Analysis of Unknown Protists and Cryptic Signatures of Pseudopod Formation Genes with Regards to Morphological Switches

Tyler Lee Johnson

A thesis submitted to the University of Huddersfield in partial fulfilment of the requirements for the degree of MSc by Research

Acknowledgements

I would like to extend my gratitude towards my family and friends for being understanding and supporting my decision to challenge myself academically. To my partner, who has supported and encouraged me through times of stress to see me achieve my goals.

I would also like to take the time to thank all my supervisors. Dr. Martin Carr for his knowledge and for his part with PCR protocols that enabled me to carry out my research. Prof. Michael Ginger, for challenging me with my writing to ensure I was concise and accurate throughout, motivating me along the way. But most importantly, for believing that I could do this during bleak times. Lastly, Dr. Jane Harmer, for her unwavering support and criticisms that have left an impact on me that will undoubtably benefit me in the near future. Your cheery personality and extensive knowledge have fundamentally got me here.

To you all, I do not think you can fathom how much this means to me. So, thank you all so much.

Abstract

Protists represent the largest proportion of eukaryotic taxa that dominate (in terms of numbers and diversity of species) freshwater and marine environments. These aquatic protists include; the free-living kinetoplastids, which are sometimes outshined by their parasitic-sisters, the trypanosomatids and *Leishmanias.*

Parasitic taxa that are capable of changing morphologies or cellular differentiation do so in various ways. Such changes may be dependent upon a shift in the host internal environments, resulting in repurposing appendages or proteins through RNA editing mechanisms which allow the parasite to change morphologies through various life-cycle stages; or by moonlighting proteins, whereby proteins characterized by their primary functions can also have a hidden secondary function (which can relate to the virulence of parasitic taxa) *Trichomonas vaginalis* (*Trichomonadidae*), *Salpingoeca rosetta* (*Salpingoecidae*), and *Naegleria gruberi (Vahlkampfiidae)* are well documented protists (amongst others) that have been observed to change morphologies. These species are largely removed from the kinetoplastids (Discoba) and each other, Metamonada, Opisthokonta, and Percolozoa respectively, yet display the same capabilities for morphological change, albeit under differing conditions. This display of function would suggest that this mechanism (either for parasitic virulence or as a 'fight-or-flight' response to changes in the environment) is likely to have evolved from a common ancestor linking all four lineages and not evolved independently after divergence.

Aims

In this MSc, I aim to highlight the defining characteristics, with the aid of confocal microscopy, in conjunction with BLASTn bioinformatics to propose the identity of an unknown kinetoplastid dubbed 'Dark Stuff' collected from Embo beach in the highlands of Scotland. I also aim to examine the candidate identities of protists collected around the lake at University of York and to provide insight into the hypothesis of cryptic signatures related to pseudopod formation genes in a range of protists. Here BLASTp analyses and protein sequence alignments will be used to identify protists that may show capabilities to switch morphologies.

Contents

Abbreviations

BLAST	Basic Local Alignment Search Tool
BLASTn	Nucleotide Basic Local Alignment Search Tool
BLASTp	Protein-protein Basic Local Alignment Search Tool
ToL	Tree of Life
eToL	Eukaryotic Tree of Life
DNA	Deoxyribonucleic Acid
rDNA	Ribosomal Deoxyribonucleic Acid
kDNA	Kinetoplast Deoxyribonucleic Acid
mtDNA	Mitochondrial Deoxyribonucleic Acid
cDNA	Complementary Deoxyribonucleic Acid
eDNA	Environmental Deoxyribonucleic Acid
gDNA	Genomic Deoxyribonucleic Acid
RNA	Ribonucleic Acid
rRNA	Ribosomal Ribonucleic Acid
mRNA	Messenger Ribonucleic Acid
eRNA	Environmental Ribonucleic Acid
AA(s)	Amino Acid(s)
NGS	Next Generation Sequencing
S	Svedberg Units
ITS	Internal Transcribed Spacer
SSU	Small Ribonucleic Acid Sub-unit
mtSSU	Mitochondrial Small Ribonucleic Acid Sub-unit
LSU	Large Ribosomal Subunit
mtLSU	Mitochondrial Large Ribosomal Sub-unit
bp	Base Pair(s)
V	Variable Regions
PCR	Polymer Chain Reaction
OUT(s)	Operational Taxonomic Unit(s)
CARD	Catalysed Reporter Deposition
FISH	Fluorescent *in situ* Hybridization
DAPI	4',6-diamidino-2-phenylindole
kb	Kilo-base(s)
Kbp	Kilo Base Pair(s)
PFR	Paraflagellar Rod
Ca. Pa. novymonadis	*Candidatus Pandoraea novymonadis*
Ca. kinetoplastibacterium	*Candidatus kinetoplastibacterium*
KP	Kinetoplastid-Posterior
NP	Nucleus-Posterior
BB	Basal Body
FP	Flagellar Pocket
Ax	Axoneme
DSPD	Deep-Sea Pelagic Diplonemids
Mbp	Mega Base-Pair(s)
COX	Cytochrome c oxidase
SIP	Stable Isotope Probing
MDS	Multi-Dimensional Scaling

SAGE	Serial Analysis of Gene Expression
CAGE	Cap Analysis of Gene Expression
RNA Seq.	Ribonucleic Acid Sequencing
EST	Expressed Sequence Tag
Tb	Tera Bases
EM	Electron Microscopy
IFT	Intraflagellar Transport
BBSome	Bardet-Biedl Syndrome
ADF	Adenosine Diphosphate
Arp	Actin-related Proteins
WASP	Wiskott-Aldrich Syndrome Protein
SCAR/WAVE	Sequence Characterized Amplified Region
ATP	Adenosine Triphosphate
NPF(s)	Nucleation-Promotion Factor(s)
Hsp(s)	Heat Shock Protein(s)
Tris	Hydroxymethyl aminomethane
NaOH	Sodium Hydroxide
EDTA	Ethylenediaminetetraacetic Acid
SDS	Sodium Dodecyl Sulphate
RNAse(s)	Ribonuclease(s)
EB	Elution Buffer
dNTP(s)	Deoxynucleotide Triphosphate(s)
TAE	Tris-Acetate-Ethylenediaminetetraacetic Acid
PBS	Phosphate Buffered Saline
UV	Ultraviolet
LB	Lysogeny Broth
DTT	DL-Dithiothreitol
EtOH	Ethanol
MIC	Micronucleus
MAC	Macronucleus

Chapter 1

Introduction

The complexity of Earth's biodiversity is not yet fully understood either with respect to the catalogue of life currently on planet Earth or the community relationships that define different ecosystems or environments. New taxa are still discovered frequently and these newly discovered species often prompt discussions as to where they belong on the Tree of Life (ToL). This is perhaps particularly so in the case of protists and other microbial eukaryotes; collectively, the breadth of biodiversity across microbial eukaryotes appears to be far from defined. There are two types of organism: the prokaryotes (*pro-,* before, *-karyote*, nucleus) and the eukaryotes (*eu-,* true). Often there is still uncertainty as to where different organisms' taxa sit on the ToL. When looking at the relationships or phylogeny between different eukaryotes, including the vast breadth of microbial eukaryotes, I will refer to this as the eukaryotic ToL (or eToL). In my MSc research I set out to achieve two broad objectives: to characterize novel protists taxa from a series of environmental samples and to look for possible cryptic signatures of different modes of cell motility in evolutionarily little-studied eukaryotes.

1.1 Current views on the Eukaryotic Tree of Life (eToL)

The eToL is based around five to eight (subject to different authors views and interpretations) major groupings. These groups were reviewed, and redefined recently in Burki, et al. (2020) and Lax, et al. (2018). Eukaryotes include multi-cellular animals, plants, fungi, and protists; much of the uncertainty as to what the eToL looks like stems from various evidence-views of where different microbial eukaryotes sit within the eToL (Burki, et al., 2020). This uncertainty is likely to continue whilst further new protist and fungal lineages are discovered. Innovations in technology have resulted in the observation of new taxa by microscopy (Tables 1-2). Yet, major advances in deoxyribonucleic acid (DNA) sequencing and the scale of what can be sequenced means that the numerous eukaryotic taxa have a known DNA signature or footprint but are yet to be isolated into culture or their morphologies observed directly by microscopy (Albuquerque, et al., 2009). Collectively, the view of the eToL has changed much during the last two decades and is very different to the original domain-based ToLs summarised in Figure 1.
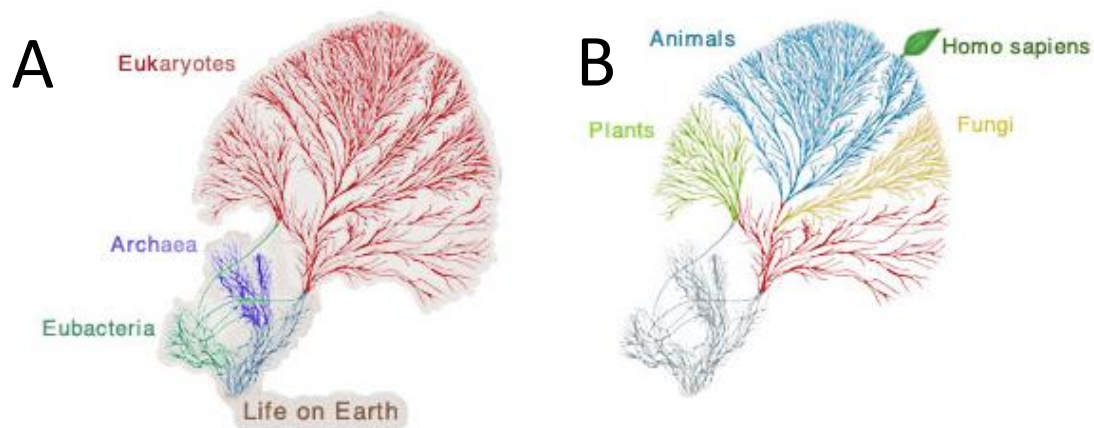
Figure 1: Hierarchy subgroup depiction of life on Earth. A) Division of all life on Earth from a common ancestor joining prokaryotes and eukaryotes, with further evolutionary divides into Archaea and Eubacteria (prokaryotes) and all other eukaryotes. B) Kingdom branching of some of the major groups observed within the tree of life. Reproduced from the Open Access source, Maddison and Schulz 2007, 'Tree of Life Web Project'.

