

The distinctive flagellar proteome of *Euglena gracilis* illuminates the complexities of protistan flagella adaptation

Michael Hammond^{1*} , Martin Zoltner^{2,3*} , Jack Garrigan², Erin Butterfield², Vladimír Varga⁴ , Julius Lukes^{1,5}  and Mark C. Field^{1,2} 

¹Institute of Parasitology, Biology Centre, Czech Academy of Sciences, České Budějovice (Budweis) 370 05, Czech Republic; ²School of Life Sciences, University of Dundee, Dundee, DD1 5EH, UK; ³Department of Parasitology, Faculty of Science, BIOCEV, Charles University, Vestec 252 50, Czech Republic; ⁴Institute of Molecular Genetics of the Czech Academy of Sciences, Prague 142 20, Czech Republic; ⁵Faculty of Sciences, University of South Bohemia, České Budějovice (Budweis) 370 05, Czech Republic

Summary

Author for correspondence:
Mark C. Field
Email: mfield@mac.com

Received: 25 May 2021
Accepted: 15 July 2021

New Phytologist (2021) 232: 1323–1336
doi: 10.1111/nph.17638

Key words: cilia, *Euglena*, evolution, flagella, pellicle, proteomics.

- The eukaryotic flagellum/cilium is a prominent organelle with conserved structure and diverse functions. *Euglena gracilis*, a photosynthetic and highly adaptable protist, employs its flagella for both locomotion and environmental sensing.
- Using proteomics of isolated *E. gracilis* flagella we identify nearly 1700 protein groups, which challenges previous estimates of the protein complexity of motile eukaryotic flagella.
- We not only identified several unexpected similarities shared with mammalian flagella, including an entire glycolytic pathway and proteasome, but also document a vast array of flagella-based signal transduction components that coordinate gravitaxis and phototactic motility. By contrast, the pellicle was found to consist of >900 protein groups, containing additional structural and signalling components.
- Our data identify significant adaptations within the *E. gracilis* flagellum, many of which are clearly linked to the highly flexible lifestyle.

Introduction

The flagellum is an important structure present across the tree of life, with roles including motility, signalling and development (Diniz *et al.*, 2012). The eukaryotic flagellum or cilium shares a universal axoneme structure, characterized by nine outer microtubule doublets which evolved independently from the prokaryotic flagellum. Motile flagella, capable of beating, possess two central microtubule singlets, tethered by radial spokes to the outer doublets which are absent from nonmotile forms, including primary cilia of metazoan cells. The outer doublets slide relative to each other with the assistance of the motor protein dynein that enables flagella motion (Hausmann *et al.*, 2014). Motile flagella are present in the majority of single-celled eukaryotes or protists for at least part of their life cycles, playing critical roles in movement and environmental sensing (Leander *et al.*, 2017). In mammals, the flagellum propels spermatozoa along the reproductive canal, whereas fallopian tube cilia beat to translocate the ovum, and later fertilized zygote, towards the uterus. Motile cilia have roles along the respiratory epithelium and ependyma. Nonmotile cilia are additionally present in various cells of multicellular organisms, where they coordinate critical signalling pathways. Cilia defects are associated with a number of human disorders referred to as ciliopathies, which include polycystic kidney disease, infertility, *situs inversus* and blindness; many of the genes

responsible are part of the intraflagellar transport (IFT) system, an ancient component of the protozoan system (Waters & Beales, 2011; Rout & Field, 2017).

Euglena gracilis is a phototrophic, ecologically important member of the phylum Euglenozoa. The lineage includes the well-known parasites within the order Kinetoplastida and free-living diplomonids (Kostygov *et al.*, 2021). As with many euglenids, *E. gracilis* has two motile flagella of substantially different lengths (Leander *et al.*, 2017); one is referred to as ‘emergent’ and the other ‘nonemergent’, depending on their retention within or extrusion from the flagellar pocket, a plasma membrane invagination (Fig. 1). The emergent flagellum is responsible for coordinating cell movement and additionally contains the paraflagellar body (PFB), an invaginated structure near the flagellar base (Verni *et al.*, 1992). The PFB is photosensitive and the emergent flagellum regulates phototaxis to optimize photosynthetic conditions (Iseki *et al.*, 2002; Yang *et al.*, 2021). The flagellum also gravitaxis for positioning within the water column (Nasir *et al.*, 2018).

Adjacent to the *E. gracilis* emergent flagellum axoneme is a lattice-like structure, the paraflagellar rod (PFR) (Hyams, 1982). PFR morphology and composition have roles in signalling, metabolism, environmental sensing and regulatory functions in the kinetoplastids (Portman & Gull, 2010; Beneke *et al.*, 2019). This feature was present in the last euglenozoan common ancestor and broadly distributed amongst the phylum (Portman &

*These authors contributed equally to this work.

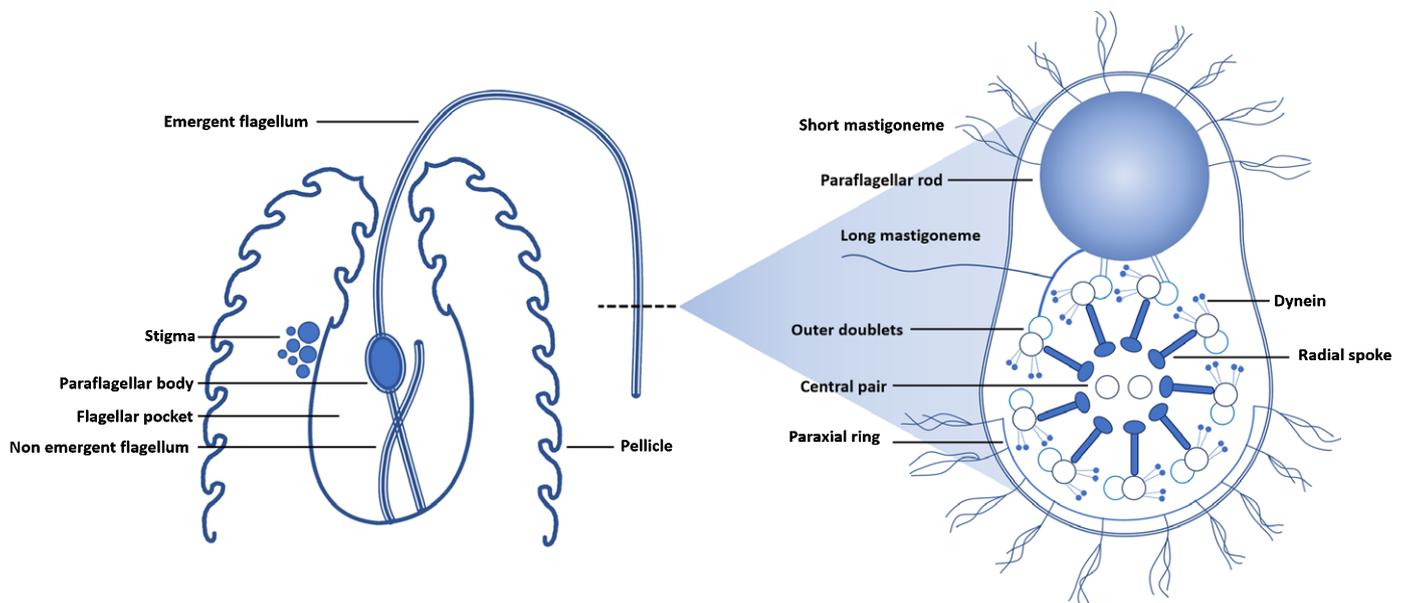


Fig. 1 (Left) Apical region of *Euglena gracilis*, displaying flagellar pocket with associated structures. (Right) Transverse section of the emergent flagellum with morphologically recognizable flagellar features.

Gull, 2010; Tashyreva *et al.*, 2018). However, unique to euglenids are series of proteinaceous strips beneath the cell membrane, referred to as the pellicle, which is underlain by a microtubule cytoskeleton (Rout & Field, 2017). Gliding of these protein strips enables undulating cell body ‘euglenoid’ movement independent of the flagella, and is employed preferentially in confined spaces (Noselli *et al.*, 2019).

The flagellum has a distinct protein composition, from both the cell body and the plasma membrane. At least two mechanisms are known that support this; transition fibres at the base maintaining a selectivity barrier and the IFT system that actively delivers specific cargo into the flagellum (Lechtreck *et al.*, 2017). The dual motile flagella of *Chlamydomonas reinhardtii* remain the most extensively studied eukaryotic unicellular flagellates, and a core of *c.* 600 proteins has been proposed for protist flagellar proteomes (Pazour *et al.*, 2005). The flagellum of the two kinetoplastids *Trypanosoma brucei* and *Leishmania major* also have been characterized extensively (Broadhead *et al.*, 2006; Dean *et al.*, 2017; Beneke *et al.*, 2019).

In order to expand our appreciation of flagellum diversity and adaptation we characterized the proteomes of the flagellum and pellicle from *E. gracilis*, selected owing to its taxonomic position, unique biology and availability of efficient purification methods. We report a large flagellar proteome and identify important structural features, including IFT, PFR and mastigoneme components. We also find signal transduction pathways coordinated by the *E. gracilis* flagella and unusual metabolic features suggesting novel functions. A small fraction of proteins are shared between the pellicle and flagellum and we conclude that the *E. gracilis* flagellum possesses surprising complexity and advances our concepts of eukaryotic flagella.

Materials and Methods

Cultivation

Euglena gracilis strain Z1 was grown at room temperature, under constant illumination as described previously (Ebenezer *et al.*, 2019). Cells were cultivated in Hutner’s medium and cell density was measured using a haemocytometer. Harvesting of cells was carried out at a density of *c.* 5×10^6 cells ml^{-1} from 500 ml cultures (2.5×10^9 cells) by centrifugation at 1500 *g* for 10 min at 4°C.

Subcellular fractionation and isolation of flagella

Cell pellets were re-suspended in 50 ml ice cold 12% ethanol (v/v) in 25 mM sodium acetate, pH 7.0, including protease inhibitor complete cocktail (Roche) (Supporting Information Fig. S1). The suspension was incubated on ice for 10 min then centrifuged at 1500 *g* at 4°C for 15 min. The supernatant containing released flagella was removed and stored on ice. The cell pellet was resuspended again in 50 ml ice cold 12% ethanol buffer, kept for 10 min on ice, centrifuged as before, and the supernatant removed and stored on ice. The pooled supernatant fractions were centrifuged three additional times, each time removing the whole-cell pellet present (indicated by the intense green colour of *E. gracilis*). The flagellar fraction then was centrifuged at 45 000 *g* in a 70 Ti rotor (Beckman Coulter) for 30 min at 4°C and flagellar pellets were resuspended in a small volume of 100 mM sodium phosphate buffer (SPB), pH 7.0, centrifuged and resuspended twice more to wash the flagella. Following removal of the supernatant, an aliquot was fixed for electron microscopy (see later). Light microscopy was used to check

for successful flagellum detachment, with treated cell pellets resuspended in 100 mM SPB, pH 7.0. Five μl of this solution were added to 595 μl SPB and 600 μl of 2% (v/v) paraformaldehyde. For untreated control cells, 50 μl culture was added to 550 μl SPB and 600 μl of 2% paraformaldehyde. Fixed cells were mounted and examined under a Zeiss Axiovert 200M microscope and images were captured with an AxioCam MRm camera and processed using ZEN PRO software (Zeiss).

