R., Hulmes, D.J.S., Theisen, M., Ruggiero, F., 2001. Unhydroxylated Triple Helical Collagen I Produced in Transgenic Plants Provides New Clues on the Role of Hydroxyproline in Collagen Folding and Fibril Formation *. J. Biol. Chem. 276, 43693–43698. https://doi.org/10.1074/jbc.M105507200

Plachetzki, D.C., Fong, C.R., Oakley, T.H., 2010. The evolution of phototransduction from an ancestral cyclic nucleotide gated pathway. Proc. R. Soc. B Biol. Sci. https://doi.org/10.1098/rspb.2009.1797
Pollard, T.D., Goldman, R.D., 2018. Overview of the Cytoskeleton from an Evolutionary Perspective. Cold Spring Harb. Perspect. Biol. 10, a030288. https://doi.org/10.1101/cshperspect.a030288
Pyagay, P., Heroult, M., Wang, Q., Lehnert, W., Belden, J., Liaw, L., Friesel, R.E., Lindner, V., 2005. Collagen triple helix repeat containing 1, a novel secreted protein in injured and diseased arteries, inhibits collagen expression and promotes cell migration. Circ. Res. 96, 261–268. https://doi.org/10.1161/01.RES.0000154262.07264.12
Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., La Scola, B., Suzan, M., Claverie, J.-M., 2004. The 1.2-Megabase Genome Sequence of Mimivirus. Science 306, 1344–1350. https://doi.org/10.1126/science.1101485
Rasmussen, M., Jacobsson, M., Björck, L., 2003. Genome-based Identification and Analysis of Collagen-related Structural Motifs in Bacterial and Viral Proteins *. J. Biol. Chem. 278, 32313–32316. https://doi.org/10.1074/jbc.M304709200
Reinhard, C.T., Planavsky, N.J., Olson, S.L., Lyons, T.W., Erwin, D.H., 2016. Earth's oxygen cycle and the evolution of animal life. Proc. Natl. Acad. Sci. U. S. A. 113, 8933–8938. https://doi.org/10.1073/pnas.1521544113
Ricard-Blum, S., 2011. The Collagen Family. Cold Spring Harb. Perspect. Biol. 3, 1–19. https://doi.org/10.1101/cshperspect.a004978
Richards, T.A., Cavalier-Smith, T., 2005. Myosin domain evolution and the primary divergence of eukaryotes. Nature 436, 1113–1118. https://doi.org/10.1038/nature03949
Richter, D.J., Fozouni, P., Eisen, M.B., King, N., 2018. Gene family innovation, conservation and loss on the animal stem lineage. eLife 7, e34226. https://doi.org/10.7554/eLife.34226
Richter, D.J., King, N., 2013. The genomic and cellular foundations of animal origins. Annu. Rev. Genet. 47.
Rodriguez-Pascual, F., Slatter, D.A., 2016. Collagen cross-linking: insights on the evolution of metazoan extracellular matrix. Sci. Rep. 6, 37374. https://doi.org/10.1038/srep37374
RStudio Team, 2020. RStudio: Integrated Development Environment for R. RStudio, PBC., Boston, MA.
Ruggiero, F., Exposito, J.-Y., Bournat, P., Gruber, V., Perret, S., Comte, J., Olagnier, B., Garrone, R., Theisen, M., 2000. Triple helix assembly and processing of human collagen produced in transgenic tobacco plants. FEBS Lett. 469, 132–136. https://doi.org/10.1016/S0014-5793(00)01259-X
Ruiz-Trillo, I., Roger, A.J., Burger, G., Gray, M.W., Lang, B.F., 2008. A phylogenomic investigation into the origin of metazoa. Mol. Biol. Evol. 25, 664–672. https://doi.org/10.1093/molbev/msn006
Ryerson, W.G., Schwenk, K., 2012. A simple, inexpensive system for digital particle image velocimetry (DPIV) in biomechanics. J. Exp. Zool. Part Ecol. Genet. Physiol. 317, 127–140. https://doi.org/10.1002/jez.725
Saitoh, M., Ishikawa, T., Matsushima, S., Naka, M., Hidaka, H., 1987. Selective inhibition of catalytic activity of smooth muscle myosin light chain kinase. J. Biol. Chem.
Sawyer, J.M., Harrell, J.R., Shemer, G., Sullivan-Brown, J., Roh-Johnson, M., Goldstein, B., 2010. Apical constriction: A cell shape change that can drive morphogenesis, Developmental Biology. https://doi.org/10.1016/j.ydbio.2009.09.009