Burki, et al., (2020) place eukaryotes into seven major groups, plus a number of orphan taxa for which nearest relations are uncertain (Figure 2). The seven major groups are Amorphea, Archaeplastida, Cryptista, Haptista, Excavata (a group which may not be monophyletic—*i.e.,* may not all share the same common, most recent ancestor), TSAR (acronym for Telonemia, Rhizaria, Alveolata, and Stramenopila), and CRuMs (acronym for *Collodictyonidae*—also known as *Diphylleidae*, Rigifilida, and *Mantamonas*) (Table 1). Several of the recently discovered orphan lineages are summarised in Table 2.
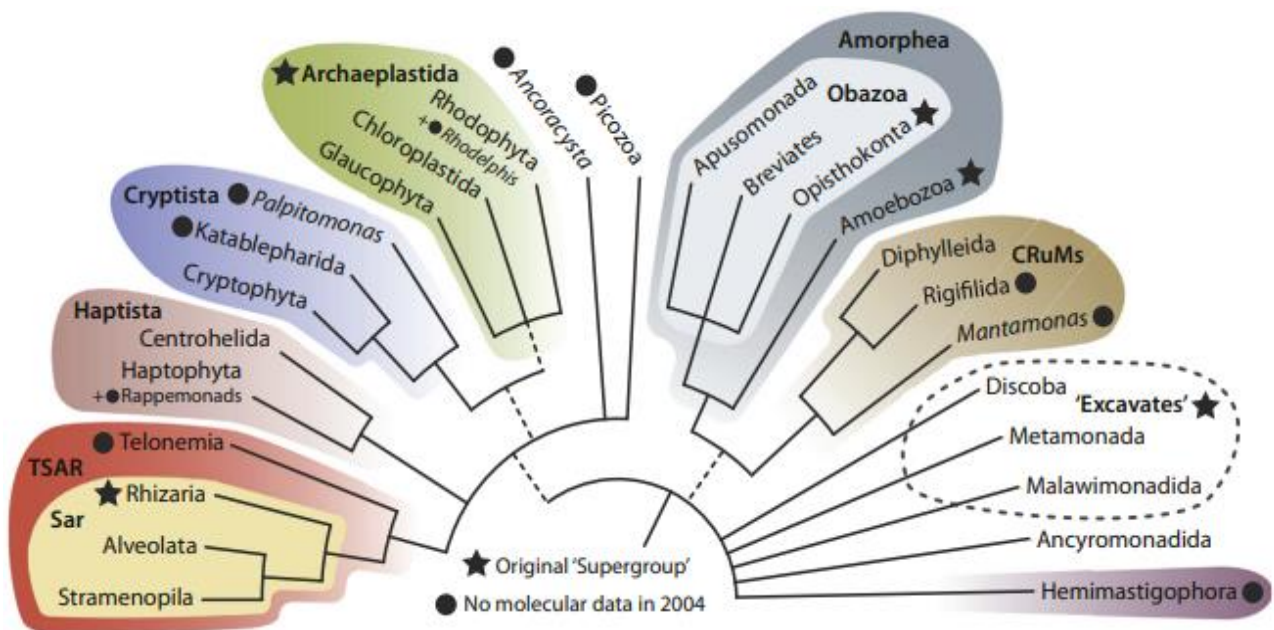
Figure 2: Summary of recent phylogenomic studies on several species of microbes. Colour groupings relate to the supergroups taxa belong to. Broken lines reflect uncertainties within groups. Filled stars denote original supergroups in earlier versions of the model which have since either disappeared or subsumed within new taxa. Filled circles show a lack of molecular data when the supergroup models emerged (due to them being undiscovered at the time). *Rappemonads* are placed based only on the available ribosomal ribonucleic acid (rRNA) data. Reproduced permissions from Fabian Burki, 2020 'The New Tree of Eukaryotes', UK, Elsevier LTD.

Supergroups	Clades	Groups
Amorphea	Amoebozoa	Lobosa
	Obazoa	Opisthokonta
		Breviata
Archaeplastida	Rhodophyta	Florideophyceae
	Picozoa	Picomonadida
	Rhodelphidia	*Rhodelphis*
Cryptista	*Katablepharidae*	*Roombia*
		Katablepharis
	Cryptophyceae	Tetragonidiales
Haptista	Centrohelids	*Raphidiophryidae*
	Haptophytes	Pavlovophyceae
Excavata	Discoba	Euglenozoa
		Heterolobosea
	Loukozoa	Metamonada
TSAR	Telonemia	*Telonema*
	Stramenopile	Opalines
	Alveolata	Ciliophora
	Rhizaria	Cercozoa
CRuMs	*Diphylleidae (Collodictyonidae)*	*Diphylleia*
	Rigifilida	*Rigifila*
	Mantamonadida	*Mantamonas*

Table 1: Breakdown of revised supergroups into clades and examples of groups within each clade.

Amended with permissions from Fabian Burki, 2020 The New Tree of Eukaryotes, UK, Elsevier LTD.

Group	Year Identified	Original Description	Molecular Data Source	Phylogenomic Confirmation
Breviates	2004	1893	Cultivation	2013
Katablepharids	2005	1939	Cultivation	2012
Telonemia	2006	1913	Cultivation	2009
Picozoa	2007	2007	Environmental PCR + FISH	2012
Rigifilids	2008	2001	Cultivation	2018
Palpitomonas	2010	2010	Cultivation	2014
Tsukubamonas	2011	2011	Cultivation	2014
Mantamonas	2011	2011	Cultivation	2014
Rappemonads	2011	2011	Environmental PCR + FISH	N/A
Microheliella	2012	2012	Cultivation	2015
Ancoracysta	2017	2009	Cultivation	2017
Anaeramoeba	2017	2017	Cultivation	N/A
Hemimastigophora	2018	1893	Single-cell Isolation	2018
Rhodelphis	2019	2019	Cultivation	2019

Table 2: Revision of clade placement using recent phylogenetic analyses. These taxa did not fall into any robust clade prior to 2004 upon when these taxa were originally described. However, due to the availability to cultivate these species in laboratory conditions more effectively, this has allowed taxonomic identification (in this case after 2004) of these species. Phylogenomic investigations have since revised the positions of these taxa into more accepted phyla. Adapted with permissions from Burki 2020, 'The New Eukaryotic Tree of Life', UK, Elsevier LTD.

1.2 Molecular approaches to taxa identifying, and categorising microbial eukaryotes

Phylogeny places organisms alongside their sister taxa. Phylogenies can be built from sequence alignments of one or a few types of sequences corresponding to either ribosomal ribonucleic acid (rRNA) or protein-coding genes. When using the latter either DNA or amino acid (AA) sequences can be used to make the multiple sequence alignment (Carr, et al., 2008). Alternatively, phylogenies can be derived from the much larger genome-wide scale by aligning many genes, which are conserved in evolution (Jackson, et al., 2016). Here, sequences used to build the multiple sequence alignment are concatenated together. To build phylogenies based on genome-wide alignments, sequences for the alignments are collected from either genomes or whole cell transcriptomes. In such studies, paralogous sequence specific to individual taxa are removed. With the advent of several tractable next-generation-sequencing (NGS) technologies (Zhou, et al., 2010), it is becoming the norm for phylogenies to be derived from large-scale or genome-wide analyses. Current views of the eToL are almost exclusively derived from multigene molecular phylogenies (Burki, et, al., 2020).

Hypervariable regions

Traditionally, eukaryotic phylogenies were often based on the alignment of the hypervariable region of 18S (18 Svedberg units, a non-SI; or non-International System of Units, for sedimentation coefficients) rRNA (also referenced as 18S rDNA) arrays (Hejazi, et al., 2010). For instance, alignment of the internal transcribed spacer (ITS), located between the small and large subunits of ribosomal transcription proteins (Hejazi, et al., 2010) was used to identify species relationships. ITS-derived phylogenies are still powerful for assessing relationships within specific taxonomic groups; for instance, in determining relationships between many small flagellates and amoebas in the Rhizaria supergroup (Burki, et al., 2021; d'Avila-Levy, et al., 2015).