Isolation of a pellicle fraction

The deflagellated cell pellet, following the above-described second wash with ethanol, was resuspended in 50 ml ice cold 100 mM SPB, pH 7.0, including two protease inhibitor cocktail tablets (Roche). The suspension was divided into two 25-ml fractions and each was subjected to probe sonication using a Qsonica125a sonicator equipped with a 3/8" disruptor horn (Qsonica, Newtown, CT, USA) with 10 bursts each of 15 s and a 1 min rest between bursts on ice throughout. Cell disruption was monitored between sonications by light microscopy until *c.* 75% of cells were broken. The sonicates then were centrifuged at 1000 *g* for 15 min at 4°C. The resultant pellets contained a lower whitish section which represents the pellicle and an upper green section, the chloroplast. The chloroplast fraction was removed and the pellicle fractions pooled and resuspended in 20 ml ice cold 100 mM SPB, including protease inhibitor cocktail tablets (Roche). Centrifugation under the same conditions was carried out twice more, removing residual chloroplasts before resuspension. The pellicle fraction was homogenized in a potter homogenizer and sonicated again with five bursts of 5 s while on ice and then loaded on top of a 75% sucrose cushion and centrifuged at 2000 *g* for 1 h at 4°C (Eppendorf 5417C). The resultant supernatant was discarded and the pellet was loaded onto a 75–100% (w/v) discontinuous sucrose gradient (75%, 85%, 90%, 95% and 100%). Gradients were centrifuged at 50 000 *g* overnight at 4°C in a SW41Ti rotor (Beckman Coulter, Carlsbad, CA, USA). The resultant gradients had four bands at varying densities (Fig. S2) and the band at the 95–100% sucrose interface was removed and diluted 10 \times using ice cold 100 mM SPB, pH 7.0. Centrifugation at 48 000 *g* for 1 h at 4°C pelleted the pellicle, which was resuspended and centrifuged twice more. After removal of the final supernatant, the pellet was resuspended in SDS-PAGE buffer and an aliquot was fixed for electron microscopy.

Production of whole-cell lysate

In order to determine the enrichment for specific proteins in the pellicle or flagellar fractions, a whole-cell lysate was prepared and analyzed. Cultures were centrifuged (cell density 5×10^6 cells ml^{-1}) and pellets were resuspended in 1.5 ml ice cold 100 mM SPB, pH 7.0, with one protease inhibitor cocktail tablet (Roche) per 20 ml. Suspensions were sonicated for 10 short bursts of 5 s each, interspersed with a 1 min pause. Centrifugation at 18 000 *g* formed pellets of chloroplast and pellicular material. Supernatants were combined and 50 μl added to 50 μl NuPAGE LDS

sample buffer before heating at 90°C for 10 min. The pelleted chloroplast and pellicles were resuspended as before and centrifuged again at 18 000 *g* to pellet once more. The pellet was resuspended in NuPAGE LDS sample buffer before a few short sonication bursts on ice to aid solubilization and heated at 90°C for 10 min before SDS-PAGE.

Transmission electron microscopy

Pellets were fixed in 4% paraformaldehyde (v/v) and 2.5% glutaraldehyde (v/v) in 0.2 M sodium cacodylate buffer, pH 7.2, for 1 h. Pellets were decimated, washed in the above buffer and post-fixed in 1.5% potassium ferricyanide (v/v) and 1% osmium tetroxide (v/v) in cacodylate buffer for 1 h, then dehydrated through an ethanol series, into propylene oxide and finally embedded in Durcupan resin. Ultrathin sections were taken from the resin and stained with 3% aqueous uranyl acetate and Reynold's lead citrate before examining on a Jeol 1200EX electron microscope. Images were collected on a SIS Megaview III camera.

Liquid chromatography/mass spectrometry

Euglena gracilis samples were supplemented with NuPAGE LDS sample buffer, sonicated (five 1-s pulses) and migrated *c.* 13 mm into a NuPAGE Bis-Tris 4–12% mini gradient polyacrylamide gel under reducing conditions (Zoltner *et al.*, 2020). Respective gel areas were cut and underwent tryptic digest and reductive alkylation. Liquid chromatography tandem mass spectrometry (LC-MS/MS) was performed in-house at the University of Dundee, UK. Samples were analyzed on a Dionex UltiMate 3000 RSLCnano System (Thermo Scientific, Waltham, MA, USA) coupled to an Orbitrap VelosPro mass spectrometer (Thermo Scientific) at the University of Dundee FingerPrints Proteomics facility and mass spectra were analyzed using MAXQUANT v.1.5 (Cox & Mann, 2008) searching the predicted *E. gracilis* proteome. Minimum peptide length was set at six amino acids, isoleucine and leucine were considered indistinguishable, and false discovery rates (FDR) of 0.01 were calculated at the levels of peptides, proteins and modification sites based on the number of hits against a reversed sequence database. Ratios were calculated from label-free quantification intensities using only peptides that could be mapped uniquely to a given protein. If the identified peptide sequence set of one protein contained the peptide set of another protein, these two proteins were assigned to the same protein group. Three thousand seven hundred and eighty seven distinct protein groups were identified in the MAXQUANT analysis. Proteomics data were deposited to the ProteomeXchange Consortium by the PRIDE (Vizcaino *et al.*, 2016) partner repository with the dataset identifier PXD024952.

Bioinformatics

Proteins identified in *E. gracilis* were queried against the NCBI nonredundant protein database using BLASTP homology searches, using the top hit (cutoff E^{-2}). BLAST2GO automatic functional

annotation (Conesa *et al.*, 2005) and GO annotations of the best BLAST results with an E -value cutoff $> 1 \times 10^{-10}$ were generated from the GO database. Flagellar proteins were additionally functionally annotated using BLASTKOALA against the eukaryote databases (Kanehisa *et al.*, 2016). The flagellar proteome of *E. gracilis* was compared against the photosensory cilium of *M. musculus* and the combined flagellar studies of *T. brucei* via ORTHOFINDER 2.2.7 (Emms & Kelly, 2015). Protein models were acquired from PFAM (Finn *et al.*, 2016) (including protein kinase (PF00069.26), blue light under flavin (PF04940.13), adenylate cyclase (PF00211.21), calmodulin (PF00612.28), transient receptor potential (PF06011.13), cyclic phosphodiesterase (PF00233.20), adenylate kinase (PF00406.23), nucleoside diphosphate kinase (PF00334.20), dynein light chain (PF01221.19), dynein heavy chain (PF03028.16), kinesin (PF00225.24), myosin head (PF00063) and actin (PF00022)), and underwent HMM-based searches using the HMMER package v.3.1 (Eddy, 2009) against transcriptome of *E. gracilis* (Ebenezer *et al.*, 2019), using a cutoff E^{-2} . Articulins, kinesin and dynein protein families underwent a muscle alignment and maximum-likelihood trees were constructed using partial deletion of $\leq 90\%$ aligned sequences with 300 bootstrap replicates. Paraflagellar rod candidates, transition zone proteins, basal body components, axonemal cap and distal tip flagellar enzymes were determined by BLASTP searching known proteins of *T. brucei* with a cut-off threshold of E^{-5} . Sequences reported previously in *E. gracilis*, including calmodulins, articulins and protein kinase A were retrieved from Genbank and BLAST-ed against the transcriptome for highest match. Mastigoneme and GTP cyclohydrolase I sequences from *C. reinhardtii* were accessed and searched for in a similar manner. The pellicle-enriched fraction of proteins was compared via ORTHOFINDER v.2.2.7 with the coding sequence libraries of diplomonids *D. japonicum* and *Hemistasia phaecysticola*, kinetoplastids *T. brucei* and *Bodo saltans*, as well as euglenids *Euglena longa*, *Eutreptiella gymnastica* and *Rhabdomonas costata*. Dyneins and kinesin families underwent muscle protein alignments (UPGMA). Maximum-likelihood trees were generated using partial deletion of positions with $< 90\%$ site coverage, with 300 replicates for bootstrap analysis were employed. Phylogenetic trees were generated via MEGAX (Kumar *et al.*, 2018). Articulins and articulins-related proteins underwent MAFFT alignment with MRBAYES NGphylogeny (Lemoine *et al.*, 2019), and was presented via JALVIEW with CLUSTAL colours (Waterhouse *et al.*, 2009). Several sequences were chosen for motif analysis via RADAR (Madeira *et al.*, 2019).

Results and Discussion

Purification of *E. gracilis* flagella and pellicle fractions

In order to complement recent whole cell (Ebenezer *et al.*, 2019), mitochondrial (Hammond *et al.*, 2020) and chloroplast proteomes of *E. gracilis* (Novak Vanclová *et al.*, 2019), we isolated protein fractions corresponding to the flagella and pellicle. Furthermore, as a representative of the eukaryotic supergroup Excavata and member of a related lineage to the kinetoplastida, as

well as possessing a highly flexible metabolic capability and distinct environmental niche we considered *E. gracilis* as an excellent choice to extend understanding of flagellum evolution and diversity.

In order to obtain high purity flagella from *E. gracilis* cells we used the method of Ngo and Bouck, with minor modifications (1995). Upon cold-shock with ice-cold 12% ethanol in sodium acetate buffer, cells rapidly lost their flagella. Efficiency was monitored by light microscopy before and following cold-shock. Cells initially possessed the typical elongated shape with the flagellum visible, but following ethanol exposure the vast majority of cells ($> 90\%$) were rounded and deflagellated (Fig. S1). Cell bodies were disrupted by sonication and the pellicular fraction isolated by density gradient centrifugation as described (Hofmann & Bouck, 1976; Nakano *et al.*, 1987) (Fig. S2). Transmission electron microscopy and subsequent LC-MS/MS confirmed that both fractions represent high levels of enrichment and are suitable for detailed analyses (Fig. S3; Tables S1, S2).

