Trends in Genetics

CellPress OPEN ACCESS

Review

Advances in euglenoid genomics: unravelling the fascinating biology of a complex clade

Oskar Fields[1,2,5], Michael J. Hammond[3,4,5], Xiao Xu[1,2,5], and Ellis C. O'Neill[1,2,*]

Euglenids have long been studied due to their unique physiology and versatile metabolism, providing underpinnings for much of our understanding of photosynthesis and biochemistry, and a growing opportunity in biotechnology. Until recently there has been a lack of genetic studies due to their large and complex genomes, but recently new technologies have begun to unveil their genetic capabilities. Whilst much research has focused on the model organism *Euglena gracilis*, other members of the euglenids have now started to receive due attention. Currently only poor nuclear genome assemblies of *E. gracilis* and *Rhabdomonas costata* are available, but there are many more plastid genome sequences and an increasing number of transcriptomes. As more assemblies become available, there are great opportunities to understand the fundamental biology of these organisms and to exploit them for biotechnology.

Highlights

Euglenids have unique physiology and versatile metabolism, with diverse lifestyles including phototrophy, osmotrophy, and predatory behaviour.

Recent work on genome sequencing of the model *Euglena gracilis* shows the complexity of the genome; it has proved challenging to assemble, and the organellar genomes also have unique features.

Transcriptomes show the diversity in other euglenids and their capabilities.

Our understanding of euglenid physiology and cell biology has been helped by comparison with related organisms.

Why study euglenids?

Euglenids are a class of protozoa that predominantly live in freshwater environments, and together with the kinetoplastids – including unicellular parasites *Trypanosoma* and *Leishmania*, diplonemids, widespread marine plankton, and the poorly defined symbiontids – make up the phylum Euglenozoa [1]. Rather than a cell wall, euglenids are surrounded by a flexible protein-based pellicle containing glycoproteins and polysaccharides [2,3]. Many possess a green plastid derived by secondary endosymbiosis of a chlorophyte alga, estimated to have occurred between 652 and 539 million years ago [4]; in some lineages the chloroplast can be lost without compromising viability [5].

Euglenids have been subject to scientific study for hundreds of years, and *Euglena gracilis* is a key model in protist genetics and cellular biology due to its unique physiological features, including distinctive euglenoid movement (metaboly), biochemistry, and production of a unique storage polysaccharide, paramylon [6]. There is growing interest in the biotechnological exploitation of euglenids [7,8] and their use to produce valuable products [9], including biofuels [10] and even cytotoxins for cancer treatment [11]. Recent advances in sequencing of the model organism *E. gracilis* have begun to shine a light on the underlying genetics that gives them these fascinating capabilities (Figure 1). Whilst this model has been the focus of most of the research, genetic research has been carried out on specific euglenids which display interesting physiology or metabolism (Table 1).

Genomes

Nuclear genomes

Euglena gracilis. The nuclear genome of *E. gracilis*, the model euglenid species, was considered an enigma for many years, but advances in technology are now starting to uncover its secrets. Slow progress in genome elucidation for *E. gracilis* has been due to the organism possessing a large genome (estimated at >1.5 Gbp), with a high prevalence of repetitive sections and the hypermodified 'base J' (β-D-glucopyranosyloxymethyluracil). Base J is a hypermodified nucleobase that is found across the phylum Euglenophyta and is thought to prevent local

[1]School of Chemistry, University of Nottingham, University Park, Nottingham, NG7 2RD, UK
[2]Biodiscovery Institute, University of Nottingham, University Park, Nottingham, NG7 2RD, UK
[3]Institute of Parasitology, Biology Centre, Czech Academy of Sciences, České Budějovice (Budweis), Czech Republic
[4]Faculty of Science, University of South Bohemia, České Budějovice (Budweis), Czech Republic
[5]These authors contributed equally.

*Correspondence:
ellis.oneill@nottingham.ac.uk (E.C. O'Neill).