'18S' codes for the small RNA sub-unit of eukaryotic ribosomes (SSU), one of the basic components of all eukaryotes (Gregory, et al., 2019) . The reliability of using 18S rRNA analysis for phylogenetics is also illustrated by reference to the kinetoplastids (Hughes & Piontkivska, 2003). The organisation of a typical rRNA gene array is shown in Figure 3 with regions important for use in sequence alignments highlighted. Variable regions (V) help better

understand the shear diversity of microbial community. V4 and V9 (expected amplicon sizes 270-387 base-pairs (bp) and 96-134bp respectively) are more commonly used over other variable regions for metabarcoding with regards to protistan lineages (Choi & Park, 2020). V4 is commonly used for studying the phylogenetic relationships of eukaryotes whilst V9 reveals the extant diversity of eukaryotes. However, despite the advantages of these regions of 18S rRNA, they have rarely been employed for environmental sample analyses due to uncertainties of which taxa are present within the environmental sample, thereby limiting the use of specific V4/V9 primers (Choi & Park, 2020). Differences from 18S rRNA sequencing are currently hugely important in the emergent area of DNA metabarcoding. Metabarcoding differs from barcoding in that many sequences are collected in parallel and compared in order to determine microbial species compositions within different environments (Sorof-Uddin & Cheng, 2015; d'Avila-Levy, et al., 2015). It is the ease of 18S rRNA hypervariable region amplification by polymer chain reaction (PCR) using universal primers that in part allows metabarcoding. The result of barcoding and metabarcoding are the identification of operational taxonomic units (OTUs) to identify the taxa present within a given environment. Limitations, however, include that those universal primers may not be suitable for all groups of eukaryotes (d'Avila-Levy, et al., 2015). Here, kinetoplastids provide a good example (Hughes & Pointkivska, 2003).
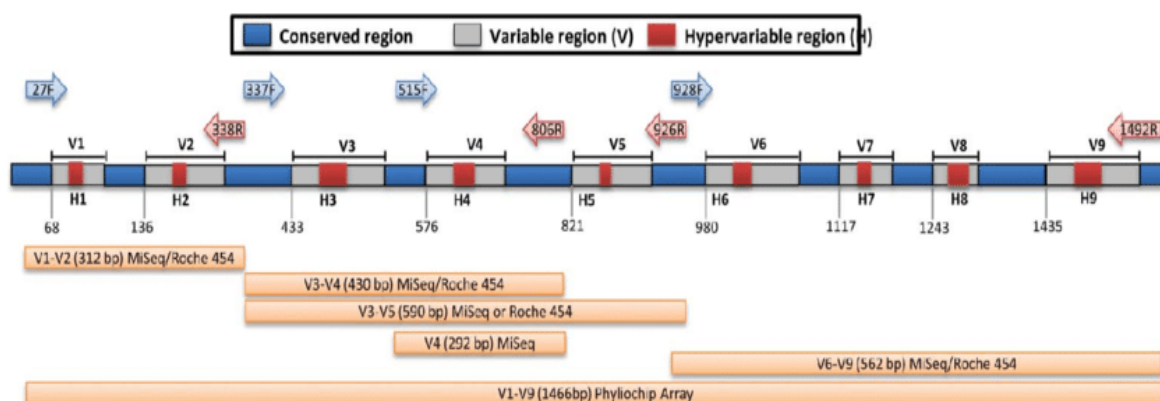
Figure 3: Illustration of the breakdown of different regions of the 16S rRNA gene. The key indicates the different regions associated with the gene. Also illustrating the differences in nucleotide lengths for the variable and hypervariable regions. Whilst not exactly what we look at in this paper (16S rRNA for prokaryotes and plastids, and 18S rRNA being the eukaryotic homologue), this helps visualise the gene and its major components and why we use these in phylogenetic studies. Adapted with permissions by Shahi, 2017, 'Gut Microbiome in MS' Vol. 8, Issue 6, USA, Taylor&Francis Online.

Returning to the wider-scale determination of relationships between different eukaryotes and indeed where the root for the last common ancestor of all eukaryotes is placed, then there is recently released work that moves beyond the consensus views discussed by Burki, et al. (2020) and Lax, et al. (2018). Still a preprint in bioRiv, Cerón-Romero, et al., 2021 performed phylogenomics with 2,786 eukaryotic-specific gene families to place the eukaryotic root between Opisthokonta (eukaryotes which are represented by supergroups Obazoa and Amorphea) and all other eukaryotic lineages. In their analysis, fungi were the earliest branching lineage within the Opisthokonta (Figure 4). In many ways, this conclusion, yet to be widely accepted (Spang, et al., 2022), is reminiscent of the root placement determined by Stechmann and Cavalier-Smith (2002) on a basis of presence/absence of a key gene fusion between dihydrofolate and thymidylate synthase.

1.3 Biodiversity of protists

Protists

The umbrella term 'protist' describes eukaryotes that are generally unicellular, but which are classified as neither animals (or metazoans), land plants, or fungi (Whittaker & Margulis, 1978). Protists are typically either heterotrophic or photoautotrophic in the manner in which they acquire organic carbon. Protists therefore include eukaryotic algae (Sanders, 2011). Whilst not possessing cell walls made of chitin or growing as filamentous hyphae (classic traits associated with fungi) (Raghukumar, 2017), protists, depending on the taxa under study, can either possess cell walls made from any range of macromolecular structures or can lack a cell wall and simply present a plasma membrane to the external environment (*e.g., Naegleria gruberi*). Among heterotrophs various feeding modes are possible, chiefly osmotrophic acquisition of nutrients via uptake through surface transporters, endocytosis, and pinocytosis (Greek; "cell drinking") or through predatory phagotrophy whereby prokaryotes, other eukaryotes (including cells from multicellular animals) are actively eaten, sometimes hunted, for nutritional purposes (Sanders, 2011; Leander, 2020).

Cellular motility

Two principal modes of cellular motility are known: flagellum- (or cilium-) mediated cell swimming and cell crawling such as that seen in amoebae. Cell crawling (also referred to as α-motility by Fritz-Laylin, et al., 2017b) is a mode of motility regulated by motility-related genes. These include actin-related protein (Arp2/3) complexes, formins, Wiskott-Aldrich Syndrome Protein (WASP), and Sequence Characterized Amplified Region (SCAR/WAVE—synonymous abbreviations) genes that work in tandem to allow a cell to form pseudopodia and therefore promote cell crawling abilities (Kaneshiro, 1995; Fritz-Laylin, et al., 2017a; Fritz-Laylin, et al., 2017b). SCAR/WAVE are synonymous with one another both being the same protein identified independently by two different groups of scientists (Bear, et al., 1998; Miki, et al., 1998).

Across the breadth of microbial eukaryotic diversity there are wide-ranging examples of multi-cellular forms whereby individual cells come together to form fruiting bodies multi-cellular aggregates (*e.g. Dictyostelium discoideum* (Kin & Schaap, 2021) or even undergo animal-like patterns of embryogenesis such as that seen in green algae *Volvox carteri* (Matt & Umen, 2016)).

Such diverse modes of nutrient acquisition and cell motility mean, perhaps unsurprisingly, there is huge morphological diversity to be seen when looking across the breadth of microbial evolution. Protists also form complex community structures, which are often ill-understood, in the environments and ecosystems in which they are found. These ecosystems are many and varied. Moist soil, fresh-water, brackish (more salinity than fresh-water, but less than typical sea water) water, and seawater are all hosts to diverse protist communities. In the surface waters of photic zone (200m below surface) of the ocean, huge diversity in the range of diatoms, dinoflagellates, and other eukaryotic algae are known (Keeling, et al., 2005) and collectively these taxa are responsible for ~50% of photosynthetic CO_2 uptake annually (Berdjeb, et al., 2018).

Protist biodiversity

The deep-sea is also home to a rich diversity of protist communities, including some communities that contain kinetoplastid species. Dry or desert soils are also home to protist communities (*e.g., Prosopis laevigata*, and *Parkinsonia praecox* (*Fabaceae*), (Pérez-Juárez, et al., 2018)) and some of the recently described protist orphan taxa have been isolated from diverse sources in the tropics (*e.g., Microheliella maris* (Yazaki, et al., 2022)). Collectively these studies point to a huge protist biodiversity that remains to be discovered. The roles that such protists play in complex community ecosystem ecology is often ill-understood. There are, however, some protists that are well-documented in their roles involved with ecological processes in their environments. Marine phytoplankton (such as *Ostreococcus lucimarinus* and *Scrippsiella trochoidea*) are increasingly appreciated for the importance they play in in the global ecological processes and biogeochemical cycles of oceanic ecosystems and the interplays of which they provide for their communities (Berdjeb, et al., 2018).

Complexity of cellular life cycles

For some protists their life cycles are complex, requiring cellular differentiation into diverse morphological forms. Nutrient availability, temperature changes, and changing pH, oxygen tension metabolite concentration, or other external cues typically provide the trigger or stimulus for cellular differentiation (Corliss, 2001; Persson, 2001; Padilla & Savedo, 2013). Again, the kinetoplastids, and indeed other taxa from the phylum Euglenozoa, provide prime examples of protists with complex cellular life cycles. Some species of protists form colonies or aggregates, allowing the pellicle to change shape and secreting a glue-like matrix so individual protists bond together. *Lenisia limosa* (a breviate), for example, has both swimming and adherent gliding forms utilising the flagellum in different ways to prey on bacteria dependent upon its life stage (Hamann, et al., 2016).