Enrichment suggests high flagellar fraction purity

Both flagellar and pellicle fractions were subject to proteomics analysis as described previously (Hammond *et al.*, 2020) (Tables S1, S2). Flagellar proteins with an enrichment > 100 -fold were defined as exclusively flagellar. Canonical flagellar IFT complexes A and B were enriched several thousand-fold (Table S1), whereas surface pellicle-associated articulins 80 and 86 were enriched up to three-fold; this lower enrichment is likely the result of both biosynthetic material and detachment of pellicle fragments contaminating the residual cell corpses (Marrs & Bouck, 1992) (Table S2). Conversely, enrichment of definitively nonflagella components belonging to chloroplast, mitochondria or cytosol indicated minimal presence within the flagellar fraction (Fig. S4), suggesting a high degree of confidence from which we can assign proteins as flagella constituents. Furthermore, some traditional cytosol markers including polyubiquitin and glycolysis enzymes are genuine flagella constituents (Mitchell *et al.*, 2005; Long *et al.*, 2015).

The flagellar proteome of *E. gracilis* consists of 1684 protein groups, corresponding to 2369 individual proteins. Two hundred and eighty six protein groups contain more than one indistinguishable protein, whereas 1397 protein groups qualify as exclusive to the flagellar fraction (Table S1). Furthermore, 1365 flagellar protein groups (81%) were identified by more than one unique peptide (maximum of 202, with 371 protein groups identified with three or two unique peptides), 304 by a single unique peptide (18%), and 15 through at least one razor peptide (1%), demonstrating broad support for the majority of designations (Table S1).

The number of proteins confidently identified in the *E. gracilis* flagellar proteome is notable when compared to other unicellular eukaryotes, including stramenopiles, algae, ciliates and kinetoplastids (*Colpomenia bullosa* (Fu *et al.*, 2014), *Dunaliella salina* (Jia *et al.*, 2010), *Tetrahymena thermophila* (Smith *et al.*, 2005), *Leishmania mexicana* (Beneke *et al.*, 2019) respectively) as well combined analyses of *C. reinhardtii* (Pazour *et al.*, 2005; Yang

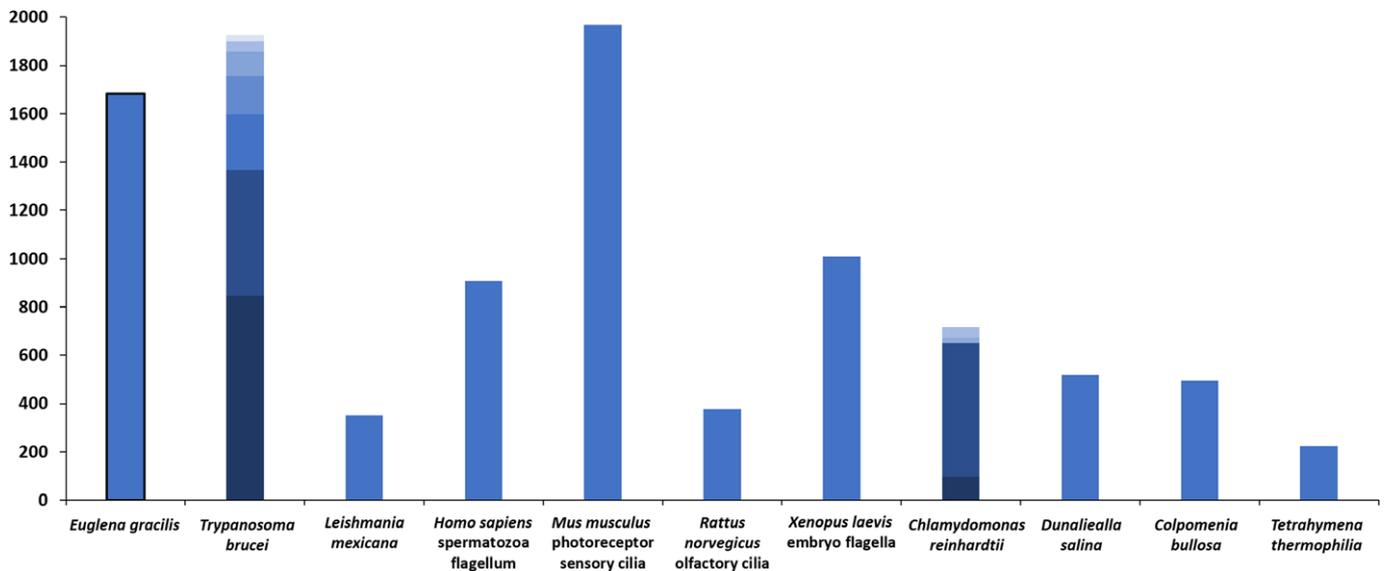


Fig. 2 Flagellar proteomes as determined for various eukaryotes. Numbers of identified proteins for *Euglena gracilis*, *Trypanosoma brucei* (Broadhead *et al.*, 2006; Zhou *et al.*, 2010; Oberholzer *et al.*, 2011; Subota *et al.*, 2014; Dean *et al.*, 2017; Velez-Ramirez *et al.*, 2021), *Leishmania mexicana* (Beneke *et al.*, 2019), *Homo sapiens* (Baker *et al.*, 2012), *Mus musculus* (Liu *et al.*, 2007), *Rattus norvegicus* (Mayer *et al.*, 2009), *Xenopus laevis* (Sim *et al.*, 2020), *Chlamydomonas reinhardtii* (Pazour *et al.*, 2005; Yang *et al.*, 2006; Lechtreck *et al.*, 2009; Zhao *et al.*, 2019), *Dunaliella salina* (Jia *et al.*, 2010), *Colpomenia bullosa* (Fu *et al.*, 2014) and *Tetrahymena thermophila* (Smith *et al.*, 2005) proteomes. Shaded regions indicate multiple studies performed in *T. brucei* and *C. reinhardtii*, with darkest lower region indicating proteins recovered in more than one analysis and lighter regions representing unique identifications from individual extractions.

et al., 2006; Lechtreck *et al.*, 2009; Zhao *et al.*, 2019), which produced proteomes between *c.* 350 and 750 proteins (Fig. 2). The *E. gracilis* flagella appears more extensive than cilia proteomes from metazoan: *c.* 1000 proteins for embryo flagella of *Xenopus laevis* and *Homo sapiens* spermatozoa (Sim *et al.*, 2020; Baker *et al.*, 2012), and >400 proteins for *Rattus norvegicus* olfactory cilia (Mayer *et al.*, 2009). This may reflect a high number of paralogous expansions (Ebenezer *et al.*, 2019) although the number of distinct gene products identified implies true functional, rather than genomic, complexity.

Comparative proteomics identifies conserved core proteins

Analyses in *T. brucei* (Broadhead *et al.*, 2006; Zhou *et al.*, 2010; Oberholzer *et al.*, 2011; Subota *et al.*, 2014; Dean *et al.*, 2017; Novák Vanclová *et al.*, 2019), identified a proteome of 1926 sequences, indicating that flagella of euglenozoan protists may generally be of considerable complexity (Fig. 2). The photosensory cilia of *Mus musculus* possess the largest current proteome of 1938, are nonmotile and specialized for light stimulated signalling (Liu *et al.*, 2007). Fewer than 500 proteins from photosensory cilia were initially reported to have orthologues among protists (Liu *et al.*, 2007). A comparison with *E. gracilis* flagella reveals an additional 219 proteins from *M. musculus* which can be added to this tally. From this it would seem that a set of 500–600 proteins constitute a conserved core eukaryotic flagella, upon which lineage-specific proteins may be added. Significantly, this number is similar to that shared between *T. brucei* and *E. gracilis* as only a small set of orthologous proteins (655) are shared,

indicating that the vast majority of components are likely parasitic vs free-living specializations.

Canonical core proteins were well-represented in the *E. gracilis* flagella, with α - and β -tubulin associated central pair proteins (Table S3). Radial spoke proteins which connect the inner and outer axonemal arms and a plethora of outer and inner dyneins, responsible for motile action are also present, though a minority of predicted dyneins and radial spoke proteins were not recovered from the flagella as well as the whole-cell lysate (Table S3). Only 698 flagellar proteins from *E. gracilis* could be assigned functional annotation through KEGG databases (Kanehisa *et al.*, 2016), which also speaks to a need for additional functional studies.

Intraflagellar transport (IFT) is an axoneme-associated bidirectional transport mechanism (Lechtreck, 2015) and was established early in eukaryogenesis and is well-represented across eukaryotes (van Dam *et al.*, 2013). Hence, together with identification of coding sequences in the genome, the IFT system can be used as a sentinel for completeness and also purity of flagellar proteomes. The IFT complex consists of three primary components: IFT-A, IFT-B and the Bardet–Biedl syndrome complex (BBSome), which recruit additional cargo for transport, as well as motor proteins of the kinesin and dynein families. This process is vital for the assembly and maintenance of flagella as well as the transport of flagellar signalling molecules and other proteins. Retrograde IFT trains additionally can be exploited for gliding motility, where flagellar regions bind external structures, temporarily recruiting attachment from nearby IFT trains, whose retrograde movement serves to pull the cell towards the substrate (Lechtreck *et al.*, 2017).