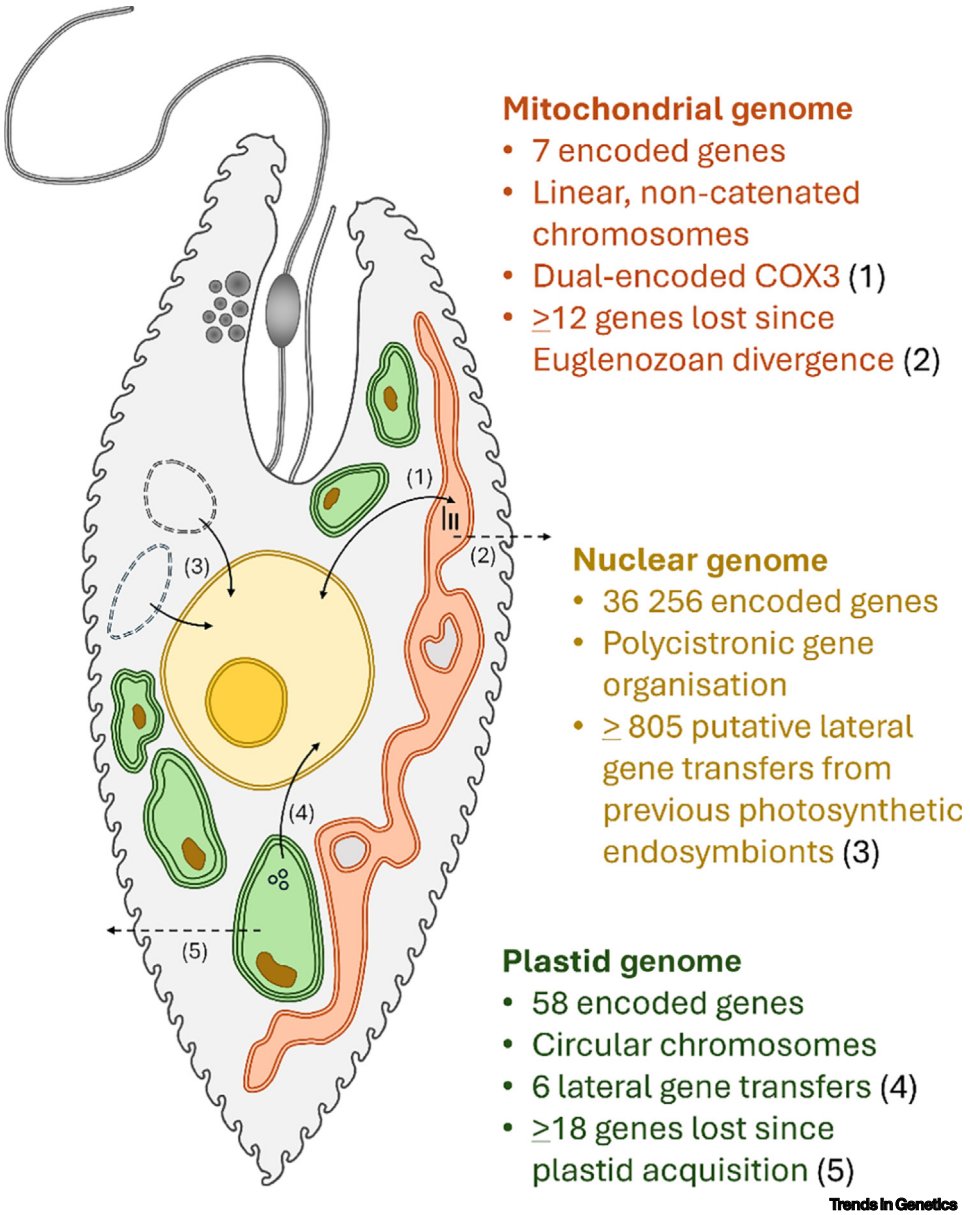

Figure 1. The genomes of *Euglena gracilis*. There is an extra copy of the COX3 gene transferred from the mitochondrial genome to the nucleus (1) and a loss of at least 12 mitochondrial genes since the last eukaryotic common ancestor (2). Many of the nuclear genes appear to have been acquired from photosynthetic endosymbionts which have since been lost (3). The chloroplast has developed from acquisition of a green algal endosymbiont, subsequently transferring six of the plastid genes to the nucleus (4) and losing at least 18 others (5).

transcription [12]. Trypanosome and euglenid genomes both contain similar rates of base J modification (replacing approximately 1% of all thymidines) located in inactive telomeric regions in trypanosomes and spread more uniformly across the *Euglena* genome [12,13].

Draft genomes have been constructed for *E. gracilis* with an estimated total size of 2.3–2.4 Gbp, which sits towards the lower end of previously speculated values (1.0–9.0 Gbp) [14,15]. The

Table 1. Progress on sequencing various euglenids (see text for details)

Species	Nuclear genome	Mitochondrial genome	Plastid genome	Transcriptome
Euglena gracilis	✓	✓	✓	✓
Rhabdomonas costata	✓		Not present?	✓
Euglena mutabilis			✓	✓
Euglena longa			✓	✓
Rapaza viridis			✓	✓
Eutreptiella gymnastica			✓	✓
Many other euglenids			✓	

nuclear genome has an estimated haploid size of 330–500 Mbp, with 140–160 Mbp of single-copy DNA [14], organised into 42–46 linear chromosomes [13,15]. When *E. gracilis* transcripts were mapped onto the early genome assemblies, initial methods produced low conformity, but when programs used for assembling polyploid genomes were utilised the conformity increased [14]. Given the estimated sizes of the haploid and total genomes, and the distribution of repetitive elements, it is currently thought that *E. gracilis* could be polypoid [8]. This may also be responsible for the lack of sexual reproduction observed in *E. gracilis*, as triploid organisms struggle to evenly distribute chromosomes, producing non-viable gametes. However, it should be noted that some species of trypanosomes display aneuploidy, an alternative possibility for *E. gracilis* [16].

The genome contains 36 526 unique coding sequences which make up <1% of the total genome size [14]. Open reading frames (ORFs) are interspersed between long repetitive sequences with high AT content, which interact to induce the hyperchromicity observed in euglena DNA [14,17]. The highly prevalent repeats also explain the disparity between transcriptome GC content (~60%) and the genome GC content (~50%) [14,18].