Identification of novel organisms

Prior to advances in metabarcoding technologies, most organisms were identified using low/medium-throughput sequencing, such as Sanger or Illumina approaches. However, since then, metabarcoding has partially helped broaden the scope of how we view the species compositions of environmental samples (Behjati & Tarpey, 2013). Consequently, this has led to new stems of biology, namely metagenomics and metatranscriptomics, and more high-throughput sequencing methods now readily available for both prokaryotes and eukaryotes. Although still in the infancy stages of applying metagenomics to eukaryotes; DNA metabarcoding (capable of examining whole communities) and fluorescent *in situ* hybridization (FISH) techniques have aided in relaxing some of the burdens often associated with culturing specific species (Keeling, 2019; Machida, & Knowlton, 2015). For example, the abundance of kinetoplastids has been expanded using catalysed reporter deposition-fluorescent *in situ* hybridization (CARD-FISH) techniques, signal amplification in junction with FISH (Neuenschwander, et al., 2015), and diversity using 18S kinetoplastid-specific primers in amplicon sequencing (Mukherjee, et al., 2015; Mukherjee, et al., 2019). Furthermore, by retaining the ecosystem that the sample is taken from (*i.e.* food sources, temperatures, salinity, or mineral content within marine ecosystems) it is possible to observe more readily the complexity of some life cycles and morphology changes in some protists (Burki, et al., 2020).

Appendage remodelling

Keeling (2019) described the complexity of protist genomes, some of which I look at in this MSc (*e.g.,* the kinetoplastids, which are capable of multi-purpose gene regulation and RNA editing events (Damasceno, et al., 2020), which complicate identifying discrete function of protein-coding genes). One conserved and iconic eukaryotic structure Keeling (2019) discussed is the eukaryotic flagellum or cilium, and the complexity of motility and function that this organelle can provide including: *e.g.* whipping or beating, propelling the organism through its environment; gliding motility, which also aids in stabilizing the organism; or completely changing the function from one of motility to sensing, anchoring (as seen in some trypanosomatids and choanoflagellates), or feeding (*e.g.* through the cytostome-cytopharynx apparatus as observed in some kinetoplastids). This ability to repurpose structures or organelles, such as the flagellum, allows one to consider the ability to transform cell structures across the eToL, thereby giving rise to morphological complexity better suited to dynamic environments or rather physical and which potentially enable these organisms to readily alter their transcriptome, and possibly their morphology, in response to environmental cues (Keeling, 2019).

1.4 Kinetoplastids, biodiversity, evolution, and ecology

Kinetoplastids are a well-studied order of flagellate protists and are among the most readily isolatable protists from diverse aquatic environments (Flegontova, et al., 2020; Mukherjee, et al., 2019). Kinetoplastids belong to the phylum Euglenozoa. Free-living kinetoplastid species that can be readily isolated from ponds, lakes, and marine environments, however, tend to be the least studied of these important, cosmopolitan protist groups. Is it the parasitic trypanosomatids that are the much more comprehensively studied kinetoplastid group owing to the medical and veterinary importance of some trypanosomatids (Medkour, et al., 2020). Thus, *Trypanosoma cruzi*, *Trypanosoma brucei*, and several of ~30 known *Leishmania* species (Jackson, et al., 2016) cause serious tropical diseases in people (Pizarro, et al., 2013). There are other *Trypanosoma* species that cause economical very significant animal diseases in Africa, and South America. There is even one trypanosomatid genus (*Phytomonas*) responsible for diseases in plants including commercially important crops (Ivanoff, 1933). These parasites are transmitted between hosts by insect vectors. The majority of trypanosomatid species simply parasitize the digestive tract of their insect host (Kaufer, et al., 2017), although the fitness cost to the insect host of this apparent parasitism is not clear.

Kinetoplast

Kinetoplastids owe their name to the highly unusual appearance of their mitochondrial genomes. In these protists, there is a huge mass of mitochondrial DNA (mtDNA) composed of two, circular DNA classes—maxicircles (20-40 kilo-bases (kb)) and minicircles (0.5-10kb, depending on the species) (de Souza, et al., 2010; Lukeš, et al., 2002). Maxicircles contain mitochondrial rRNA genes and genes encoding hydrophobic mitochondrial proteins, predominantly involved with oxidative phosphorylation whilst mini-circles contribute guide RNAs (gRNA) that direct the modification of maxicircle transcripts by uridylate insertion or deletion via RNA editing mechanisms (Maslov & Simpson, 2007; Michaeli, 2015).

In trypanosomatids (examples of which are found in Figure 5), maxicircles (30-40 copies of the 20-40kb molecule) are catenated with ~5000 minicircles forming the kinetoplast (kDNA) structure that is readily viewed using standard nucleic acid stains (e.g. 4',6-diamidino-2-phenylindole, or DAPI), and is physically attached as discrete structures to the flagellar basal bodies via a series of fibres that extend from the kinetoplast through inner- and outer-mitochondrial membranes to the proximal end of the flagellar basal bodies (Lukeš, et al., 2018; Ogbadoyi, et al., 2003). In non-trypanosomatid kinetoplastids maxi-circles and mini-circles are not catenated (i.e. do not form a chain or repetitive series). Rather they are 'free' DNA molecules present in huge copy numbers and can be found in a variety of organisations either in one region of the mitochondrion as a pro-kDNA network (as found in *Bodo saltans*, pro- denoting a single globular bundle), in several discrete regions of the mitochondrion (poly-kDNA as seen in *Dimastigella trypaniformis*, poly- denoting multiple distinct foci throughout the mitochondrial lumen) or dispersed throughout the mitochondrion as pan- or mega-kDNA (seen in *Cryptobia helicis* or *Trypanoplasma borreli*, respectively, pan- denoting supercoil forms and mega- denoting that the minicircle-like sequences are tandemly linked into larger molecules) (Lukeš, et al., 2002; Lukeš, et al., 2018; Midha, et al., 2021). In Table 3, the different kinetoplastid subclasses in which these different mitochondrial DNA arrangements can be seen, together with identities of some of the key kinetoplastid genera. In Figure 4 phylogenetic relationships between different kinetoplastids as determined from 18S rRNA maximum likelihood phylogeny is shown. Within this figure there is also a cartoon illustration insert of some of the morphological diversity that is seen within euglenozoan protists.

Phylum	Class	Subclass	Order	Genera
Euglenozoa	Kinetoplastea Euglenoidea Diplonemea Symbiontida	Prokinetoplastina	Prokinetoplastida (Polykinetoplastic kDNA)	*Ichthyobodo*
				Perkinsela
		Metakinetoplastina	Trypanosomatida (Eukinetoplastic kDNA)	*Trypanosoma*
				Paratrypanosoma
				Leishmania
				Herpetomonas
				Rhynchoidomonas
			Neobodonida (Eu-/Polykinetoplastic kDNA)	*Neobodo*
				Cruzella
				Rhynchobodo
				Rhynchomonas
			Eubodonida (Eukinetoplastic kDNA)	*Bodo*
			Parabodonida (Pankinetoplastic kDNA)	*Parabodo*
				Trypanoplasma