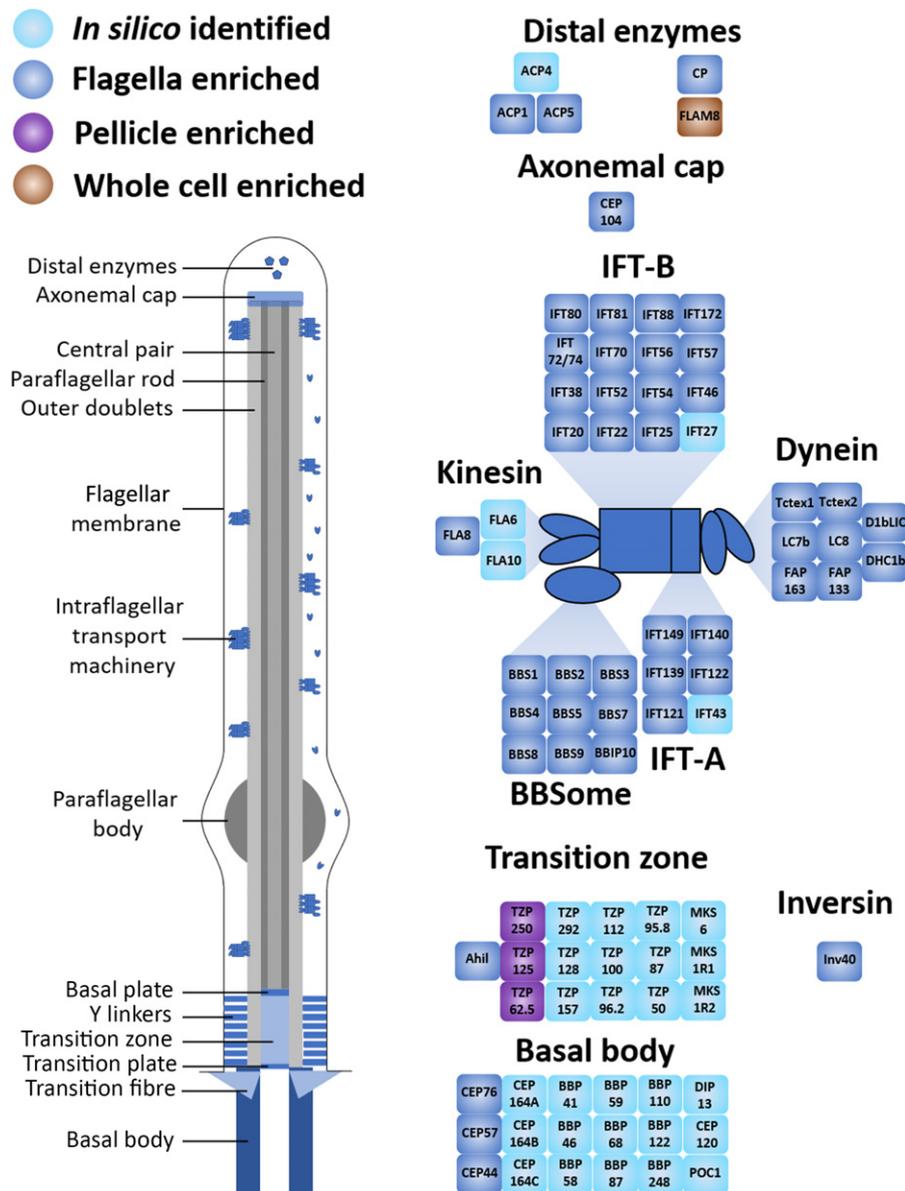


Fig. 3 Distal and basal proteins with intraflagellar transport machinery of *Euglena gracilis*. Protein coloration indicates enrichment in flagella fraction (blue), pellicle fraction (purple), whole-cell fraction (brown) or only identified *in silico* (light blue).

All canonical subunits of the cargo scaffold BBSome were recovered (Fig. 3), including GTPase BBS3, which mediates entry into the flagellum through transition fibres (Lechtreck, 2015). Only one subunit each was absent from IFT-A and IFT-B (Fig. 3). Furthermore, the multisubunit motor protein dynein was entirely recovered, whereas two subunits were not identified for kinesin (Fig. 3). Therefore, of 48 IFT and IFT-associated proteins, all but six were significantly enriched, representing 87.5% identification.

Proximity-based protein identification defines the point of deflagellation

We searched for *E. gracilis* orthologues of proteins with known locations in *T. brucei* (Varga *et al.*, 2017; Velez-Ramirez *et al.*,

2021). The nonemergent flagellum of *E. gracilis* frequently is observed connected to the emergent flagellum within the flagella pocket, which raises the possibility that, like in trypanosomatids, a connector protein complex may be present (Varga *et al.*, 2017) tethering the flagella together. However, no orthologues of this complex were recovered within the flagellum proteome (Table S3). Likewise, although no orthologues of the flagellum attachment zone (FAZ) were recovered in the proteome, *in silico* identification of several orthologous sequences in *E. gracilis* raises the possibility of these ancestral genes being repurposed within kinetoplastids (Table S3) or failing to be recovered by extraction.

A single protein of the axonemal cap, a conserved structure in all eukaryotic flagella, was identified, indicating that the flagellum tip was not disrupted during preparation (Fig. 3). Several

traditionally distal flagella enzymes were recovered including adenylate cyclases and cysteine peptidases (Saada *et al.*, 2014), whereas ‘flagellar member 8’ (Subota *et al.*, 2014) was enriched in the whole-cell lysate, suggesting alternate localization in *E. gracilis* (Fig. 3). By contrast, the majority of basal body (Dang *et al.*, 2017) and transition zone (Dean *et al.*, 2016) components were not recovered with only five of 35 components found (14%), implying that the basal region of the flagellum had not efficiently detached during deflagellation procedures. Furthermore, significant enrichment of inversin, which typically manifests immediately beyond the transition zone (Dean *et al.*, 2016), suggests a point of deflagellation distal to the transition zone (Fig. 3). Some transition zone proteins also were enriched in the pellicle extractions, indicating connections between these structures (Fig. 3).

Coordination of phototaxis, gravitaxis and free swimming

The *E. gracilis* flagella coordinate several distinct motility modes, some of which have been partially investigated and characterized (Noselli *et al.*, 2019; Yang *et al.*, 2021). Phototactic responses primarily involve stimulation of a photoactivatable adenylate cyclase (PAC) (Koumura *et al.*, 2004), which generates cyclic AMP (cAMP), stimulating protein kinase A (PKA) to phosphorylate

flagella motor proteins (Fig. 4) (Daiker *et al.*, 2011). PAC is composed of two α - and β -subunits per complex and localized to the flagellum and paraflagellar body (PFB), an invaginated region housing the photoreceptor and connecting with the flagellum via the PFR (Ntefidou & Häder, 2005). A combination of specific gene silencing (Iseki *et al.*, 2002) and mutant *Euglena* strains (Ntefidou & Häder, 2005) demonstrate that PAC mediates increased swimming velocity away from light, whereas PFB-located PAC coupled with pterins synthesized by GTP cyclohydrolase I coordinates both positive and negative phototaxis (Häder *et al.*, 2005; Ntefidou & Häder, 2005).

PKA is significantly enriched within the flagellar fraction (Fig. 4), whereas PAC subunits were recovered, but depleted in comparison to the whole-cell lysate (Fig. 4), suggesting that parts of the PFB may not purify well within the flagellar extractions, potentially remaining within the flagellar pocket upon the deflagellation procedure. GTP cyclohydrolase I, necessary only for phototaxis, was exclusively in the whole-cell lysate (Fig. 4). Previous studies have observed *Euglena* shedding emergent flagella above the PFB, near the pellicle-bearing region of the flagella pocket (referred to as the canal) (Blum, 1971; Ngo & Bouck, 1995), so it is likely that this occurred here also, which is supported by images of detached flagella (Fig. S3). Proteins coordinating a

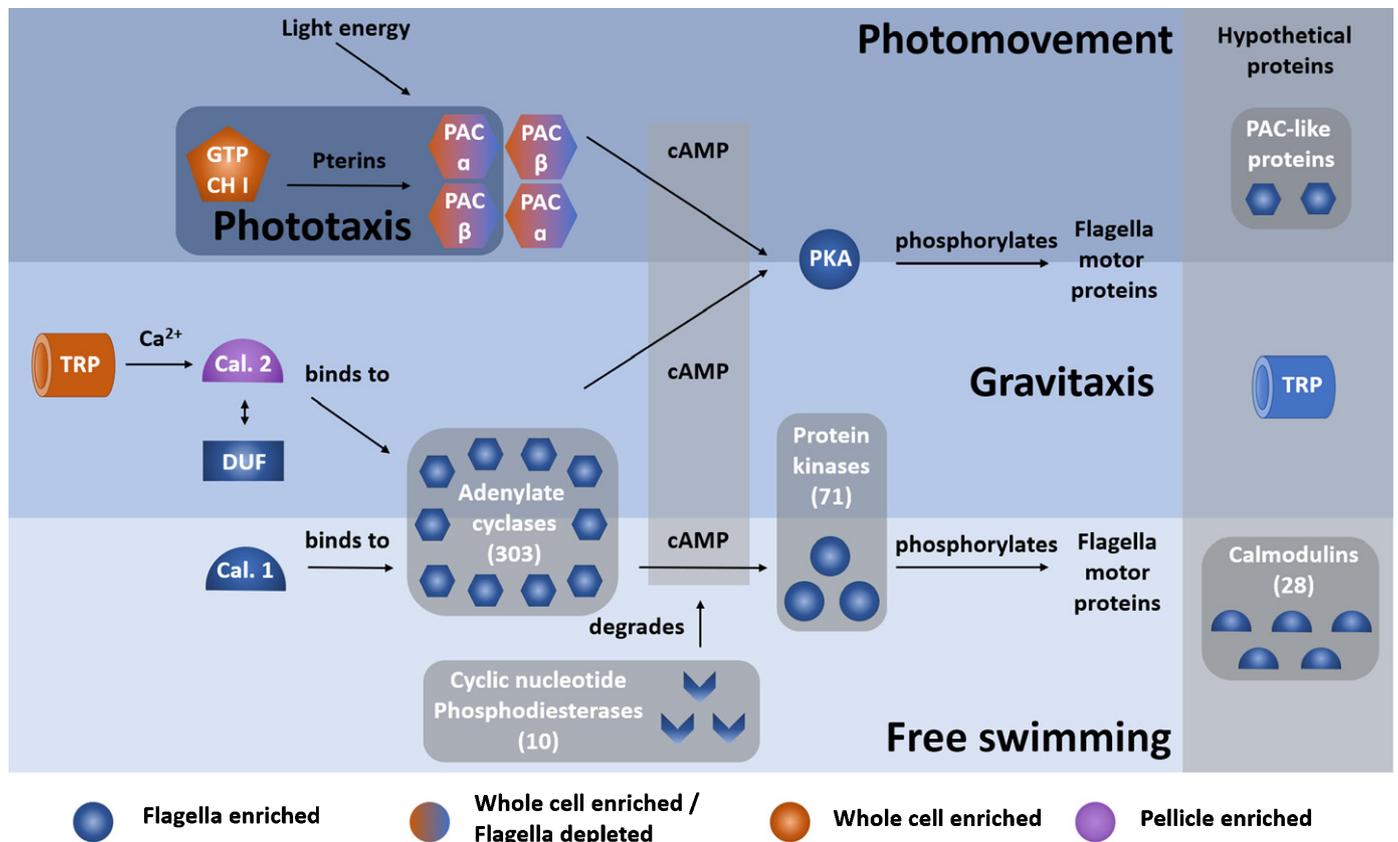


Fig. 4 Signal transduction pathways for photomovement, gravitaxis and free swimming in *Euglena*. Protein colouration indicates enrichment in flagella (blue), whole-cell lysate (brown), pellicle (purple) and whole-cell lysate enriched with flagella depleted (blue/brown). Hypothetical flagellar proteins with putative functions are indicated on the right. Guanosine triphosphate cyclohydrolase I (GTPCH I), photoactivated adenylate cyclases (PAC), protein kinase A (PKA), calmodulin (Cal), domain of unknown function (DUF) and transient receptor potential (TRP). Annotations of hypothetical PAC proteins are available in Supporting Information Fig. S5.

decrease in the swimming speed in response to light are not coordinated by PAC and are undefined. We searched for PAC paralogues with similar domain architecture within the flagellum proteome, namely dual sensors of blue light using FAD (BLUF) domains succeeded by adenylate cyclase domains (Koumura *et al.*, 2004). We identified two proteins satisfying these criteria (EG_transcript_3256 and EG_transcript_1219), which have highest similarity to PAC α - and β -subunits (Fig. S5), and represent promising candidates for coordination of this response.