Unlike most organisms, euglenozoan protein abundance is not regulated by gene expression, but predominantly through post-transcriptional factors [14]. This may have affected a number of genetic traits such as polycistronic genes, which are usually uncommon in eukaryotes, and the presence of non-conventional introns, which have only been found in euglenozoans [14,19,20]. The polycistronic genes in trypanosomes, relatives of *E. gracilis*, encode proteins with unrelated biosynthesis pathways or molecular functions [21]. Trypanosomes also process their polycistronic genes post-transcriptionally using trans splicing to produce monocistronic mRNA [21]. This process could also occur in *E. gracilis* as conventional and non-conventional, cis and trans splicing have been observed, giving *E. gracilis* a large degree of optionality regarding its mRNA processing [14].

Rhabdomonas costata. Genomic analysis has also been performed for *R. costata*, a heterotrophic euglenid [22]. Unfortunately, the genome that was produced was fragmented, most likely due to bias during whole genome amplification (WGA) or the presence of highly repetitive regions [22]. This meant that no estimates could be made for chromosome number or ploidy for the organism, although the estimated size of the haploid genome is 250 Mbp [22]. This is significantly smaller than that of *E. gracilis*, although it still encodes a predicted 39 456 proteins which is comparable with its photosynthetic relative [22]. The *R. costata* genome was also found to contain novel non-conventional introns similar to those in *E. gracilis*. It has not been possible to confirm whether there is any link between gene and protein expression or whether these introns had any effect on it [20,22].

Mitochondrial genomes

The mitochondrial genome of *E. gracilis* was recently sequenced and found to be composed of linear fragments of 5–8 kbp, including only seven protein-coding genes, fragmented across these pieces of DNA [23]. All of these proteins are respiratory complex subunits, with an additional copy of cox3 encoded in the nucleus, with the expected mRNA processing [24]. This genome is extremely small when compared with that of many other organisms, and indicates large amounts of gene transfer to the nuclear genome. Although the related kinetoplastids and diplonemids carry out extensive RNA editing in their mitochondria, there is no evidence for this in *E. gracilis* [23]. Given that euglenid evolution has involved multiple endosymbiotic events, perhaps this has induced the alga to develop efficient gene transfer mechanisms which are responsible for the lack of mitochondrial genes.

Chloroplast genomes

The *E. gracilis* chloroplast genome was one of the first chloroplast genomes to be isolated and characterised, with its high GC content making it more buoyant and thus easier to separate from the residual DNA [25]. The circular genome is approximately 143 kbp in length and encodes 58 proteins and 30 unique RNAs [26]. This is 23 fewer protein-coding genes than in the genome of the chloroplast of *Pyramimonas parkeae*, the closest living algal relative of the euglenid plastid [27]. Six of these genes have been transferred to the nucleus of the euglenid host, while 18 have apparently been lost. The chloroplast genome possesses high numbers of group II (82) and group III (64) introns, with many of these forming twintrons (intron-inside-introns) [26]. Group III introns were first discovered from euglenid chloroplasts and have been identified in only a few other alga chloroplasts to date. They appear to have evolved from group II introns and act using the same catalytic mechanism but are much more compact, requiring only 95–110 nt compared with 277 nt (the smallest group II intron found in *E. gracilis*) [26].

There are currently 39 full or partial sequences of euglenid chloroplasts currently available, with substantially similar gene content, barring the loss of some photosynthetic elements in non-photosynthetic lineages, such as *Euglena* (*Astasia) longa* [28], which still retains a functional ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBISCO). However, the genomes vary substantially in size, with the *Eutreptiella gymnastica* genome, at 67 kbp, less than half the size of that of *Euglena gracilis*, containing one more coding sequences (CDS) [27]. The rDNA operon does show variations in inverted and tandem repeats between different species, with the plastid genome complexity possibly related to the expansion of group II introns [29].

Transcriptomes

Euglena gracilis

Prior to the availability of the *E. gracilis* genome, various groups undertook *de novo* transcriptome sequencing [14,18,30]. These various studies confirmed that protein expression does not mirror RNA abundance and is instead controlled post-transcriptionally [31]. Analysis of the *E. gracilis* transcriptome supports the 'shopping bag' model of plastid acquisition, demonstrating the incorporation of genes from diverse photosynthetic lineages [14]. This model suggests that *E. gracilis* has integrated genes from various algal sources through multiple endosymbiotic events [32], forming a versatile metabolic framework that supports its survival in diverse ecological niches. Furthermore, *E. gracilis* exhibits sophisticated post-transcriptional regulation, similar to the mechanisms observed in kinetoplastids [18], which plays an important role in protein expression control and is essential for environmental adaptability.

Euglena mutabilis

E. mutabilis is often found in acid mine drainage and is able to resist low pH and high levels of heavy metals. Transcriptomic analyses have revealed that *E. mutabilis* possesses advanced

systems to manage environmental challenges. These include proton pumps and membrane transport proteins that regulate cellular pH, alongside mechanisms for sulfur metabolism and metal detoxification [33]. This extensive suite of stress response genes, including those for DNA repair and detoxification, underscores the ability of *E. mutabilis* to maintain cellular integrity under extreme environmental pressures [34]. The circadian rhythm of *E. mutabilis* was found to control gene expression for photosystem II repair and oxidative stress management [34]. Pulse-amplitude-modulation (PAM) fluorometry indicated rapid photosynthetic adjustments to fluctuating light conditions [35], a capability less pronounced in other euglenids.