Table 3: Taxonomic model of Euglenozoa. Depicting the breakdown of Euglenozoa into the classes Kinetoplastea, Euglenoidea, Diplonemea, and Symbiontida. Kinetoplastea is further broken down into the Pro/Metakinetoplastina yet does not depict the Basal kinetoplastids outlined in a more recent Tikhonenkov's 2021 revision. Orders in this table explain the type of kDNA associated with each Order and subsequent Genera. For ease of use, some species have been omitted with focus presented to more commonly referenced kinetoplastids. Amended and adapted with permission from d'Avila-Levy 'Exploring the Environmental Diversity of Kinetoplastid Flagellates in the High-Throughput DNA Sequencing Era', 2015, Oswaldo Cruz Institution.
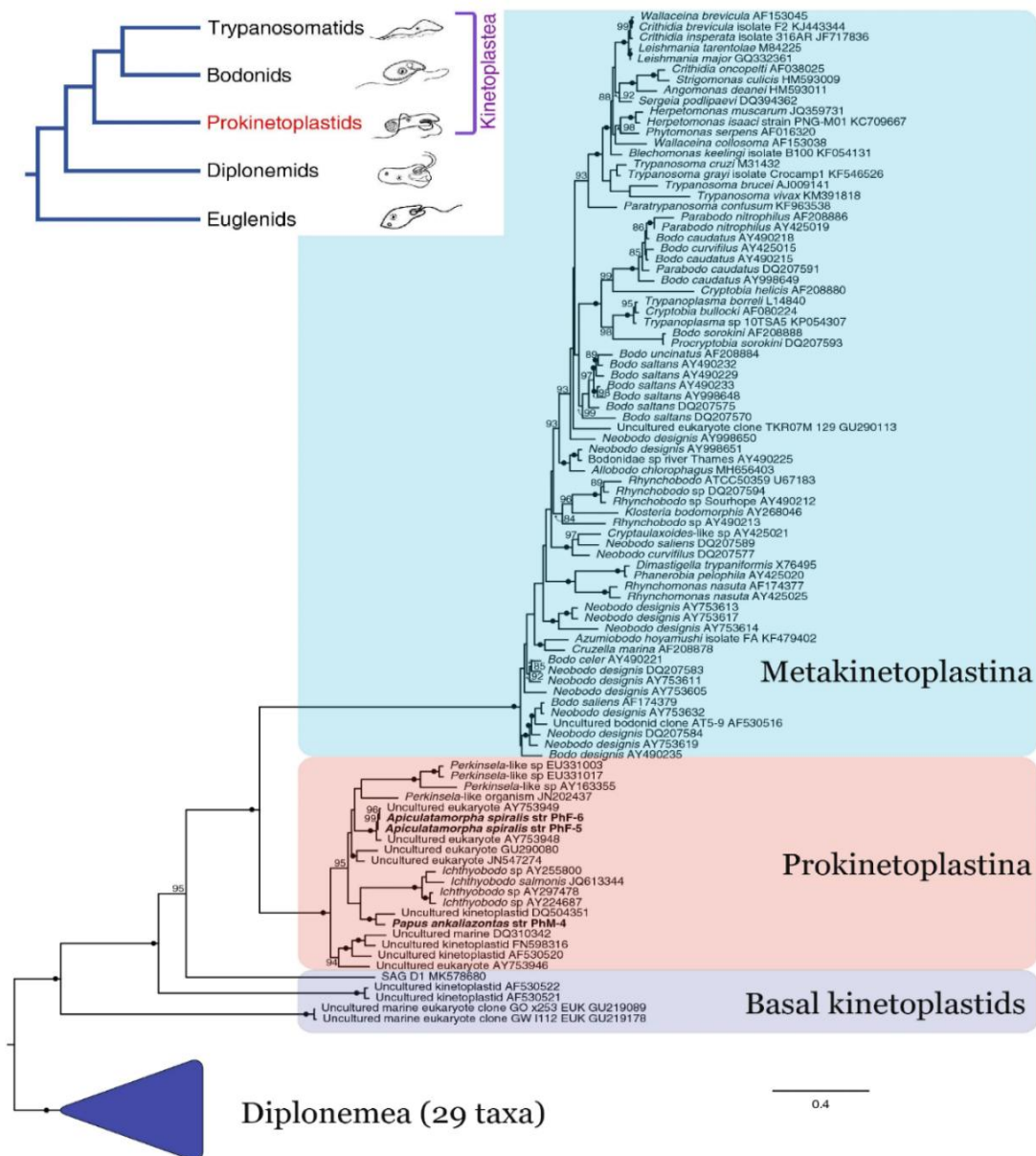
Figure 4: Maximum likelihood phylogeny parsimony tree of the 18S ribosomal RNA gene from kinetoplastids and diplonemids. build using ultrafast bootstrapping methods on IQ-TREE software (phologenomic inference software). This breakdown of kinetoplastids shows the relation between species and divergence patterns that can be attributed to genome streamlining observed within parasitic members. Open Access provided by Tikhonenkov (2021), 'First finding of free-living representatives of Prokinetoplastina and their nuclear and mitochondrial genomes', Springer Nature publishing, Creative Common License, no changes have been made to the original source.

Among euglenids, the second of three major lineages in Euglenozoa, their cellular forms are defined in large part by the organisation of extra-cellular pellicles. In contrast, the kinetoplastids, microtubule-based cytoskeletons determine overall cell morphologies and placement of the flagellar basal bodies from which either two flagella, in the free-living kinetoplastids (with the exception of intracellular, aflagellate *Perkinsela* (Tanifuji, et al., 2017)), or generally one flagellum (the trypanosomatids) are built. *Vickermania* is the one known example of a biflagellate trypanosomatid, its two flagella being an adaptation to life inside its insect host (Kostygov, et al., 2020; Clayton, 2016).

Morphological complexity of kinetoplastids

Within trypanosomatids, cell morphology (or form) is classically described as function of kinetoplastid placement relevant to the nucleus. Across the kinetoplastids, there is a little evidence for the formation of resting cysts except possibly for a few taxa (Dias, et al., 2014). There is currently little insight into possible examples of cellular differentiation in free-living kinetoplastids. Yet, in trypanosomatids there is considerable understanding of how environmental cues encountered when moving between host and vector or migration within the digestive tract of the insect vector lead to striking examples of cellular differentiation and changed morphological appearance (Dias, et al., 2014). In some instances, the molecular or environmental triggers for trypanosomatid differentiation are understood. For instance, *T. brucei* and *T. cruzi* are well established due to medical importance as stated above and as such, are readily researched for an effective cure (Mansur-Pontes, et al., 2021). Whilst there remains little evidence as yet for morphological complexity occurring within the life cycles of free-living kinetoplastids, it is also fair to say that this possibility has been seldom addressed (Butenko, et al., 2021). Moreover, stepping outside of the kinetoplastids to their nearest evolutionary relations in the Euglenozoa, the diplonemids, then some species of these abundant marine flagellates (Tashyreva, et al., 2022; Butenko, et al., 2020) have recently been observed to undertake surprisingly complex life cycles.

As with examples of trypanosomatid differentiation, it appears to be external environmental cues that trigger cellular differentiation, including some complex shifts in the appearance of diplonemids cytoskeletal architecture (Tashyreva, et al., 2022). In particular, *T. brucei*, *Angomonas deanei*, and *Strigomonas culicis* (Votýpka, et al, 2014) have been observed to build a paraflagellar rod (PFR) at only some stages in their life cycles. The PFR is considered a well-known extra-axonemic structure found in many euglenozoan protists; it runs alongside and is physically attached to the flagellar axoneme in euglenozoans where it is found (Alves, et al., 2020).

Bacterial endosymbionts

Exploring further cellular diversity in the kinetoplastids, then a few taxa are characterised by the presence of endosymbiotic bacteria which tend to lay in the anterior region of the cytoplasm, proximal to the kinetoplasts, but well away from the cytostome-cytopharynx (see Figures 6 & 10) (Harmer, 2018; Butenko, 2021). Of the trypanosomatids with bacterial endosymbionts (e.g., *Candidatus Pandoraea novymonadis (Ca. Pa. novymonadis)*, and *Candidatus Kinetoplastibacterium (Ca. kinetoplastibacterium)* (Silva, et al., 2018)), the endosymbionts have several interesting characteristics which have been documented (Kostygov, et al., 2016) and attributed to obligate endosymbiotic lifestyles. These include: a reduction in GC-content (in comparison to closely-related, free-living bacterial relatives); a reduction in both genome size and gene content, which would suggest streamlining of the proteome and/or removal of processes that other closely-related species genomes code for; and lastly a rarity of mobile elements (Harmer, et al., 2018) which are associated with gene duplication events and mutations in protein coding regions, altering protein functionality (Singh, et al., 2014). These streamlining events become more apparent the more we understand how an endosymbiont, such as *Ca. Pa. novymonadis,* benefits its host. The endosymbiont provides the host with heme (also spelt haem), several essential AAs, purines, coenzymes, and vitamins and in return the host offers: several nonessential AAs, phospholipids, and enzymes for carbohydrate metabolism that are otherwise missing from the endosymbiont (Zakharova, et al., 2021). This relationship, over time, could have impacted gene expression of the endosymbiont-bearing trypanosomatids, leading to the subsequent silencing or out-right deletion of unnecessary or 'abandoned' genes improving overall fitness and streamlining of the trypanosomatid genome in this manner.

Endosymbiotic bacteria have been observed in only a few kinetoplastid species: among free-living kinetoplastids, *Cryptobia,* and *Bodo,* and in the parasitic trypanosomatids they have been found in *Strigomonadinae*, and *Novymonas*. The phylogenetic positions of these endosymbiont-containing kinetoplastids relative to other kinetoplastids is shown in Figures 5 & 6 below.

Kinetoplastid hosts and endosymbiont line up their replication and division together allowing the former to entrain the latter. For the *Ca.* kinetoplastibacterium, a bacterial endosymbiont found within Strigomonadinae, endosymbiont duplication happens early within the host cell cycle, occurring following kinetoplasts replication (Harmer, et al., 2018). Endosymbiont division then follows by movement of the endosymbiont to the opposite outer-face of the nucleus where mitosis and new flagellum elongation beyond the flagellar pocket then conclude the latter stages of the host cell cycle prior to cytokinesis generating new daughter cells (Harmer, et al., 2018).

Figure 5: use of aligned and alignment-free analysis of the maxicircle coding region (mtDNA within the kinetoplast) between species of the trypanosomatid family. A) Use of PhyML-derived phylogeny showing relationships between trypanosomatids using multiple sequence alignment analysis. B) Feature Frequency Profiles-estimated phylogeny, again, using alignment-free analysis of the maxicircle coding region. Red branches indicate a different clustering from (A). A black diamond represents a node with 100% bootstrap for both (A) and (B). the scalebars indicate the number of nucleotide substitutions per position. Open Access permitted from the Kaufer, 2019, 'Evolutionary Insight into the *Trypanosomatidae* Using Alignment-Free Phylogenomics of the Kinetoplast', MDPI AG publishing. No changes have been made to the original source.

Figure 6: Kinetoplastid phylogeny and history of endosymbiosis. Taxa with bacterial endosymbionts are boldened. •= presence of cytostome-cytopharynx complex. O/○= uncertainty from lack of data or unlikeliness based on published electron microscopy studies. Reproduced with permission from Harmer, 2018, 'Farming, Slaving, and Enslavement: Histories of endosymbiosis during kinetoplastid evolution' paper, UK, Cambridge University Press 2018.