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D.J., Hartenstein, V., Eliceiri, K., Tomancak, P., Cardona, A., 2012. Fiji: an open-source platform for biological-image analysis. Nat. Methods 9, 676–682. https://doi.org/10.1038/nmeth.2019

Sebé-Pedrós, A., Degnan, B.M., Ruiz-Trillo, I., 2017. The origin of Metazoa: A unicellular perspective, Nature Reviews Genetics. https://doi.org/10.1038/nrg.2017.21

Sebé-Pedrós, A., Grau-Bové, X., Richards, T.A., Ruiz-Trillo, I., 2014. Evolution and classification of myosins, a paneukaryotic whole-genome approach. Genome Biol. Evol. 6, 290–305. https://doi.org/10.1093/gbe/evu013

Sebé-Pedrós, A., Zheng, Y., Ruiz-Trillo, I., Pan, D., 2012. Premetazoan origin of the hippo signaling pathway. Cell Rep. 1, 13–20. https://doi.org/10.1016/j.celrep.2011.11.004

Sheppard, D., 2003. Functions of Pulmonary Epithelial Integrins: From Development to Disease. Physiol. Rev. 83, 673–686. https://doi.org/10.1152/physrev.00033.2002

Sheppard, D., 1996. Epithelial integrins. BioEssays News Rev. Mol. Cell. Dev. Biol. 18, 655–660. https://doi.org/10.1002/bies.950180809

Shoulders, M.D., Raines, R.T., 2009. Collagen Structure and Stability. Annu. Rev. Biochem. 78, 929–958. https://doi.org/10.1146/annurev.biochem.77.032207.120833

Showalter, A.M., 1993. Structure and function of plant cell wall proteins. Plant Cell 5, 9.

Sigg, M.A., Menchen, T., Lee, C., Johnson, J., Jungnickel, M.K., Choksi, S.P., Garcia, G., Busengdal, H., Dougherty, G., Pennekamp, P., Werner, C., Rentzsch, F., Florman, H.M., Krogan, N., Wallingford, J.B., Omran, H., Reiter, J.F., 2017. Evolutionary proteomics uncovers ancient associations of cilia with signaling pathways. Dev. Cell 43, 744-762.e11. https://doi.org/10.1016/j.devcel.2017.11.014

Spudich, J.L., 2006. The multitalented microbial sensory rhodopsins, Trends in Microbiology. https://doi.org/10.1016/j.tim.2006.09.005

Srivastava, M., Simakov, O., Chapman, J., Fahey, B., Gauthier, M.E.A., Mitros, T., Richards, G.S., Conaco, C., Dacre, M., Hellsten, U., Larroux, C., Putnam, N.H., Stanke, M., Adamska, M., Darling, A., Degnan, S.M., Oakley, T.H., Plachetzki, D.C., Zhai, Y., Adamski, M., Calcino, A., Cummins, S.F., Goodstein, D.M., Harris, C., Jackson, D.J., Leys, S.P., Shu, S., Woodcroft, B.J., Vervoort, M., Kosik, K.S., Manning, G., Degnan, B.M., Rokhsar, D.S., 2010. The Amphimedon queenslandica genome and the evolution of animal complexity. Nature 466, 720–726. https://doi.org/10.1038/nature09201

Steinmetz, P.R.H., Kraus, J.E.M., Larroux, C., Hammel, J.U., Amon-Hassenzahl, A., Houliston, E., Wörheide, G., Nickel, M., Degnan, B.M., Technau, U., 2012. Independent evolution of striated muscles in cnidarians and bilaterians. Nature. https://doi.org/10.1038/nature11180

Stern, D.L., 2013. The genetic causes of convergent evolution. Nat. Rev. Genet. 14, 751–764. https://doi.org/10.1038/nrg3483

Streuli, C.H., 2016. Integrins as architects of cell behavior. Mol. Biol. Cell 27, 2885–2888. https://doi.org/10.1091/mbc.E15-06-0369

Streuli, C.H., Bailey, N., Bissell, M.J., 1991. Control of mammary epithelial differentiation: basement membrane induces tissue-specific gene expression in the absence of cell-cell interaction and morphological polarity. J. Cell Biol. 115, 1383–1395. https://doi.org/10.1083/jcb.115.5.1383