Euglena longa

E. longa provides a unique perspective on euglenid evolution due to its non-photosynthetic plastid [36], which diverges significantly from that of its photosynthetic relatives such as *Euglena gracilis* [37]. *E. longa* lacks the genomic sequences typically associated with photosynthetic processes, emphasising its adaptation to a heterotrophic lifestyle [28]. Illumina HiSeq sequencing technology was employed to comprehensively map the transcriptome of *E. longa* [36], enabling a detailed elucidation of its genomic architecture.

Transcriptomic analysis indicates a significant absence of genes encoding photosystems and light-harvesting complexes [36], which are essential components in photosynthetic euglenids. Even without photosynthetic functions, *E. longa* still has plastid-localised proteins with N terminal bipartite signals, indicating a preserved mechanism for importing proteins into plastids [38]. There is also a novel plastid-targeted homologue of the bacterial transcription termination factor Rho [39], acquired through horizontal gene transfer. This indicates an innovative adaptation in transcriptional regulation within the plastid of *E. longa*, a feature unprecedented in other euglenids [40].

Rapaza viridis

R. viridis exhibits a mixotrophic lifestyle, combining both phototrophic and phagotrophic nutritional modes [41]. It utilises kleptoplasts (kP), unique photosynthetic organelles transiently acquired from the green alga *Tetraselmis* sp., enabling photosynthesis without permanent endosymbiosis [42]. *R. viridis* differs from its close relatives in the Euglenophyceae as it does not possess canonical plastids [43–45].

Transcriptomic analysis of *R. viridis* identified 276 sequences encoding putative plastid-targeting proteins and 35 sequences identified as kleptoplast transporters [42]. These genes have diverse origins, reflecting extensive horizontal gene transfer from various algal species [46]. This underscores the complex evolutionary processes that have shaped the *R. viridis* genome. Many of these plastid-targeted proteins are homologous with those found in *E. gracilis*, indicating a shared evolutionary heritage and suggesting that early plastid acquisition involved kleptoplasty before the establishment of permanent plastids [47].

Furthermore, the transcriptome identified kleptoplast-targeted proteins with euglenid-type bipartite targeting signals [47,48]. This indicates a system for photosynthetic integration despite its phagotrophic ancestry. This genetic complexity supports its dual nutritional strategies and offers significant insights into plastid evolution and euglenid diversity.

Eutreptiella gymnastica

E. gymnastica is a marine photosynthetic euglenid and plays an important role in primary production and nutrient cycling. The transcriptome of *E. gymnastica* was sequenced as part of the Marine Microbial Eukaryote Transcriptome Sequencing Project [49], though further study has been limited.

Microbiomes

Microbes such as euglenids have evolved in diverse environments, often having to deal with different challenges and competition between species. This has often given rise to symbiotic relationships between organisms that co-occur in a bid to increase their survival chances. Symbiosis is well known to cause genome reduction and streamlining as well as morphological changes, thus it is difficult to assess the evolution of any euglenids without also considering the symbionts that they form productive relationships with.

Euglenid microbiomes

The term phycosphere has been used to describe phytoplankton–bacteria relationships in which metabolites and infochemicals are used to facilitate interactions between an alga and other organisms inhabiting the same environment [50] which may enable organisms to gain competitive advantages. A range of euglenids have been found to co-occur with specific microorganisms, without investigation of the roles each partner plays.

Functional bacterial communities have been associated with *Euglena sanguinea* and its toxic blooms, with 16S rDNA amplicon sequencing revealing a preponderance of *Deinococcus–Thermus* bacteria [51]. A bacterium was isolated from the cell surface and cytostome of *E. gracilis*, and whole-genome sequencing revealed it to be a *Paenibacillus* sp., although no further work has been carried out to identify the nature of their interactions [52].

Increased stress tolerance. Many *Euglena* species are capable of thriving in harsh environments, such as acidic mine drainage, which often contain high concentrations of heavy metals. *E. mutabilis* has been found inhabiting acid ponds in the Yukon, existing in mutualistic relationships with *Cryptococcus* spp. that appeared to increase the acid and heavy metal tolerance of the alga [53]. Full-length amplicon sequencing and targeted Sanger sequencing identified the fungus *Talaromyces* sp. and the bacterium *Acidiphilium acidophilum*, which combine to form a fungal, algal, and bacterial (FAB) consortium with *E. mutabilis* that enhances its cadmium tolerance [54].