Obligate lifestyles

A key question in kinetoplastid evolution is how the transition from a free-living status to obligate parasitism occurred. To understand how parasitism may have evolved, in trypanosomatids, Jackson, et al., (2016) compared nuclear genomes of free-living *B. saltans*, the then closest known relative of trypanosomatid family, and *T. borreli,* a parabodonid fish parasite, with streamlined trypanosomatid genomes. Parasitic lineages show functional complexity with species-specific features, including an absence of biosynthetic pathways for heme, purines, and aromatic AAs which can be obtained from the host.

Results from Jackson, et al., (2016) showed the parasite genomes were 18-34% smaller than that of *B. saltans*, but there was a 41-56% reduction in gene content, indicating parasite genomes are perhaps surprisingly less dense (Lukeš, et al., 2018). This implies *B. saltans* has a compact genome, compared to trypanosomatid models (Opperdoes, et al., 2016), conserving space by reducing the size of intrinsic (coding) regions whilst the parasitic trypanosomatid genomes contain more non-coding (extrinsic) DNA. This results in *B. saltans* having twice as many coding genes as their parasitic counterparts (Jackson, et al., 2016).

Trypanosomatid cell variation

Some variations in cell form that have been described in trypanosomatids are shown in Figures 7 and 8 below. Here, placement of the kinetoplast posterior to the nucleus with an attached flagellum running along the cell body is known as the trypomastigote form; a kinetoplast anterior to the nucleus with an attached flagellum is called an epimastigote; a kinetoplast at the anterior cell end with a free flagellum extending from the cell body is typically called a promastigote; an anterior cell end kinetoplast but with no flagellum extending from the flagellum pocket gives rise to an amastigote. Such cellular forms were originally and classically described in the 1960s (Hoare & Wallace, 1966).

In a wide-ranging analysis of trypanosomatid morphologies in ~250 trypanosomatid isolates, Wheeler (2013) proposed constraints and extrinsic selection pressures that limit minimum and maximum cell dimensions available to trypanosomatid parasites, and which thus define the full diversity of trypanosomatid morphologies that are possible. Wheeler (2013), performed meta-analysis on the limits of trypanosomatid cell shapes associated with constrains or selective pressures using previously published trypanosomatid morphometric data to generate a database. Their research yielded ~250 references of trypanosomatids from different hosts; he recorded cell body length and width, free and total flagellum length, and kinetoplast to posterior distance (Figures 8 & 9). The data collected focused on motile life cycle stages and represents coverage of the *Trypanosoma, Phytomonas*, Leishmaniinae*, Blastocrithidia, Herpetomonas, Paratrypanosoma*, and the endosymbiont-bearing clades. Their findings revealed two distinct classes of life cycles: those that transition between trypomastigote, epimastigote, and/or amastigote (juxtaform- Latin, *juxta* meaning beside), in reference to the flagellum is attached to a length of the cell body following exiting from the flagellar pocket, and those that transition between promastigote, choanomastigote, opisthomastigote, and/or amastigote (liberform- Latin, *liber* meaning free), in reference to the flagellum where it is not attached to an extended region of the cell body following exit from the flagellar pocket (Wheeler, et al., 2013).
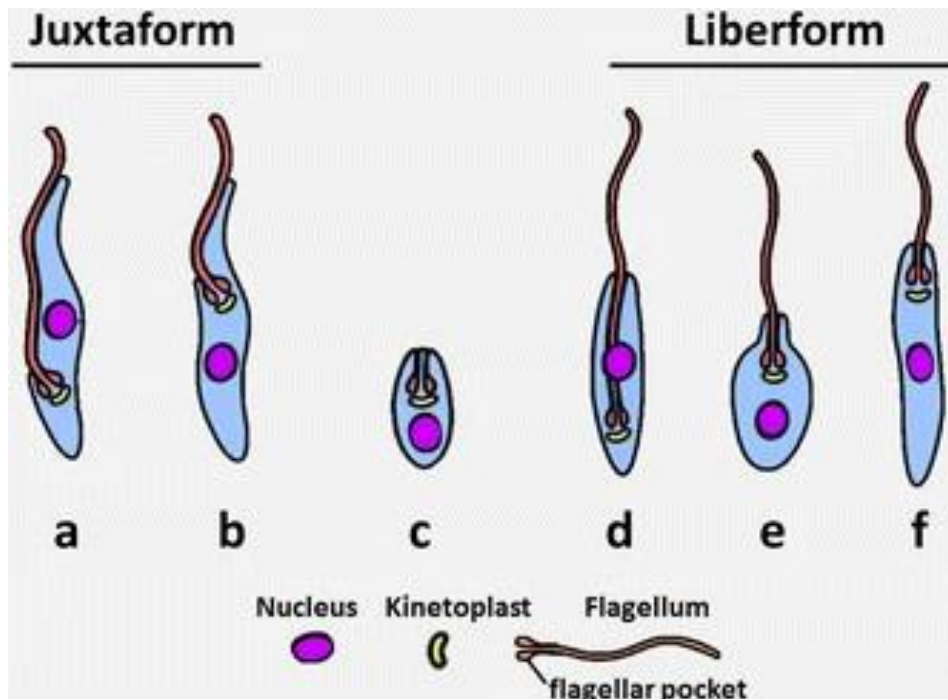
Figure 7: Juxtaform and Liberform general morphologies. Illustrating the different forms of which are observed in juxta- and liberforms of the trypanosomatid cell cycles. Depicting a) trypomastigote, b) epimastigote, c) amastigote, d) promastigote, e) choanomastigote, and f) opisthomastigote. Open Access permission from Kaufer (2017), 'The Evolution of Trypanosomatid Taxonomy'. BioMed Central; BIOMED CENTRAL LTD., No changes have been made to the original source.
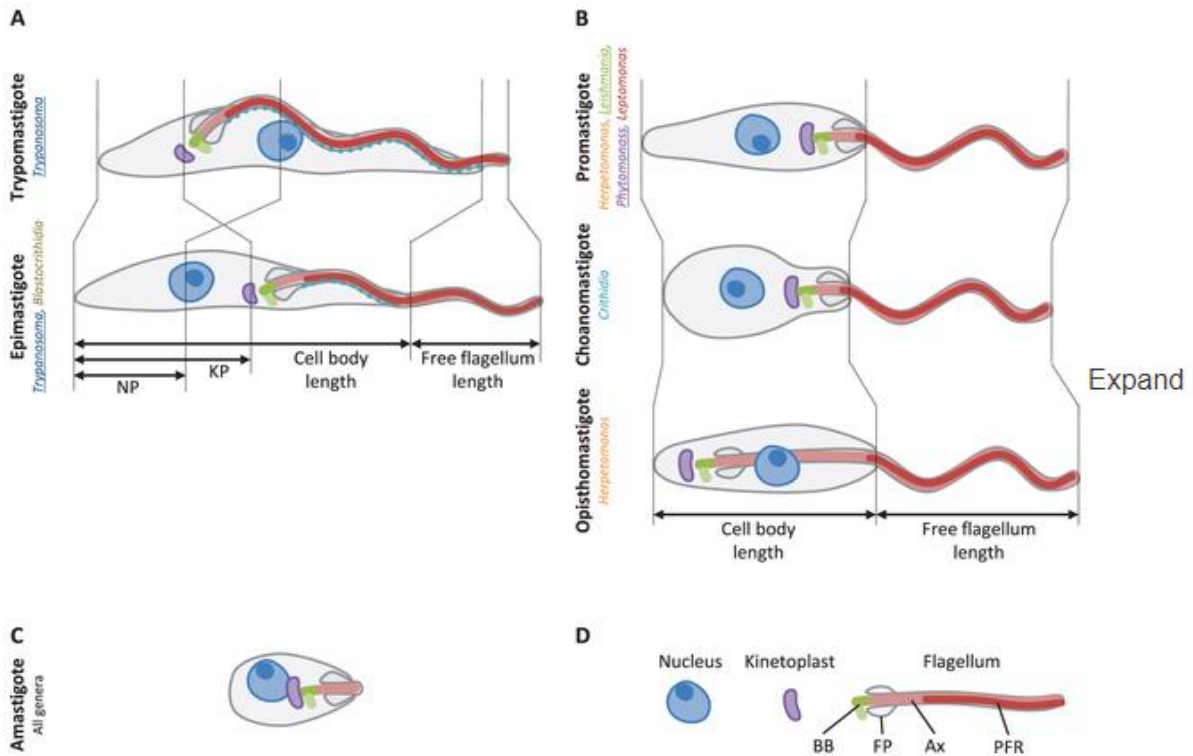
Figure 8: Positions and length of cytoplasmic cell structure identified within the life cycle of different Kinetoplastids. This highlights both the Juxtaform and Liberforms of each in pictures A) and B) respectively. KP denotes the kinetoplastid-posterior distance whilst NP represents the nucleus-posterior distance. Picture C) shows the Amastigote morphology, a form taken by parasitic trypanosomes that allow them to replicate within the host cell. D) outlines the key structures associated with the flagellum. These include the basal body (BB), flagellar pocket (FP), the axoneme (Ax) and the paraflagellar rod (PFR). Reproduced with permissions from the Wheeler, (2013), 'The Limits on Trypanosomatid Morphological Diversity', Open access provided by PLoS publishing, no changes have been made to the original source.
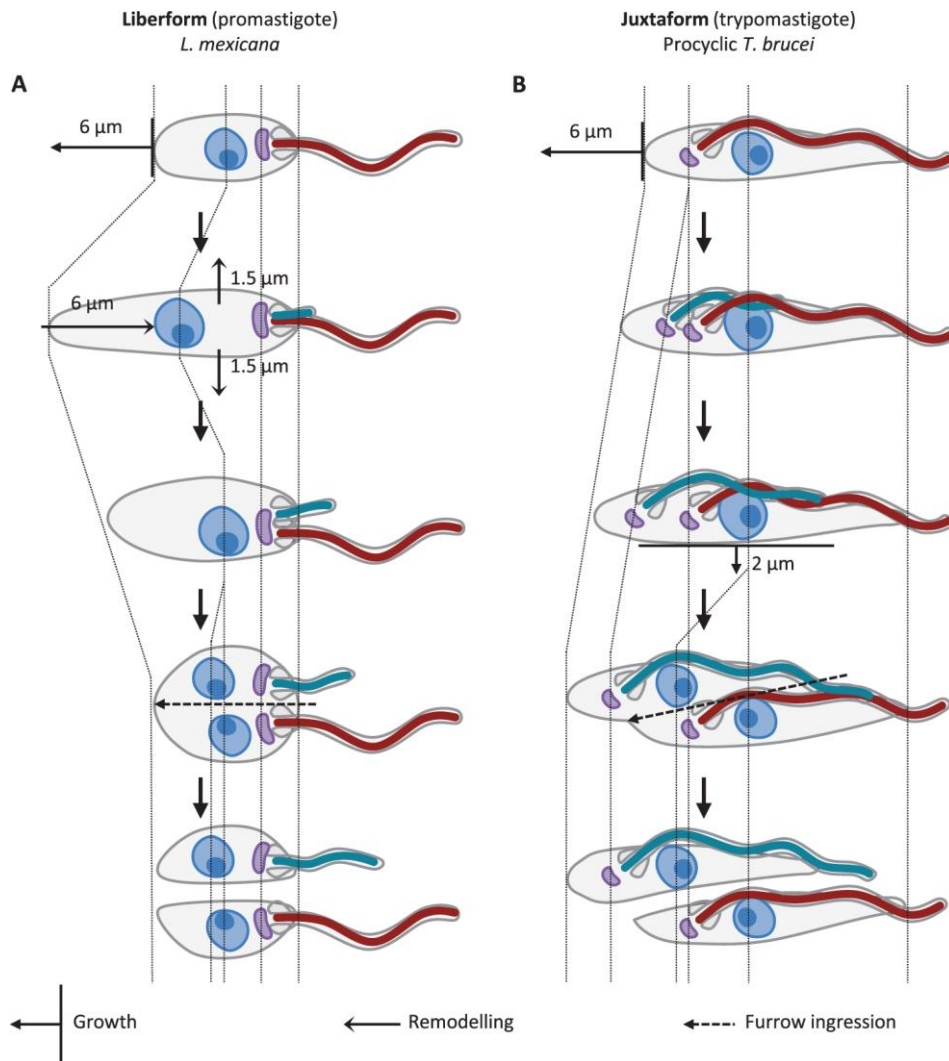
Figure 9: Cell body morphogenesis defined by which form the species (promastigote and trypomastigote) through the cell cycle. Indicating which areas grow and retract undergoing cell remodelling. Old flagellum and new flagellum are coloured red and blue respectively indicating the formation of new cellular components within the cell. The dashed line indicates where the cell splits. Credit to the Wheeler, et al., 2017 paper on trypanosomatid morphological diversity. Illustration sourced from Kaufer (2019), 'Evolutionary Insight into the Trypanosomatidae Using Alignment-Free Phylogenomics of the Kinetoplast'. , 'Evolutionary Insight into the Trypanosomatidae Using Alignment-Free Phylogenomics of the Kinetoplast', MDPI AG publishing. No changes have been made to the original source.