Gravitaxis involves gravity-induced calcium ion flux through transient receptor potential (TRP) pores. This stimulates calmodulin-2 (Häder *et al.*, 2009), which is associated with a partner containing a domain of unknown function (DUF) (Nasir *et al.*, 2018). This protein binds to an adenylate cyclase, driving production of cAMP and stimulating PKA, the same kinase as involved in phototaxis (Daiker *et al.*, 2011), to phosphorylate flagellar motor proteins (Fig. 4). The adenylate cyclase responsible for gravitaxis-driven cAMP production has not yet been identified, but we find a large repertoire of > 300 adenylate cyclases enriched in the flagellum. Interestingly, calmodulin-2 was exclusively identified in the pellicle fraction, whereas the DUF-containing partner was present only in the flagella, raising questions over how these two proteins interact and bind adenylate cyclase. Gravity-induced calcium flux has been observed near the tip of *Euglena* cells, close to the flagellar pocket (Richter *et al.*, 2001). The TRP ion pore responsible for gravitaxis has been characterized (Häder *et al.*, 2009) and localized in this study to the whole-cell fraction (Fig. 4). Interestingly, a single TRP transmembrane containing protein was recovered within the flagella (Fig. 4).

The free-swimming transduction pathway remains the least characterized. Here, calmodulin-1 binds an adenylate cyclase which presumably activates a protein kinase, in turn phosphorylating the actino-myosin system (Daiker *et al.*, 2010). Accordingly, we identified > 70 protein kinases within the flagella (Fig. 4). Regulating signal transduction pathways for flagellar motion are the cyclic nucleotide phosphodiesterases, which degrade cAMP operating as a second messenger, to prevent indefinite flagellar motion, of which 10 are present in the flagellum (Fig. 4). An additional 28 calmodulin domain-bearing proteins with uncharacterized functions also were identified, indicating extremely complex Ca^{++} signalling pathways (Fig. 4).

Euglena gracilis flagella exhibit unique metabolic capability

Flagella require a significant amount of ATP to drive signalling pathways and motility, yet the extreme length and low volume of the flagellar lumen poses a challenge to ATP demands. Flagella employ two main mechanisms to solve this problem, either by localizing steps of the glycolytic pathway within the lumen to generate ATP *in situ* or energy shuttles to enable rapid transfer of ATP from the cell body. Although *T. brucei* makes use of three adenylate kinases for a phospho-transfer-relay system (Ginger *et al.*, 2005), nine adenylate kinases were enriched within the *E. gracilis* flagella. We additionally report four nucleoside diphosphate kinases, which in *T. brucei* and mammalian flagella

maintain the ATP/ADP/AMP homeostasis (Oberholzer *et al.*, 2007) and likely have the same function in *E. gracilis*.

The *E. gracilis* flagella possesses a full glycolytic pathway, where glycolysis is initiated by phosphoglucomutase and glucose-1-phosphate. Hexokinase, the conventional first glycolytic enzyme is seemingly absent (Fig. 5). Aside from *E. gracilis*, mammalian sperm are the only other flagella known to employ a full glycolytic pathway for energy generation, with enrichment of sperm mitochondria within the adjacent flagellar midpiece providing a clear rationale for the pathway being present; thus, to provide intermediate metabolites for the mitochondrial TCA cycle (Visconti, 2012). This prompted consideration of why *E. gracilis* employs a full glycolytic pathway in its flagella, as opposed to other organisms such as *C. reinhardtii*, with only the last three glycolytic enzymes within the flagellum (Mitchell *et al.*, 2005). A full glycolytic pathway produces two additional ATP molecules, although two more ATP are equally consumed within the entire pathway, with one ATP being consumed by hexokinase to phosphorylate glucose and another in the third step by phosphofructokinase. We suggest that *E. gracilis* deliberately utilizes phosphoglucomutase within its flagella as opposed to hexokinase, to avoid this initial ATP expenditure; by starting with phosphorylated glucose, it achieves an intraflagellar generation of net three ATP molecules per substrate, as opposed to two by other characterized flagella.

A class 2 fructose biphosphate aldolase (FBA) that catalyzes the fourth step of glycolysis is specifically enriched in the flagella (Fig. 5), whereas class 1 FBAs are present in both whole-cell lysates and plastids of *E. gracilis* (Novak Vanclová *et al.*, 2019). The presence of a class 2 FBA in certain photosynthetic protists is a hallmark of lateral gene transfer from an ancient red algal endosymbiont, as outlined in the 'chromalveolate hypothesis' (Maruyama *et al.*, 2011). The presence of multiple transient endosymbionts within the evolutionary history of euglenids, likewise argued for in the 'shopping bag hypothesis', producing a mosaic genome following a series of transient endosymbiotic events and transfer of several genes to the nuclear genome (Howe *et al.*, 2008). Thus, the flagellar FBA represents a clear example of protein retargeting from an ephemeral plastid to the nuclear genome and the flagella. A glyceraldehyde 3-phosphate dehydrogenase (GAPDH) of red algal origin that mediates the sixth step of glycolysis is likewise present in *E. gracilis*, but is localized to the plastid (Novak Vanclová *et al.*, 2019), with the flagella possessing only two conventional euglenid GAPDH orthologues (Fig. 5).

We also detected the majority of 26S proteasome subunits, involved in degradation of ubiquitinated proteins, within the flagellar fraction. A cohort of ubiquitin E1, E2 and E3 ligases, similar in number to those reported in other flagella were found, consistent with reports of ubiquitinated flagellar proteins in *Euglena* (Long *et al.*, 2015). Proteasomes are present in spermatozoa flagella, with roles in spermatogenesis and capacitation (the final maturation of released sperm for fertilization) (Kerns *et al.*, 2016). In *Giardia lamblia* proteasome subunits are found only in flagellar pores which precede flagellar assembly (Sinha *et al.*, 2015) and no proteasome subunits have been detected in the

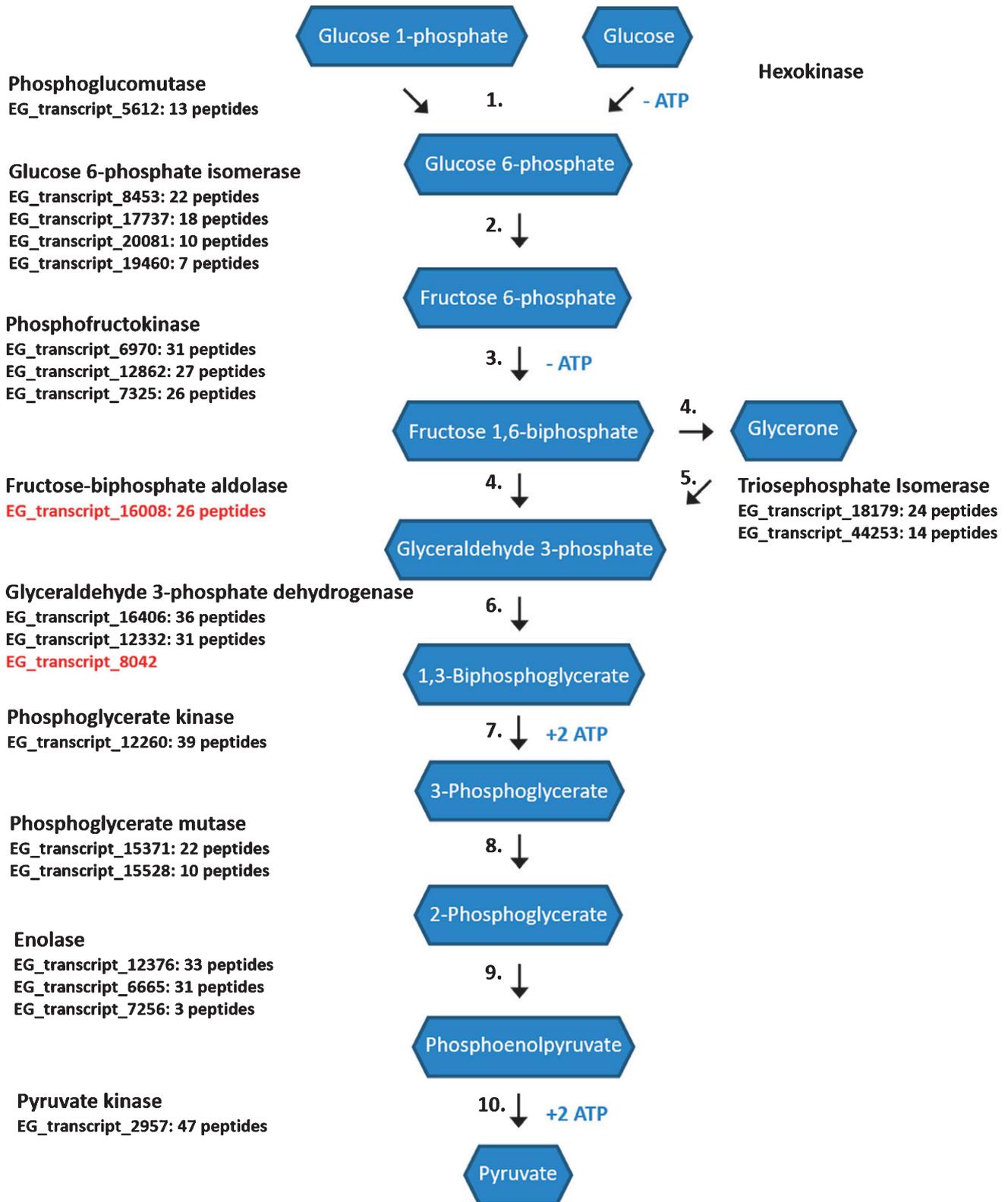


Fig. 5 Glycolytic pathway with enzymes recovered in the *Euglena* flagella. Proteins enriched in the flagella are shown with the number of corresponding peptides recovered. Enzymes with red algal ancestry are indicated in red.

C. reinhardtii flagella, suggesting that proteins ubiquitinated within the flagellum are transported to the cytoplasm via IFT for degradation (Long *et al.*, 2015). Hence detection of proteasomes

in the fully assembled flagella of *E. gracilis* suggests distinct mechanisms for flagella initiation and maturation processes representing an alternate strategy of protein degradation to *C. reinhardtii*.

The presence of proteasomes prompts speculation on how degradative activity is controlled to avoid catastrophic flagella disassembly. Additionally, as vertebrate primary cilia use specialized proteasomes to regulate ciliary signalling, we speculate that a similar mechanism may operate in *Euglena* (Gerhardt *et al.*, 2016).