Increased growth and production. Various microorganisms have been investigated in the hopes that co-cultivation with *E. gracilis* will improve the algal growth and production of high-value compounds that the alga synthesises. The bacterial species *Enterobacter* sp. CA3 and *Emticicia* sp. CN5 promoted both algal growth and production of the valuable carbohydrate paramylon in *E. gracilis* [55]. Additionally, *Emticicia* sp. EG3 was found to promote biomass and lipid yield during co-cultivation [56]. Co-culturing *E. gracilis* with species like these could provide a low-tech solution to improve yields of high-value compounds. The fungus *Cladosporium westerdijkiae* and bacterial species *Lysinibacillus boronitolerans* and *Pseudobacillus badius* have been shown to provide *E. gracilis* with vitamins B1 and B12 that are required for its growth and survival [57]. This reduces the cost of vitamins required in the media during large-scale cultivation which could improve the economic viability of industrial production of compounds derived from *E. gracilis*.

Microbial relationships of related organisms. *R. viridis,* a euglenoid species, has lost its *Pyramimonas*-related plastid but now fosters transient endosymbiotic plastids from a species of prey alga (*Tetraselmis* sp.). The alga is phagocytosed by *R. viridis*; however, the chloroplast is retained [41]. Investigating this relationship may help to inform how endosymbiotic relationships in euglenids and other host organisms have evolved.

Two relatives of the Euglenoida from the class Symbiontida, *Calkinsia aureus* and *Bihospites bacati*, form symbiotic relationships with epibiontid bacteria [58]. These bacteria are related to sulfur and sulfide-oxidising epsilon proteobacteria and are thought to support the survival of

the unicellular flagellates in the low-oxygen and high-sulfur environments where they are found. Epibiontid relationships may also play a part in the euglenoid phycospheres and could be an area for future study.

How diplonemid and kinetoplastid genomic advances have affected our understanding of euglenids

Euglenids and their sister clades of kinetoplastids and diplonemids in the phylum of Euglenozoa share many genomic oddities, including the presence of base J and trans-splicing of the spliced leader sequence, allowing informative comparisons with these more thoroughly studied organisms.

Kinetoplastids include well-known mammalian parasites *Trypanosoma* spp. and *Leishmania* spp., as well as more basal, free-living representatives such as *Bodo* spp. and *Neobodo* spp. Kinetoplastid nuclear genomes, by contrast with euglenids, are generally small and streamlined, almost entirely lacking introns. The relative ease of sequencing kinetoplastid genomes has resolved certain genetic trends when compared with available datasets from euglenids. For instance, the extensive genome reduction in kinetoplastids, presumed to represent an adaptive response to the parasitic lifestyles, was in fact steadily occurring in free-living ancestors of this clade as well [59].

Kinetoplastids are notable for their radically divergent kinetochore complex, used to mediate chromosome segregation during cell division. Only a handful of free-living members retain a single conventional kinetochore component (cenH3), and instead employ over two dozen kinetoplastid kinetochore proteins (KKTs) and kinetoplastid kinetochore-interacting proteins (KKIPs) [60]. Euglenid kinetochores in comparison represent a mosaic composition, possessing up to four of 11 conventional components as well as up to three KKTs and one KKIP. It is likely that several uncharacterised proteins, specific to euglenids, are additionally incorporated to complete this complex.

Diplonemids represent primarily marine protists, constituting abundant and diverse members of the ocean's plankton communities [61]. While entirely lacking photosynthetic organelles, their heterotrophic lifestyles are considerably less resolved than the varied strategies documented across euglenids. Recently generated datasets for diplonemids suggest even larger protein-coding capacities (~37 000–52 000) than those of euglenids (~34 000–37 000), arising from enhanced gene family gains and expansions such as for peptidases and sphingolipid and amino acid metabolism, following the divergence of this group [59,60]. Diplonemids similarly possess an abundance of non-canonical introns, lacking typical GT-AG splice-site motifs [62].

Like euglenids, the presence of many transposons and long, repetitive genomic regions, make long read sequencing technologies essential for chromosome-level assemblies for diplonemids. Model organism *Paradiplonema papillatum* (formerly of the genus *Diplonema)* represents the first comprehensive nuclear assembly [62], although other species are currently being processed through the Aquatic Symbiosis Genome Project. The ~280-Mb genome of *P. papillatum* encodes approximately 37 000 genes, and appears to be in a haploid state, but with the potential to form diploid zygotes.

Kinetoplastids and diplonemids are together subclassified as glycomonada due to the shared presence of modified glycolytic peroxisomes, and additionally both possess complex mitochondrial genomes. By contrast, the simplified and small mitochondrial genomes of euglenids, usually present as linear chromosomes, represent a dramatic point of divergence from the baroque glycomonad accumulations of interlinked mitochondrial DNA circles, which are further

complicated by obligatory editing to render these mitochondrial transcripts functional [63,64]. In turn, the mitochondrial proteome of *E. gracilis* revealed a base selection of conserved nucleic acid processing components, including RNA helicases, exonucleases, and RNA-binding proteins from which mitochondrial RNA editing was likely developed independently across both kinetoplastid and diplonemid lineages [24].