Cytostome-cytopharynx

It seems likely that osmotrophic trypanosomatids (*e.g., T. brucei*) evolved from a phagotrophic feeding, free-living, ancestor suggested by the presence of a cytostome-cytopharynx complex; a monolayer of cross-linked microtubules forming a sub-pellicular corset, that acts as a mouth, absorbing nutrients which also prevents general endocytosis or membrane invagination across the cell surface (Harmer, et al., 2018; Butenko, et al., 2021) in free-living kinetoplastids and its retention in some trypanosomatids (Figures 6 & 10). Membrane invagination can, however, occur at points in the life cycle where the sub-pellicular corset is absent as observed in the African trypanosomes and *Leishmania* where the flagellar pocket forms around the singular flagellum which, in these trypanosomatids, is the site of feeding on bacteria and have observed to distend to absorb larger prey (Harmer, et al., 2018; Butenko, et al., 2021).

Enzymatic machinery used for digesting more complex macromolecular structures appears to have been lost at an early point during divergence of the last trypanosomatid ancestor (Butenko, et al., 2021) indicating that uptake of a bacterium by a cytostome-bearing trypanosomatid does not incline immediate digestion. In the basal trypanosomatid *Paratrypanosoma confusum*, a monoxenous kinetoplastid and the most basal branch of trypanosomatids, the cytostome-cytopharynx complex is retained and as it shares a close relationship with the *Leishmanias* and *Crithidia fasciculata* it is indictive that organelle degradation and loss was likely complex (Harmer, et al., 2018). This, coupled with *T. bruci* having stage-regulated cytostome-cytopharynx assembly shows possibility of a cryptic/hidden cytostome in other extant trypanosomatids implying that endosymbiont uptake through cytostome cell entry could be how *Novymonas* and strigomonads evolved (Harmer, et al., 2018; Skalický, et al., 2017).
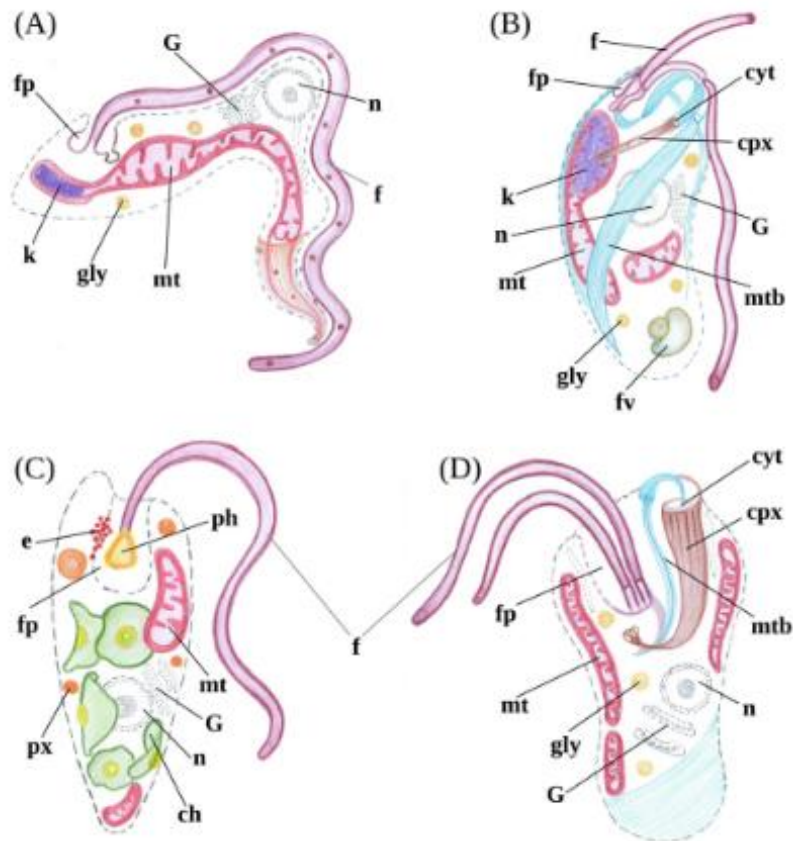
Figure 10: Schematic Morphologies of several flagellated species. These show a Trypanosomatid, *T. brucei* (A), a Bodonid, *B. saltans* (B), a Euglenid, *E. gracilis* (C), and a Diplonemid, *D. japonicum* (D). annotation abbreviations: ch- chloroplast; cpx- cytopharynx; cyt- cytostome; e- eyespot (associated with light reception); f- flagellum; fp- flagellar pocket; fv- food vacuole; G- Golgi apparatus; gly- glycosome; k- kinetoplast; my- mitochondrion; mtb- microtubules; n- nucleus; ph- photoreceptor; px- peroxisome. Reprinted with Open Access permissions from Butenko (2021), 'Reductionist Pathways for Parasitism in Euglenozoans? Expanded Datasets Provide New Insights'. Elsevier publishing. No changes have been made to source.

1.5 Diplonemea is the third major lineage in Euglenozoa

Within the Euglenozoa, Euglenida and Diplonemea are the two other major lineages. Euglenids (Euglenida) are single-flagellated (but can be biflagellated), eukaryotic protists with phototrophic, osmotrophic, and phagotrophic members. Phagotrophic euglenids (Petalomonadida, Ploeotiidae, and Spirocuta) are poor swimmers, motile through α-motility or coordinated body deformations (Lax, et al., 2021) and like their kinetoplastid counterparts, are predominantly found in fresh water sources with only a few species found in marine environments.