Sumper, M., Hallmann, A., 1998. Biochemistry of the Extracellular Matrix of Volvox. Int. Rev. Cytol. 180, 51–85. https://doi.org/10.1016/S0074-7696(08)61770-2

Sutherland, T.D., Peng, Y.Y., Trueman, H.E., Weisman, S., Okada, S., Walker, A.A., Sriskantha, A., White, J.F., Huson, M.G., Werkmeister, J.A., Glattauer, V., Stoichevska, V., Mudie, S.T., Haritos, V.S., Ramshaw, J.A.M., 2013. A new class of animal collagen masquerading as an insect silk. Sci. Rep. 3, 2864. https://doi.org/10.1038/srep02864
Tian, Y., Gao, S., Yang, S., Nagel, G., 2018. A novel rhodopsin phosphodiesterase from Salpingoeca rosetta shows light-enhanced substrate affinity. Biochem. J. 475, 1121–1128. https://doi.org/10.1042/BCJ20180010
Tidona, C.A., Darai, G., 1997. The complete DNA sequence of lymphocystis disease virus. Virology 230, 207–216. https://doi.org/10.1006/viro.1997.8456
Towe, K.M., 1970. Oxygen-Collagen Priority and the Early Metazoan Fossil Record. Proc. Natl. Acad. Sci. 65, 781–788. https://doi.org/10.1073/pnas.65.4.781
Tuntevski, K., Durney, B.C., Snyder, A.K., Lasala, P.R., Nayak, A.P., Green, B.J., Beezhold, D.H., Rio, R.V.M., Holland, L.A., Lukomski, S., 2013. Aspergillus collagen-like genes (acl): identification, sequence polymorphism, and assessment for PCR-based pathogen detection. Appl. Environ. Microbiol. 79, 7882–7895. https://doi.org/10.1128/AEM.02835-13
Tyler, S., 2003. Epithelium--the primary building block for metazoan complexity. Integr. Comp. Biol. 43, 55–63. https://doi.org/10.1093/icb/43.1.55
UniProt Consortium, 2021. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res. 49, D480–D489. https://doi.org/10.1093/nar/gkaa1100
van Hulten, M.C.W., Witteveldt, J., Peters, S., Kloosterboer, N., Tarchini, R., Fiers, M., Sandbrink, H., Lankhorst, R.K., Vlak, J.M., 2001. The White Spot Syndrome Virus DNA Genome Sequence. Virology 286, 7–22. https://doi.org/10.1006/viro.2001.1002
Varoqueaux, F., Williams, E.A., Grandemange, S., Truscello, L., Kamm, K., Schierwater, B., Jékely, G., Fasshauer, D., 2018. High Cell Diversity and Complex Peptidergic Signaling Underlie Placozoan Behavior. Curr. Biol. https://doi.org/10.1016/j.cub.2018.08.067
Vasquez, C.G., de la Serna, E.L., Dunn, A.R., 2021. How cells tell up from down and stick together to construct multicellular tissues - interplay between apicobasal polarity and cell-cell adhesion. J. Cell Sci. 134, jcs248757. https://doi.org/10.1242/jcs.248757
Velle, K.B., Fritz-Laylin, L.K., 2019. Diversity and evolution of actin-dependent phenotypes. Curr. Opin. Genet. Dev., Evolutionary genetics 58–59, 40–48. https://doi.org/10.1016/j.gde.2019.07.016
Vijayraghavan, D.S., Davidson, L.A., 2017. Mechanics of neurulation: From classical to current perspectives on the physical mechanics that shape, fold, and form the neural tube. Birth Defects Res. 109, 153–168. https://doi.org/10.1002/bdra.23557
Wakatsuki, T., Schwab, B., Thompson, N.C., Elson, E.L., 2001. Effects of cytochalasin D and latrunculin B on mechanical properties of cells. J. Cell Sci. https://doi.org/10.1016/s0955-0674(98)80085-9
Walmsley, A.R., Batten, M.R., Lad, U., Bulleid, N.J., 1999. Intracellular retention of procollagen within the endoplasmic reticulum is mediated by prolyl 4-hydroxylase. J. Biol. Chem. 274, 14884–14892. https://doi.org/10.1074/jbc.274.21.14884
Wang, C., Leger, R.J.S., 2006. A collagenous protective coat enables Metarhizium anisopliae to evade insect immune responses. Proc. Natl. Acad. Sci. 103, 6647–6652. https://doi.org/10.1073/pnas.0601951103
Watari, M., Ikuta, T., Yamada, D., Shihoya, W., Yoshida, K., Tsunoda, S.P., Nureki, O., Kandori, H., 2019. Spectroscopic study of the transmembrane domain of a rhodopsin-