The paraflagellar rod: an evolutionary flexible platform for flagellar functions

The PFR is an extra-axonemal structure of euglenozoans and participates in many roles, including metabolic, regulatory and signalling aspects. Forming a lattice-like structure on the emergent flagellum, the PFR runs parallel to the axoneme and attaches directly to outer doublets (Melkonian *et al.*, 1982). The PFR interaction with the axoneme serves to generate the distinctive *E. gracilis* flagellar wave pattern, or spinning lasso (Cicconofri *et al.*, 2021), and additionally forms a connected structure with the PFB (Hyams, 1982). In kinetoplastids, the PFR has species-specific size and structure (Maslov *et al.*, 2013), with additional functions including tethering the flagellum to the cell body, which in turn is required for proper cell morphology and replication (Kohl *et al.*, 2003). The PFR also appears to play a role in trypanosomatid attachment to insect tissues, facilitating infectious transmission (Maga & LeBowitz, 1999), and is a platform for ATP transfer to distal parts of the flagella (Portman & Gull, 2010).

In *T. brucei* > 200 proteins constitute the PFR (Maharana *et al.*, 2015; Dean *et al.*, 2017). Comparison against the *E. gracilis* transcriptome identified 94 coding sequences with high similarity to *T. brucei* queries. Of these, 55 are represented in subcellular fractions analysed here, with 45 specifically enriched in the flagellum. A further 1656 candidate PFR proteins were identified in *E. gracilis* based on BLAST searches against PFR components, which exceeded a threshold of e^{-05} . Of this larger cohort, 335 are enriched in flagellar extractions which we consider as likely PFR proteins (Table S3). This suggests a massively expanded repertoire of components constituting the PFR of *E. gracilis*, potentially reflecting a free-living lifestyle and ecological versatility. We additionally suggest that tethering PFR and PFB led to increased structural and signalling roles within phototrophic euglenids.

Mastigoneme orthologues to *C. reinhardtii*

Mastigonemes are hair-like projections located on the outside of flagella in certain eukaryotes. Their presence on flagella can serve to propel cells in the direction of the flagella tip (Hausmann *et al.*, 2014), which serves as the direction of *E. gracilis* swimming. Euglenids typically have thick investments of these hairs along their flagella (Leander *et al.*, 2017), and *E. gracilis* specifically displays two types – long and short mastigonemes – along the emergent flagellum which manifest outside of the flagella pocket. Short mastigonemes are arranged helically in two groups around both the axoneme and the PFR, whereas long mastigonemes attach to the connecting filaments between the two structures (Fig. 1). Although the composition of these mastigonemes remains elusive, they have been speculated to

be heavily composed of glycoproteins (Rosati *et al.*, 1991). Mastigonemes have been most extensively studied in stramenopiles, but differ structurally from those seen on *Euglena*, being stiff and tubular by contrast with the finer hairs seen in euglenids (Leander *et al.*, 2017). Accordingly, no protein orthologues could be identified against stramenopile mastigonemes. However, a search against the limited repertoire of *C. reinhardtii* mastigoneme components identified appropriate orthologues against mastigoneme components that were equally present in the flagellome, as well as polycystin-2 which in *C. reinhardtii* serves to anchor mastigoneme proteins to the flagella surfaces (Liu *et al.*, 2020), raising the possibility of common ancestral development shared between these two protist branches (Table S3).

Pellicle content illuminates lineage-specific proteins

Many protozoan lineages possess specialized subsurface cortical structures, classically referred to as the epiplasm. In kinetoplastids the epiplasm is composed primarily of an array of microtubules, with additional and poorly characterized proteins acting to connect with the plasma membrane. In *Toxoplasma* the epiplasm also contains actin, which has not been documented in kinetoplastids. In euglenids the epiplasm is more elaborate and consists of a series of proteinaceous interlocking bands forming a ribbed or articulated surface structure or pellicle, which is itself underlain with a microtubule corset and endoplasmic reticulum membrane (Leander *et al.*, 2017; Kostygov *et al.*, 2021).

Our purified pellicle proteome consists of 937 protein groups, corresponding to 1201 proteins (Table S2); 23% of pellicle fraction proteins contain one or more *trans*-membrane domains, contrasting with 18% for the flagella proteome and 20% for all proteins extracted from whole-cell lysates. *Euglena gracilis* articulins were the first protein components of any epiplasm to be characterized, with 80 and 86 kDa paralogues originally identified (Marrs & Bouck, 1992). Articulins are organized as an N-terminal head domain, a central VPV-repeat domain and a C-terminal tail, with repeats sometimes extending into the terminal domains. As expected, both 80 and 86 kDa articulins are enriched in the pellicle fraction, alongside a variety of > 20 divergent articulins-related paralogs (Fig. S6). Although all retain the VPV motif, additional repeat variants are clearly quite common, including ERV, VEV and VPV(I)EKIVE (Fig. S7). Moreover, there is evidence for alternating repeats in some articulins paralogues (Fig. S8). Hence the expressed articulins repertoire is considerably more complex than earlier studies suggested. Significantly, if hetero-oligomers can form, the potential complexity of articulins assembly is immense.

Other notable pellicular proteins include calmodulin 2 and 5, 18 protein kinases and 10 cyclic nucleotide phosphodiesterases (Table S3), suggestive of additional pellicle-based signal transduction pathways. Furthermore, a selection of 26 kinesins, five dyneins, two myosin motor domains and one actin also were found within the pellicle, some of which likely coordinate euglenoid movement on the microtubule corset underlying the pellicle itself, a statement supported by most of these proteins

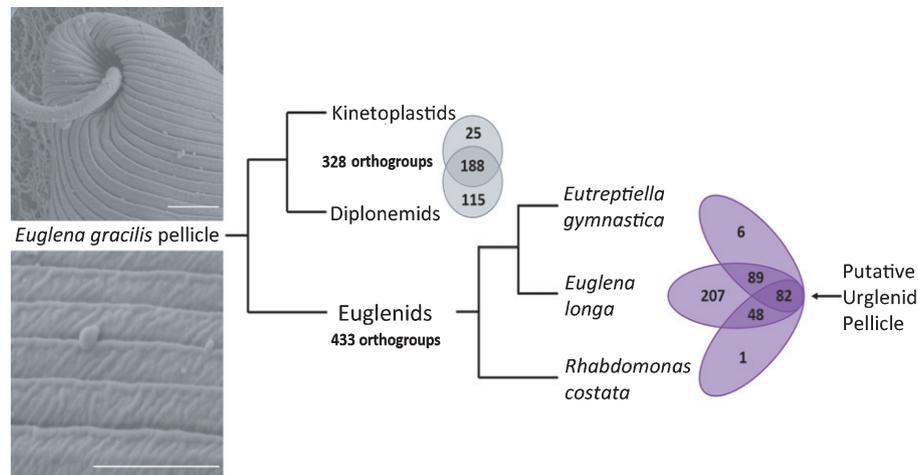


Fig. 6 Distribution of orthologues of the *Euglena gracilis* pellicle proteome in the sister groups Kinetoplastea and Diplonemea, as well as in Euglenida, for which transcriptomes are available. Proteins shared by all examined euglenids represent most likely candidates that comprised the pellicle of the last ureuglenid. Sequence distribution is available in Supporting Information Table S4. Images of *E. gracilis* provided by Eva Lacová Dobáková; bars, 1 μm .

showing exclusive pellicle enrichment (Table S3). Phylogenetic analysis of both flagella and pellicle motor proteins indicates that dyneins generally are interspersed with each other (Fig. S9), whereas a few small clades of pellicle and flagella kinesins appeared distinct from the majority of sequences, suggesting potential paralogue expansion for the kinetics connected with specific compartmental functions (Fig. S10).

Twenty-six protein groups are enriched in both the pellicle and flagellar fractions, predictably including both α - and β -tubulins (Table S3), as well as ATPases, vesicle-associated proteins, ABC transporters, an amastin and IP39. Amastins are surface glycoproteins of unknown function, initially discovered in trypanosomes, but recently demonstrated in euglenids as well (Butenko *et al.*, 2021). IP39 is a highly abundant integral membrane protein that anchors skeletal membrane proteins of *E. gracilis*, forming a crystal lattice of protein strands across the cytoplasmic side of the cell membrane (Suzuki *et al.*, 2013). IP39 enrichment in the flagellar fraction is unexpected as previous reports indicated that IP39 is not associated with the flagellar membrane (Rosiere *et al.*, 1990).

In order to determine which pellicular features are absent from kinetoplastids and diplomemids amongst Euglenozoa, we performed an extensive comparison of the pellicle proteome with transcriptomes available for the euglenids *Euglena longa*, *Eutreptiella gymnastica* and *Rhabdomonas costata*, the kinetoplastids *T. brucei* and *Bodo saltans*, as well as the diplomemids *Hemistasia phaeocysticola* and *Diplonema japonicum*. We determined 433 orthogroups present in at least one surveyed euglenid and absent from other euglenozoans (Fig. 6). This group includes articulins and the majority of articulins-related proteins, which suggests their confinement to euglenids. A total of 82 orthogroups were present in all surveyed euglenids, signifying a set of proteins that were likely present in the last common euglenid ancestor (Fig. 6). KEGG annotation of these proteins included a number of phosphorylating kinases, a single kinesin coenriched in both flagella and pellicle fractions (Fig. S8), a microtubule-associated protein and a number of traditionally nuclear-associated proteins such a

nucleoredoxin, nuclear pore complex and neuroblast differentiation-associated proteins. Conversely, a number of orthogroups conserved in diplomemids and/or kinetoplastids showed functional representation in membrane trafficking, endoplasmic reticulum and glycerophospholipid metabolism, likely indicating proteins distributed across all Euglenozoa that over time have been recruited into a close association with the pellicle (Fig. 6).

Conclusion

Comparative analyses have highlighted similarities between the flagella of mammalian sperm and trypanosomes, indicating deep and ancient homology (Oberholzer *et al.*, 2007). Our analysis of *E. gracilis* not only greatly strengthens this paradigm, but also demonstrates a dramatically increasing number of functions shared across the entire eukaryote domain, encompassing a full glycolytic pathway, the presence of the proteasome complexes and signalling pathways associated with multiple modes of locomotion. The flagella of *E. gracilis* are highly complex, likely allowing a wide range of behaviours and supporting the exceptionally broad environmental versatility of euglenids.

Acknowledgements

This work was supported by grants from the Wellcome Trust (204697/Z/16/Z to MCF), the Medical Research Council (MR/P009018/1 to MCF), the Czech Science Foundation (18-15962S to JL), ERC CZ LL1601 (to JL), the ERD project OPVVV 16_019/0000759 (to JL and MZ), and the Gordon and Betty Moore Foundation (to JL). We thank Eva Lacová Dobáková for provision of the *Euglena gracilis* images shown in Fig. 6, as well as ThankGod Ebenezer (EBI, Cambridge) for early contributions towards sequence analysis of the flagellum fraction, Lael Barlow (University of Dundee) for phylogenetic advice, Alan Prescott (University of Dundee) for his support with electron microscopy,

and also Douglas Lamont and the Fingerprints team in Dundee for excellent proteomic support.