Concluding remarks

Progress in euglenid genetics has started to accelerate with the new sequencing technologies, but many questions remain (see Outstanding questions). Although much work has focused on *E. gracilis*, the euglenids are a much broader class and have a great deal of unexplored potential. Fundamental questions remain in the model *E. gracilis*, but a much broader sequencing effort is required to explore the diversity of euglenids, probing class-specific features and diversity within the class. A comprehensive program of targeted mutagenesis, complementation, and heterologous expression is needed to understand their cellular biology, particularly focusing on the lineage-specific genes. This requires a reliable genetic manipulation toolkit, including clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein (Cas), heterologous expression, and protein-tagging tools. This will also allow the development of these organisms through engineering biology for exploitation as biotechnology platforms for the production of high-value products, nutraceuticals, and biofuels.

Whilst there has been isolated research by individual groups into various aspects of *Euglena* biology and genomics, there has recently been a push to integrate the international efforts, culminating in the formation of the Euglena International Network (EIN) [8]. This network focuses on bringing the community together through the biennial International Congress on Euglenoids, and a science committee, to collaborate on research goals and share data. This has included the formation of the Euglenoid Genomes Project (EGP) with the goal of producing 1000 reference genomes of euglenids. Initially this has contributed to prioritising species and integrating with the Darwin Tree of Life Project [65] and the Earth BioGenome [66]. The other goal is to develop tools and techniques for genetic interrogation and modification based on the new genomic data. Whilst some tools are available for genetic engineering of *E. gracilis* [67], these fall short of the ideals for synthetic biology and have not even begun development in other euglenids. These tools would allow forward and reverse genetics techniques to interrogate the fundamental biology of different euglenids and facilitate their use for commercial application and production of high-value compounds and therapeutics.

Outstanding questions

Do euglenids undergo sexual reproduction and, if so, under what circumstances?

How are genes arranged on the chromosomes, and is there gene clustering?

How are gene expression and protein levels controlled?

Where is the hypermodified base J in the genome? What is its role, and how is it regulated?

What is the role of the >50% of protein-coding genes with no predicted function?

How many chromosomes do different species contain, and what is the ploidy?

What genes are found across euglenids, but not in other eukaryotic classes?

What genes are found in different lifestyles (such as phototrophic versus predatory) and ecological niches (such as marine or low pH)?

Do other euglenids have a fragmented mitochondrial genome structure? How are these fragments replicated and assembled?

Acknowledgments

O.F. is supported by the Nottingham BBSRC Doctoral Training Partnership (BB/T008369/1/2748026).

Declaration of interests

The authors declare no conflicts of interest.

References

1. Kostygov, A.Y. *et al.* (2021) Euglenozoa: taxonomy, diversity and ecology, symbioses and viruses. *Open Biol.* 11, 200407
2. O'Neill, E. *et al.* (2017) Exploring the glycans of *Euglena gracilis*. *Biology* 6, 45
3. O'Neill, E.C. (2023) Glycosylated proteins in the protozoan alga *Euglena gracilis*: a proteomic approach. *FEMS Microbiol. Lett.* 370, fnac120
4. Jackson, C. *et al.* (2018) Plastid phylogenomics with broad taxon sampling further elucidates the distinct evolutionary origins and timing of secondary green plastids. *Sci. Rep.* 8, 1523
5. Zakryś, B. *et al.* (2017) Evolutionary origin of *Euglena*. In *Euglena: Biochemistry, Cell and Molecular Biology* (Schwartzbach, S.D. and Shigeoka, S., eds), pp. 3–17, Springer
6. Schwartzbach, S.D. and Shigeoka, S. (2017) *Euglena: Biochemistry, Cell and Molecular Biology*, Springer
7. Gissibl, A. *et al.* (2019) Bioproducts from *Euglena gracilis*: synthesis and applications. *Front. Bioeng. Biotechnol.* 7, 108
8. Ebenezer, T.E. *et al.* (2022) Euglena International Network (EIN): driving euglenoid biotechnology for the benefit of a challenged world. *Biol. Open* 11, bio059561