phosphodiesterase fusion protein from a unicellular eukaryote. J. Biol. Chem. https://doi.org/10.1074/jbc.RA118.006277
Weishaar, R.E., Cain, M.H., Bristol, J.A., 1985. A new generation of phosphodiesterase inhibitors: multiple molecular forms of phosphodiesterase and the potential for drug selectivity. J. Med. Chem. 28, 537–545.
Wenger, M.P.E., Bozec, L., Horton, M.A., Mesquida, P., 2007. Mechanical properties of collagen fibrils. Biophys. J. 93, 1255–1263. https://doi.org/10.1529/biophysj.106.103192
Wetzel, L.A., Hulett, R.E., Sigg, M.A., Aldayafleh, R., Booth, D.S., Chan, D., King, N., Levin, T.C., King, G.A., 2018. Predicted glycosyltransferases promote development and prevent spurious cell clumping in the choanoflagellate S. rosetta. eLife. https://doi.org/10.7554/elife.41482
Whitelock, J.M., Melrose, J., Iozzo, R.V., 2008. Diverse cell signaling events modulated by perlecan. Biochemistry 47, 11174–11183. https://doi.org/10.1021/bi8013938
Woznica, A., Cantley, A.M., Beemelmanns, C., Freinkman, E., Clardy, J., King, N., 2016. Bacterial lipids activate, synergize, and inhibit a developmental switch in choanoflagellates. Proc. Natl. Acad. Sci. U. S. A. 113, 7894–7899. https://doi.org/10.1073/pnas.1605015113
Woznica, A., Gerdt, J.P., Hulett, R.E., Clardy, J., King, N., 2017. Mating in the Closest Living Relatives of Animals Is Induced by a Bacterial Chondroitinase. Cell 170, 1175-1183.e11. https://doi.org/10.1016/j.cell.2017.08.005
Yamada, Y., Doi, T., Hamakubo, T., Kodama, T., 1998. Scavenger receptor family proteins: roles for atherosclerosis, host defence and disorders of the central nervous system. Cell. Mol. Life Sci. CMLS 54, 628–640. https://doi.org/10.1007/s000180050191
Yau, S., Lauro, F.M., DeMaere, M.Z., Brown, M.V., Thomas, T., Raftery, M.J., Andrews-Pfannkoch, C., Lewis, M., Hoffman, J.M., Gibson, J.A., Cavicchioli, R., 2011. Virophage control of antarctic algal host–virus dynamics. Proc. Natl. Acad. Sci. U. S. A. 108, 6163–6168. https://doi.org/10.1073/pnas.1018221108
Yeh, Y.-C., Lin, H.-H., Tang, M.-J., 2012. A tale of two collagen receptors, integrin β1 and discoidin domain receptor 1, in epithelial cell differentiation. Am. J. Physiol. Cell Physiol. 303, C1207-1217. https://doi.org/10.1152/ajpcell.00253.2012
Yoshida, K., Tsunoda, S.P., Brown, L.S., Kandori, H., 2017. A unique choanoflagellate enzyme rhodopsin exhibits light-dependent cyclic nucleotide phosphodiesterase activity. J. Biol. Chem. 292, 7531–7541. https://doi.org/10.1074/jbc.M117.775569
Yu, Z., An, B., Ramshaw, J.A.M., Brodsky, B., 2014. Bacterial collagen-like proteins that form triple-helical structures. J. Struct. Biol. 186, 451–461. https://doi.org/10.1016/j.jsb.2014.01.003
Yurchenco, P.D., 2011. Basement membranes: cell scaffoldings and signaling platforms. Cold Spring Harb. Perspect. Biol. 3, a004911. https://doi.org/10.1101/cshperspect.a004911
Zhang, Q.-Y., Xiao, F., Xie, J., Li, Z.-Q., Gui, J.-F., 2004. Complete Genome Sequence of Lymphocystis Disease Virus Isolated from China. J. Virol. 78, 6982–6994. https://doi.org/10.1128/JVI.78.13.6982-6994.2004
Zhang, X., Cote, R.H., 2005. cGMP signaling in vertebrate retinal photoreceptor cells. Front Biosci 10, 1191–1204.