Author contributions

MCF, MZ, MH and JL conceived the study; MZ and JG collected proteomics data; MZ, MH, EB, JG and VV analyzed the data; MH, MZ and MCF wrote the paper; and MCF, MZ, MH, JL and VV contributed to editing the paper. MH and MZ contributed equally to this work.

ORCID

Mark C. Field  <https://orcid.org/0000-0002-4866-2885>
 Michael Hammond  <https://orcid.org/0000-0001-7406-0717>
 Julius Lukeš  <https://orcid.org/0000-0002-0578-6618>
 Vladimir Varga  <https://orcid.org/0000-0002-0384-1281>
 Martin Zoltner  <https://orcid.org/0000-0002-0214-285X>

References

- Baker MA, Nixon B, Naumovski N, Aitken RJ. 2012. Proteomic insights into the maturation and capacitation of mammalian spermatozoa. *Systems Biology in Reproductive Medicine* 58: 211–217.
- Beneke T, Demay F, Hookway E, Ashman N, Jeffery H, Smith J, Valli J, Becvar T, Myskova J, Lestinova T *et al.* 2019. Genetic dissection of a *Leishmania* flagellar proteome demonstrates requirement for directional motility in sand fly infections. *PLoS Pathogens* 15: e1007828.
- Blum JJ. 1971. Existence of a breaking point in cilia and flagella. *Journal of Theoretical Biology* 33: 257–263.
- Broadhead R, Dawe HR, Farr H, Griffiths S, Hart SR, Portman N, Shaw MK, Ginger ML, Gaskell SJ, McKean PG *et al.* 2006. Flagellar motility is required for the viability of the bloodstream trypanosome. *Nature* 440: 224–227.
- Butenko A, Hammond M, Field MC, Ginger ML, Yurchenko V, Lukeš J. 2021. Reductionist pathways for parasitism in euglenozoans? Expanded datasets provide new insights. *Trends in Parasitology* 37: 100–116.
- Cicconofri G, Noselli G, DeSimone A. 2021. The biomechanical role of extra-axonemal structures in shaping the flagellar beat of *Euglena gracilis*. *eLife* 10: e58610.
- Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. 2005. Blast2go: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21: 3674–3676.
- Cox J, Mann M. 2008. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology* 26: 1367–1372.
- Daiker V, Häder DP, Richter PR, Lebert M. 2011. The involvement of a protein kinase in phototaxis and gravitaxis of *Euglena gracilis*. *Planta* 233: 1055–1062.
- Daiker V, Lebert M, Richter P, Häder DP. 2010. Molecular characterization of a calmodulin involved in the signal transduction chain of gravitaxis in *Euglena gracilis*. *Planta* 231: 1229–1236.
- van Dam TJP, Townsend MJ, Turk M, Schlessinger A, Sali A, Field MC, Huynen MA. 2013. Evolution of modular intraflagellar transport from a coatomer-like progenitor. *Proceedings of the National Academy of Sciences, USA* 110: 6943–6948.
- Dang HQ, Zhou Q, Rowlett VW, Hu H, Lee KJ, Margolin W, Li Z. 2017. Proximity interactions among basal body components in *Trypanosoma brucei* identify novel regulators of basal body biogenesis and inheritance. *MBio* 8: e02120–e2216.
- Dean S, Moreira-Leite F, Varga V, Gull K. 2016. Cilium transition zone proteome reveals compartmentalization and differential dynamics of ciliothy complexes. *Proceedings of the National Academy of Sciences, USA* 113: E5135–E5143.
- Dean S, Sunter JD, Wheeler RJ. 2017. TryTag.org: a trypanosome genome-wide protein localisation resource. *Trends in Parasitology* 33: 80–82.
- Diniz MC, Pacheco ACL, Farias KM, de Oliveira DM. 2012. The eukaryotic flagellum makes the day: novel and unforeseen roles uncovered after post-genomics and proteomics data. *Current Protein & Peptide Science* 13: 524–546.
- Ebenezer TE, Zoltner M, Burrell A, Nenarokova A, Novák Vanclová A, Prasad B, Soukal P, Santana-Molina C, O'Neill E, Nankissoor NN *et al.* 2019. Transcriptome, proteome and draft genome of *Euglena gracilis*. *BMC Biology* 17: 11.
- Eddy SR. 2009. A new generation of homology search tools based on probabilistic inference. *Genome Informatics Japan* 23: 205–211.
- Emms DM, Kelly S. 2015. Orthofinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology* 16: 157.
- Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A *et al.* 2016. The pfam protein families database: towards a more sustainable future. *Nucleic Acids Research* 44 (D1): D279–D285.
- Fu G, Nagasato C, Oka S, Cock JM, Motomura T. 2014. Proteomics analysis of heterogeneous flagella in brown algae (stramenopiles). *Protist* 165: 662–675.
- Gerhardt C, Wiegering A, Leu T, Rütger U. 2016. Control of hedgehog signalling by the cilia-regulated proteasome. *Journal of Developmental Biology* 4: 27.
- Ginger ML, Ngazoa ES, Pereira CA, Pullen TJ, Kabiri M, Becker K, Gull K, Steverding D. 2005. Intracellular positioning of isoforms explains an unusually large adenylate kinase gene family in the parasite *Trypanosoma brucei*. *Journal of Biological Chemistry* 280: 11781–11789.
- Häder DP, Ntefidou M, Iseki M, Watanabe M. 2005. Phototaxis photoreceptor in euglena gracilis. In: Wada M, Shimazaki K, Lino M, eds. *Light sensing in plants*. Tokyo, Japan: Springer, 223–229.
- Häder DP, Richter PR, Schuster M, Daiker V, Lebert M. 2009. Molecular analysis of the graviperception signal transduction in the flagellate *Euglena gracilis*: involvement of a transient receptor potential-like channel and a calmodulin. *Advances in Space Research* 43: 1179–1184.
- Hammond MJ, Nenarokova A, Butenko A, Zoltner M, Dobáková EL, Field MC, Lukeš J. 2020. A uniquely complex mitochondrial proteome from *Euglena gracilis*. *Molecular Biology and Evolution* 37: 2173–2191.
- Hausmann KR, Alimenti R, Antipa C, Boenigk G, Fujishima J, Hans-Dieter M, Kodama Y, Luporini P, Lynn D, Macherem H *et al.* 2014. Motility. In: Hausmann KR, Renate R, eds. *Cilia and flagella – ciliates and flagellates: ultrastructure and cell biology, function and systematics, symbiosis and biodiversity*. Berlin, Germany: Schweizerbart Science, 121–147.
- Hofmann C, Bouck GB. 1976. Immunological and structural evidence for patterned instusussceptive surface growth in a unicellular organism – postulated role for submembranous proteins and microtubules. *Journal of Cell Biology* 69: 693–715.
- Howe CJ, Barbrook AC, Nisbet RER, Lockhart PJ, Larkum AWD. 2008. The origin of plastids. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 363: 2675–2685.
- Hyams JS. 1982. The *Euglena* paraflagellar rod: structure, relationship to other flagellar components and preliminary biochemical characterization. *Journal of Cell Science* 55: 199–210.
- Iseki M, Matsunaga S, Murakami A, Ohno K, Shiga K, Yoshida K, Sugai M, Takahashi T, Hori T, Watanabe M. 2002. A blue-light-activated adenylyl cyclase mediates photoavoidance in *Euglena gracilis*. *Nature* 415: 1047–1051.
- Jia YL, Xue LX, Li J, Liu HT. 2010. Isolation and proteomic analysis of the halotolerant alga *Dunaliella salina* flagella using shotgun strategy. *Molecular Biology Reports* 37: 711–716.
- Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and ghostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *Journal of Molecular Biology* 428: 726–731.
- Kerns K, Morales P, Sutovsky P. 2016. Regulation of sperm capacitation by the 26s proteasome: an emerging new paradigm in spermatology. *Biology of Reproduction* 94: 117.