9. Bedard, S. *et al.* (2024) The biomolecules of *Euglena gracilis*: harnessing biology for natural solutions to future problems. *Protist* 175, 126044
10. Kim, S. *et al.* (2023) Biofuel production from *Euglena*: current status and techno-economic perspectives. *Bioresour. Technol.* 371, 128582
11. Aldholmi, M. *et al.* (2022) Euglenatides, potent antiproliferative cyclic peptides isolated from the freshwater photosynthetic microalga *Euglena gracilis*. *Angew. Chem. Int. Ed.* 61, e202203175
12. Assis, L.H.d.C. *et al.* (2023) Behind base J: the roles of JBP1 and JBP2 on Trypanosomatids. *Pathogens* 12, 467
13. Dooijes, D. *et al.* (2000) Base J originally found in Kinetoplastida is also a minor constituent of nuclear DNA of *Euglena gracilis*. *Nucleic Acids Res.* 28, 3017–3021
14. Ebenezer, T.E. *et al.* (2019) Transcriptome, proteome and draft genome of *Euglena gracilis*. *BMC Biol.* 17, 11
15. Chen, Z. *et al.* (2024) A chromosome-level genome assembly for the paramylon-producing microalga *Euglena gracilis*. *Sci. Data* 11, 780
16. Reis-Cunha, J.L. *et al.* (2024) Ancestral aneuploidy and stable chromosomal duplication resulting in differential genome structure and gene expression control in trypanosomatid parasites. *Genome Res.* 34, 441–453
17. Falchuk, K. *et al.* (1975) Role of zinc in cell division of *Euglena gracilis*. *J. Cell Sci.* 17, 57–78
18. O'Neill, E.C. *et al.* (2015) The transcriptome of *Euglena gracilis* reveals unexpected metabolic capabilities for carbohydrate and natural product biochemistry. *Mol. BioSyst.* 11, 2808–2820
19. Milanowski, R. *et al.* (2014) Distribution of conventional and nonconventional introns in tubulin (α and β) genes of euglenids. *Mol. Biol. Evol.* 31, 584–593
20. Gumińska, N. *et al.* (2021) A new type of circular RNA derived from nonconventional introns in nuclear genes of euglenids. *J. Mol. Biol.* 433, 166758
21. De Gaudenzi, J.G. *et al.* (2011) Gene expression regulation in trypanosomatids. *Essays Biochem.* 51, 31–46
22. Soukal, P. *et al.* (2021) Heterotrophic euglenid *Rhabdomonas costata* resembles its phototrophic relatives in many aspects of molecular and cell biology. *Sci. Rep.* 11, 13070
23. Dobáková, E. *et al.* (2015) Unexpectedly streamlined mitochondrial genome of the euglenozoan *Euglena gracilis*. *Genome Biol. Evol.* 7, 3358–3367
24. Hammond, M.J. *et al.* (2020) A uniquely complex mitochondrial proteome from *Euglena gracilis*. *Mol. Biol. Evol.* 37, 2173–2191
25. Manning, J.E. *et al.* (1971) Circular chloroplast DNA from *Euglena gracilis*. *Proc. Natl. Acad. Sci.* 68, 1169–1173
26. Hallick, R.B. *et al.* (1993) Complete sequence of *Euglena gracilis* chloroplast DNA. *Nucleic Acids Res.* 21, 3537–3544
27. Hrdá, Š. *et al.* (2012) The plastid genome of *Eutreptiella* provides a window into the process of secondary endosymbiosis of plastid in euglenids. *PLoS One* 7, e33746
28. Gockel, G. and Hachtel, W. (2000) Complete gene map of the plastid genome of the nonphotosynthetic euglenoid flagellate *Astasia longa*. *Protist* 151, 347–351
29. Maciszewski, K. *et al.* (2022) Challenging the importance of plastid genome structure conservation: new insights from Euglenophytes. *Mol. Biol. Evol.* 39, msac255
30. Yoshida, Y. *et al.* (2016) *De novo* assembly and comparative transcriptome analysis of *Euglena gracilis* in response to anaerobic conditions. *BMC Genomics* 17, 182
31. Cordoba, J. *et al.* (2021) *De novo* transcriptome meta-assembly of the mixotrophic freshwater microalga *Euglena gracilis*. *Genes (Basel)* 12, 842
32. O'Neill, E.C. *et al.* (2015) Euglena in time: evolution, control of central metabolic processes and multi-domain proteins in carbohydrate and natural product biochemistry. *Perspect. Sci.* 6, 84–93
33. Perry, J.J. *et al.* (2010) The structural biochemistry of the superoxide dismutases. *Biochim. Biophys. Acta* 1804, 245–262
34. Puente-Sánchez, F. *et al.* (2016) Solar radiation stress in natural acidophilic biofilms of *Euglena mutabilis* revealed by metatranscriptomics and PAM fluorometry. *Protist* 167, 67–81
35. Tyystjärvi, E. (2013) Photoinhibition of photosystem II. *Int. Rev. Cell Mol. Biol.* 300, 243–303
36. Záhonová, K. *et al.* (2018) Peculiar features of the plastids of the colourless alga *Euglena longa* and photosynthetic euglenophytes unveiled by transcriptome analyses. *Sci. Rep.* 8, 17012
37. Hadariová, L. *et al.* (2017) An intact plastid genome is essential for the survival of colorless *Euglena longa* but not *Euglena gracilis*. *Curr. Genet.* 63, 331–341
38. Záhonová, K. *et al.* (2016) RuBisCO in non-photosynthetic alga *Euglena longa*: divergent features, transcriptomic analysis and regulation of complex formation. *PLoS One* 11, e0158790
39. Kriner, M.A. *et al.* (2016) Learning from the leaders: gene regulation by the transcription termination factor Rho. *Trends Biochem. Sci.* 41, 690–699
40. Leander, B.S. *et al.* (2017) Euglenida. In *Handbook of the Protists* (Archibald, J.M. *et al.*, eds), pp. 1–42, Springer
41. Yamaguchi, A. *et al.* (2012) Morphostasis in a novel eukaryote illuminates the evolutionary transition from phagotrophy to phototrophy: description of *Rapaza viridis* n. gen. et sp. (Euglenozoa, Euglenida). *BMC Evol. Biol.* 12, 29
42. Karnkowska, A. *et al.* (2023) Euglenozoan kleptoplasty illuminates the early evolution of photoendosymbiosis. *Proc. Natl. Acad. Sci. U. S. A.* 120, e2220100120
43. Maeda, T. *et al.* (2021) Chloroplast acquisition without the gene transfer in kleptoplastic sea slugs, *Plakobranchus ocellatus*. *eLife* 10, e60176
44. Hehenberger, E. *et al.* (2019) A kleptoplastidic dinoflagellate and the tipping point between transient and fully integrated plastid endosymbiosis. *Proc. Natl. Acad. Sci. U. S. A.* 116, 17934–17942
45. Onuma, R. *et al.* (2020) Changes in the transcriptome, ploidy, and optimal light intensity of a cryptomonad upon integration into a kleptoplastic dinoflagellate. *ISME J.* 14, 2407–2423
46. Howe, C.J. *et al.* (2008) The origin of plastids. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 363, 2675–2685
47. Novák Vanclová, A.M.G. *et al.* (2020) Metabolic quirks and the colourful history of the *Euglena gracilis* secondary plastid. *New Phytol.* 225, 1578–1592
48. Durnford, D.G. and Gray, M.W. (2006) Analysis of *Euglena gracilis* plastid-targeted proteins reveals different classes of transit sequences. *Eukaryot. Cell* 5, 2079–2091
49. Keeling, P.J. *et al.* (2014) The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biol.* 12, e1001889
50. Seymour, J.R. *et al.* (2017) Zooming in on the phycosphere: the ecological interface for phytoplankton–bacteria relationships. *Nat. Microbiol.* 2, 1–12
51. Zheng, X. *et al.* (2020) Size-fractioned aggregates within phycosphere define functional bacterial communities related to *Microcystis aeruginosa* and *Euglena sanguinea* blooms. *Aquat. Ecol.* 54, 609–623
52. Shtratnikova, V.Y. *et al.* (2020) Complete genome assembly data of *Paenibacillus* sp. RUD330, a hypothetical symbiont of *Euglena gracilis*. *Data Brief* 32, 106070
53. Nakatsu, C. and Hutchinson, T.C. (1988) Extreme metal and acid tolerance of *Euglena mutabilis* and an associated yeast from Smoking Hills, Northwest Territories, and their apparent mutualism. *Microb. Ecol.* 16, 213–231
54. Kaszecki, E. *et al.* (2024) *Euglena mutabilis* exists in a FAB consortium with microbes that enhance cadmium tolerance. *Int. Microbiol.* 27, 1249–1268
55. Rubiyatno *et al.* (2021) Isolation and characterization of *Euglena gracilis*-associated bacteria, *Enterobacter* sp. CA3 and *Emticicia* sp. CN5, capable of promoting the growth and paramylon production of *E. gracilis* under mixotrophic cultivation. *Microorganisms* 9, 1496
56. Toyama, T. *et al.* (2019) Enhanced production of biomass and lipids by *Euglena gracilis* via co-culturing with a microalga growth-promoting bacterium, *Emticicia* sp. EG3. *Biotechnol. Biofuels* 12, 1–12
57. Lukáčová, A. *et al.* (2022) *Euglena gracilis* can grow in the mixed culture containing *Cladosporium westerdijkiae, Lysinibacillus boronitolerans* and *Pseudobacillus badius* without the addition of vitamins B1 and B12. *J. Biotechnol.* 351, 50–59
58. Edgcomb, V.P. *et al.* (2010) Identity of epibiotic bacteria on symbiontid euglenozoans in O_2-depleted marine sediments:

evidence for symbiont and host co-evolution. *ISME J.* 5, 231–243
59. Butenko, A. *et al.* (2021) Reductionist pathways for parasitism in euglenozoans? Expanded datasets provide new insights. *Trends Parasitol.* 37, 100–116
60. Butenko, A. *et al.* (2020) Evolution of metabolic capabilities and molecular features of diplonemids, kinetoplastids, and euglenids. *BMC Biol.* 18, 23
61. Tashyreva, D. *et al.* (2022) Diplonemids – a review on 'new' flagellates on the oceanic block. *Protist* 173, 125868
62. Valach, M. *et al.* (2023) Recent expansion of metabolic versatility in *Diplonema papillatum*, the model species of a highly speciose group of marine eukaryotes. *BMC Biol.* 21, 99
63. Read, L.K. *et al.* (2016) Trypanosome RNA editing: the complexity of getting U in and taking U out. *Wiley Interdiscip. Rev. RNA* 7, 33–51
64. Kaur, B. *et al.* (2020) Gene fragmentation and RNA editing without borders: eccentric mitochondrial genomes of diplonemids. *Nucleic Acids Res.* 48, 2694–2708
65. The Darwin Tree of Life Project Consortium (2022) Sequence locally, think globally: The Darwin Tree of Life Project. *Proc. Natl. Acad. Sci.* 119, e2115642118
66. Lewin, H.A. *et al.* (2018) Earth BioGenome Project: sequencing life for the future of life. *Proc. Natl. Acad. Sci. U. S. A.* 115, 4325–4333
67. Chen, Z. *et al.* (2022) A synthetic biology perspective on the bioengineering tools for an industrial microalga: *Euglena gracilis*. *Front. Bioeng. Biotechnol.* 10, 882391