- Kohl L, Robinson D, Bastin P. 2003. Novel roles for the flagellum in cell morphogenesis and cytokinesis of trypanosomes. *EMBO Journal* 22: 5336–5346.
- Kostygov AY, Karnkowska A, Votýpka J, Tashyreva D, Maciszewski K, Yurchenko V, Lukeš J. 2021. Euglenozoa: taxonomy, diversity and ecology, symbioses and viruses. *Open Biology* 11: 200407.
- Koumura Y, Suzuki T, Yoshikawa S, Watanabe M, Iseki M. 2004. The origin of photoactivated adenylyl cyclase (PAC), the euglena blue-light receptor: phylogenetic analysis of orthologues of pac subunits from several euglenoids and trypanosome-type adenylyl cyclases from *Euglena gracilis*. *Photochemical & Photobiological Sciences* 3: 580–586.
- Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. Mega X: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* 35: 1547–1549.
- Leander BS, Lax G, Karnkowska A, Simpson AGB. 2017. Euglenida. In: Archibald JM, Simpson AGB, Slamovits CH, eds. *Handbook of the protists*. Cham, Switzerland: Springer International, 1047–1088.
- Lechtreck KF. 2015. IFT-cargo interactions and protein transport in cilia. *Trends in Biochemical Sciences* 40: 765–778.
- Lechtreck KF, Johnson EC, Sakai T, Cochran D, Ballif BA, Rush J, Pazour GJ, Ikebe M, Witman GB. 2009. The *Chlamydomonas reinhardtii* BBSome is an IFT cargo required for export of specific signaling proteins from flagella. *Journal of Cell Biology* 187: 1117–1132.
- Lechtreck KF, Van De Weghe JC, Harris JA, Liu PW. 2017. Protein transport in growing and steady-state cilia. *Traffic* 18: 277–286.
- Lemoine F, Correia D, Lefort V, Doppelt-Azeroual O, Mareuil F, Cohen-Boulakia S, Gascuel O. 2019. Ngphylogeny: Fr: new generation phylogenetic services for non-specialists. *Nucleic Acids Research* 47: W260–W265.
- Liu PW, Lou XC, Wingfield JL, Lin JF, Nicastro D, Lechtreck K. 2020. *Chlamydomonas* pkd2 organizes mastigonemes, hair-like glycoprotein polymers on cilia. *Journal of Cell Biology*. 219: e202001122.
- Liu Q, Tan G, Levenkova N, Li T, Pugh EN, Rux JJ, Speicher DW, Pierce EA. 2007. The proteome of the mouse photoreceptor sensory cilium complex. *Molecular & Cellular Proteomics* 6: 1299–1317.
- Long H, Wang QY, Huang KY. 2015. Ciliary/flagellar protein ubiquitination. *Cells* 4: 474–482.
- Madeira F, Park Ym, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD *et al.* 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research* 47: W636–W641.
- Maga JA, LeBowitz JH. 1999. Unravelling the kinetoplastid paraflagellar rod. *Trends in Cell Biology* 9: 409–413.
- Maharana BR, Tewari AK, Singh V. 2015. An overview on kinetoplastid paraflagellar rod. *Journal of Parasitic Diseases* 39: 589–595.
- Marrs JA, Bouck GB. 1992. The 2 major membrane skeletal proteins (articulins) of *Euglena gracilis* define a novel class of cytoskeletal proteins. *Journal of Cell Biology* 118: 1465–1475.
- Maruyama S, Suzuki T, Weber APM, Archibald JM, Nozaki H. 2011. Eukaryote-to-eukaryote gene transfer gives rise to genome mosaicism in euglenids. *BMC Evolutionary Biology* 11: 105.
- Maslov DA, Votýpka J, Yurchenko V, Lukeš J. 2013. Diversity and phylogeny of insect trypanosomatids: all that is hidden shall be revealed. *Trends in Parasitology* 29: 43–52.
- Mayer U, Küller A, Daiber PC, Neudorf I, Warnken U, Schnolzer M, Frings S, Mohrlen F. 2009. The proteome of rat olfactory sensory cilia. *Proteomics* 9: 322–334.
- Melkonian M, Robenek H, Rassat J. 1982. Flagellar membrane specializations and their relationship to mastigonemes and microtubules in *Euglena gracilis*. *Journal of Cell Science* 55: 115–135.
- Mitchell BF, Pedersen LB, Feely M, Rosenbaum JL, Mitchell DR. 2005. Atp production in *Chlamydomonas reinhardtii* flagella by glycolytic enzymes. *Molecular Biology of the Cell* 16: 4509–4518.
- Nakano Y, Urade Y, Urade R, Kitaoka S. 1987. Isolation, purification, and characterization of the pellicle of *Euglena gracilis* z. *Journal of Biochemistry* 102: 1053–1063.
- Nasir A, Le Bail A, Daiker V, Klima J, Richter P, Lebert M. 2018. Identification of a flagellar protein implicated in the gravitaxis in the flagellate *Euglena gracilis*. *Scientific Reports* 8: 7605.
- Ngo HM, Bouck GB. 1995. Isolation of *Euglena* flagella. *Methods in Cell Biology* 47: 25–29.
- Noselli G, Beran A, Arroyo M, DeSimone A. 2019. Swimming *Euglena* respond to confinement with a behavioural change enabling effective crawling. *Nature Physics* 15: 496–502.
- Novák Vanclová AMG, Zoltner M, Kelly S, Soukal P, Záhonová K, Füssy Z, Ebenezer TE, Lacová Dobáková E, Eliáš M, Lukeš J *et al.* 2019. Metabolic quirks and the colourful history of the *Euglena gracilis* secondary plastid. *New Phytologist* 225: 1578–1592.
- Ntefidou M, Häder DP. 2005. Photoactivated adenylyl cyclase (pac) genes in the flagellate *Euglena gracilis* mutant strains. *Photochemical & Photobiological Sciences* 4: 732–739.
- Oberholzer M, Bregy P, Marti G, Minca M, Peier M, Seebeck T. 2007. Trypanosomes and mammalian sperm: one of a kind? *Trends in Parasitology* 23: 71–77.
- Oberholzer M, Langousis G, Nguyen HT, Saada EA, Shimogawa MM, Jonsson ZO, Nguyen SM, Wohlschlegel JA, Hill KL. 2011. Independent analysis of the flagellum surface and matrix proteomes provides insight into flagellum signaling in mammalian-infectious *Trypanosoma brucei*. *Molecular & Cellular Proteomics* 10: M111.010538.
- Pazour GJ, Agrin N, Leszyk J, Witman GB. 2005. Proteomic analysis of a eukaryotic cilium. *Journal of Cell Biology*. 170: 103–113.
- Portman N, Gull K. 2010. The paraflagellar rod of kinetoplastid parasites: from structure to components and function. *International Journal for Parasitology* 40: 135–148.
- Richter P, Lebert M, Tahedl H, Häder DP. 2001. Calcium is involved in the gravitactic orientation in colorless flagellates. *Journal of Plant Physiology* 158: 689–697.
- Rosati G, Verni F, Barsanti L, Passarelli V, Gualtieri P. 1991. Ultrastructure of the apical zone of *Euglena gracilis* – photoreceptors and motor apparatus. *Electron Microscopy Reviews* 4: 319–342.
- Rosiere TK, Marrs JA, Bouck GB. 1990. A 39-kD plasma membrane protein (ip39) is an anchor for the unusual membrane skeleton of *Euglena gracilis*. *Journal of Cell Biology* 110: 1077–1088.
- Rout MP, Field MC. 2017. The evolution of organellar coat complexes and organization of the eukaryotic cell. *Annual Review of Biochemistry* 86: 637–657.
- Saada EA, Kabutu ZP, Lopez M, Shimogawa MM, Langousis G, Oberholzer M, Riestra A, Jonsson ZO, Wohlschlegel JA, Hill KL. 2014. Insect stage-specific receptor adenylyl cyclases are localized to distinct subdomains of the *Trypanosoma brucei* flagellar membrane. *Eukaryotic Cell* 13: 1064–1076.
- Sim HJ, Yun S, Kim HE, Kwon KY, Kim G-H, Yun S, Kim BG, Myung K, Park TJ, Kwon T. 2020. Simple method to characterize the ciliary proteome of multiciliated cells. *Journal of Proteome Research* 19: 391–400.
- Sinha A, Datta SP, Ray A, Sarkar S. 2015. A reduced VWA domain-containing proteasomal ubiquitin receptor of *Giardia lamblia* localizes to the flagellar pore regions in microtubule-dependent manner. *Parasites & Vectors* 8: 120.
- Smith JC, Northey JGB, Garg J, Pearlman RE, Siu KWM. 2005. Robust method for proteome analysis by MS/MS using an entire translated genome: demonstration on the ciliome of *Tetrahymena thermophila*. *Journal of Proteome Research* 4: 909–919.
- Subota I, Julkowska D, Vincensini L, Reeg N, Buisson J, Blisnick T, Huet D, Perrot S, Santi-Rocca J, Duchateau M *et al.* 2014. Proteomic analysis of intact flagella of procyclic *Trypanosoma brucei* cells identifies novel flagellar proteins with unique sub-localization and dynamics. *Molecular & Cellular Proteomics* 13: 1769–1786.
- Suzuki H, Ito Y, Yamazaki Y, Mineta K, Uji M, Abe K, Tani K, Fujiyoshi Y, Tsukita S. 2013. The four-transmembrane protein ip39 of *Euglena* forms strands by a trimeric unit repeat. *Nature Communications* 4: 1766.
- Tashyreva D, Prokopchuk G, Yabuki A, Kaur B, Faktorová D, Votýpka J, Kusaka C, Fujikura K, Shiratori T, Ishida K-I *et al.* 2018. Phylogeny and morphology of new diplomonids from Japan. *Protist* 169: 158–179.
- Varga V, Moreira-Leite F, Portman N, Gull K. 2017. Protein diversity in discrete structures at the distal tip of the trypanosome flagellum. *Proceedings of the National Academy of Sciences, USA* 114: E6546–E6555.
- Vélez-Ramírez DE, Shimogawa MM, Ray SS, Lopez A, Rayatpisheh S, Langousis G, Gallagher-Jones M, Dean S, Wohlschlegel JA, Hill KL. 2021.

- APEX2 proximity proteomics resolves flagellum subdomains and identifies flagellum tip-specific proteins in *Trypanosoma brucei*. *mSphere* 6: 33568455.
- Verni F, Rosati G, Lenzi P, Barsanti L, Passarelli V, Gualtieri P. 1992. Morphological relationship between paraflagellar swelling and paraxial rod in *Euglena gracilis*. *Micron and Microscopica Acta* 23: 37–44.
- Visconti PE. 2012. Sperm bioenergetics in a nutshell. *Biology of Reproduction* 87: 72.
- Vizcaíno JA, Csordas A, del-Toro N, Dienes JA, Griss J, Lavidas I, Mayer G, Perez-Riverol Y, Reisinger F, Ternent T *et al.* 2016. 2016 update of the pride database and its related tools. *Nucleic Acids Research* 44: D447–D456.
- Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25: 1189–1191.
- Waters AM, Beales PL. 2011. Ciliopathies: an expanding disease spectrum. *Pediatric Nephrology* 26: 1039–1056.
- Yang PF, Diener DR, Yang C, Kohno T, Pazour GJ, Dienes JM, Agrin NS, King SM, Sale WS, Kamiya R *et al.* 2006. Radial spoke proteins of *Chlamydomonas* flagella. *Journal of Cell Science* 119: 1165–1174.
- Yang SY, Huang MJ, Zhao YF, Zhang HP. 2021. Controlling cell motion and microscale flow with polarized light fields. *Physical Review Letters* 126: 058001.
- Zhao L, Hou YQ, Picariello T, Craige B, Witman GB. 2019. Proteome of the central apparatus of a ciliary axoneme. *Journal of Cell Biology*. 218: 2051–2070.
- Zhou Q, Gheiratmand L, Chen YX, Lim TK, Zhang J, Li SW, Xia NS, Liu BH, Lin QS, He CY. 2010. A comparative proteomic analysis reveals a new bi-lobe protein required for bi-lobe duplication and cell division in *Trypanosoma brucei*. *PLoS ONE* 5: 12.
- Zoltner M, Del Pino RC, Field MC. 2020. Sorting the muck from the brass: analysis of protein complexes and cell lysates. *Methods in Molecular Biology* 2116: 645–653.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Light microscopy of aflagellated *E. gracilis* cells after cold-shock ethanol treatment.

Fig. S2 Density gradient centrifugation for isolation of the pellicle fraction.

Fig. S3 Transmission electron microscopy isolated flagellum and pellicle fractions.

Fig. S4 Volcano plots of flagella and pellicle proteomes.

Fig. S5 Domain architecture of photoactivated adenylyl cyclase subunits, as well as hypothetical candidate subunits for mediating light based ‘step down’ response.

Fig. S6 Phylogenetic distribution of articulins and articulins-related proteins of *E. gracilis* pellicle.

Fig. S7 Alignment of articulins and articulins-related proteins of *E. gracilis*.

Fig. S8 Repeated motifs and sequences displayed in select articulins and articulins-related proteins.

Fig. S9 Phylogenetic distribution of dynein motor proteins from *E. gracilis* flagella and pellicle fractions.

Fig. S10 Phylogenetic distribution of kinesin motor proteins from *E. gracilis* flagella and pellicle fractions.

Table S1 *Euglena gracilis* protein groups enriched in flagella fraction.

Table S2 *Euglena gracilis* protein groups enriched in pellicle fraction.

Table S3 Identified *E. gracilis* proteins of interest from study, enriched in either the flagella, pellicle or the whole-cell lysate.

Table S4 ORTHOFINDER distribution of pellicle orthogroups across euglenozoan species.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.