Diplonemid morphology

Diplonemids have varied morphologies and have a versatile way of life which allows them to adapt to numerous oceanic niches. Diplonemids fall into four major lineages; 1. Classical diplonemids (*Diplonemidae*) which include benthic and planktonic species: *Diplonema, Rhynchopus, Lacrimia, Flectonema,* and *Sulcionema.* 2. Hemistasiids (*Hemistasiidae*) that are a small planktonic clade composed of: *Hemistasia, Artemidia*, and *Namystynia*. 3. Diverse, deep-sea pelagic diplonemids (DSPD I) (Eupelagenemidae), and 4. A smaller, novel, deep-branching diplonemid clade; DSPD II (Lara, et al., 2009).

Diplonemids are described as cylindrical or flatten-bodied, bi-flagellated, free-living protists that can be heterotrophic, or bacterial phagotrophs and represent a large proportion of planktonic microbial species found in the Earth's oceans (Schoenle et al., 2021; Flegontova, et al., 2020). The class Diplonemea is comprised of two families including: *Diplonemidae* and *Hemistasiidae*.

And in contrast to the well-documented kinetoplastids and euglenids, remain relatively overlooked in terms of molecular phylogenetic characterisation, in part attributed to technical limitations associated with SSU rRNA V4 region. Recent papers suggest that targeting the V4 region in conjunction with the V5 and V9 regions of the 18S rRNA gene detectable in virtually all samples of seawater, by metabarcoding techniques could be used to observe the abundance and diversity of diplonemids (Kaur, et al., 2020; Burki, et al., 2021). Experiments to date have shown that diplonemids are a diverse group, and an extremely abundant member of marine ecosystems, including in the deep sea.

Results from Butenko, et al., (2020) have shown that diplonemids and euglenids differ, in addition to the obvious morphological characteristics, from kinetoplastids through their repertoires of protein kinases and phosphatases, and modes of nucleotide metabolism, and lipid biosynthesis. Metabolic gene loss in kinetoplastids mainly comprise of those involved metabolism of AAs, nucleotides, cofactors, vitamins, and lipids reflected in the repertoire of protein kinases, phosphates, peptidases, glycosyltransferases (Figure 11). As previously mentioned, *B. saltans*, related to parasitic trypanosomatids, lost several complete metabolic pathways of AAs, purine, and ubiquinone biosynthesis; but this does not represent parasitic reduction (Butenko, et al., 2020). The Butenko, et al., (2020), data reveals that certain pathways were probably lost even earlier in the evolution of kinetoplastids and not tied to changes in obligate lifestyles. They suspect that there is some predisposition to losing metabolic pathways which could make the representative of these lineages prone to switching lifestyles.

RNA editing

Diplonemids have a mitochondrial genome capable of systematic gene fragmentation and RNA editing nucleobases (adenine-A, cytosine-C, guanine-G, thymine-T, uracil-U, and inosine-I) including C-to-U, A-to-I, G-to-A substitutions (not previously detected in literature at the time of Kaur, (2020)) and 3'-uridine/adenosine additions (U/A appendages) (Figure 12) conserved across the diplonemids (Lukeš, et al., 2018; Butenko, et al., 2020). C-to-U interconversion can proceed via transamination or deamination (U-to-C and C-to-U respectively). This deamination could also be used for A-to-I conversion following adenine appendage editing similar to kinetoplastids with U-insertion/deletion. Euglenids, such as *Euglena gracilis*, showed no sign of the aforementioned substitutions or insertion/deletions described in kinetoplastids and diplonemids, although discussions point to an uncertainty of the splicing mechanisms involved (Dobáková, et al., 2015; McWatters, & Russel, 2017). Euglenids harbour a small compliment of supercoiled, non-catenated, linear DNA in their mitochondria. In comparison to kinetoplastids, the amount of kDNA is invariably expanded leading to higher percentages of cellular DNA being mitochondrial (Butenko, et al., 2021; Faktorová, et al., 2016).
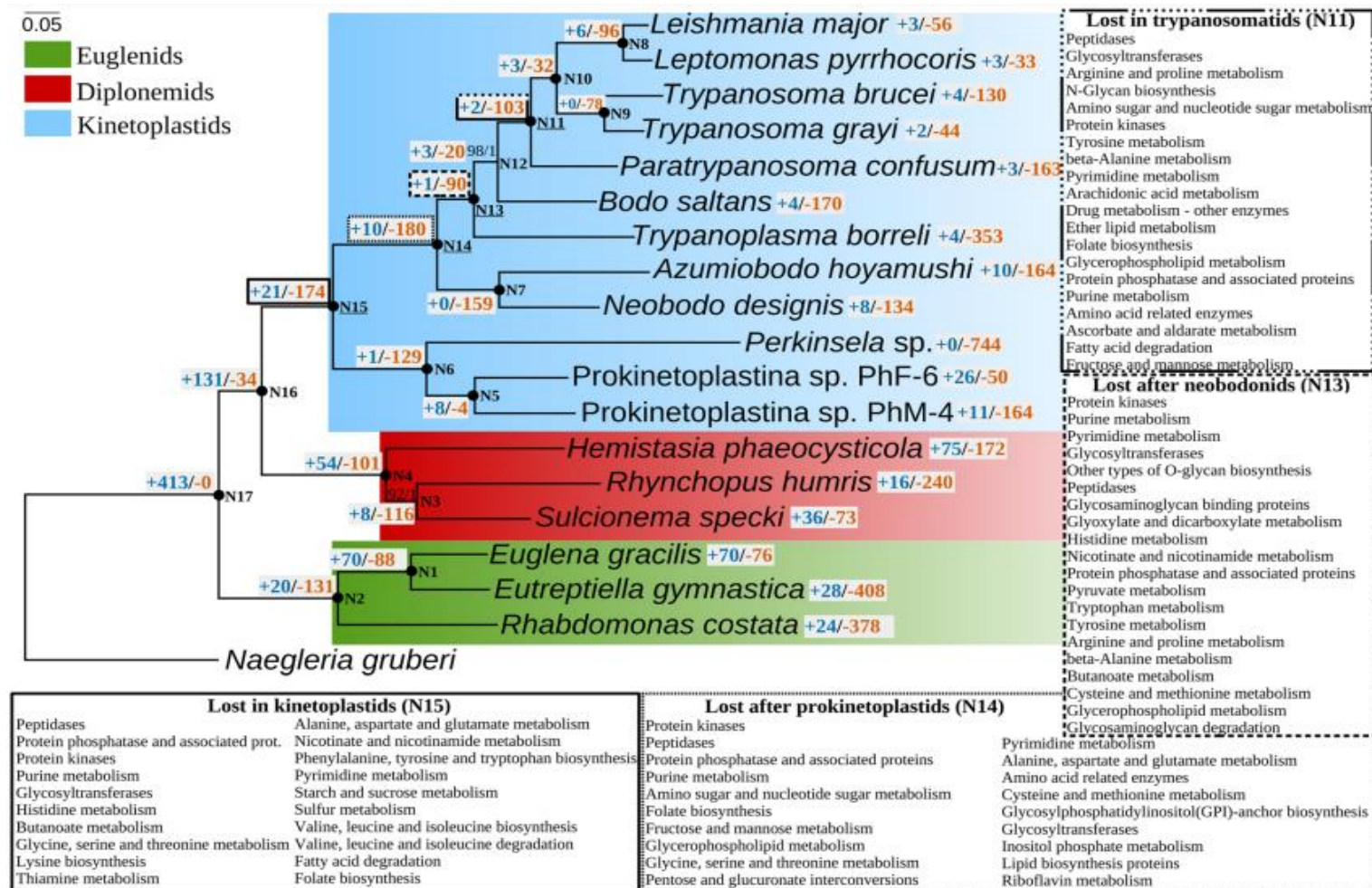
Figure content (phylogenetic tree with node annotations):

- *Leishmania major* +3/-56
- *Leptomonas pyrrhocoris* +3/-33
- *Trypanosoma brucei* +4/-130
- *Trypanosoma grayi* +2/-44
- *Paratrypanosoma confusum* +3/-163
- *Bodo saltans* +4/-170
- *Trypanoplasma borreli* +4/-353
- *Azumiobodo hoyamushi* +10/-164
- *Neobodo designis* +8/-134
- *Perkinsela* sp. +0/-744
- Prokinetoplastina sp. PhF-6 +26/-50
- Prokinetoplastina sp. PhM-4 +11/-164
- *Hemistasia phaeocysticola* +75/-172
- *Rhynchopus humris* +16/-240
- *Sulcionema specki* +36/-73
- *Euglena gracilis* +70/-76
- *Eutreptiella gymnastica* +28/-408
- *Rhabdomonas costata* +24/-378
- *Naegleria gruberi*

Node annotations: N8 +6/-96; N10 +3/-32; N9 +0/-78; N11 +2/-103; N12 +3/-20 (98/1); N13 +1/-90; N14 +10/-180; N15 +21/-174; N7 +0/-159; N6 +1/-129; N5 +8/-4; N4 +54/-101 (92/1); N3 +8/-116; N1 +70/-88; N2 +20/-131; N16 +131/-34; N17 +413/-0

Scale bar: 0.05

Key (top left):
- Euglenids (green)
- Diplonemids (red)
- Kinetoplastids (blue)

Lost in trypanosomatids (N11)
Peptidases
Glycosyltransferases
Arginine and proline metabolism
N-Glycan biosynthesis
Amino sugar and nucleotide sugar metabolism
Protein kinases
Tyrosine metabolism
beta-Alanine metabolism
Pyrimidine metabolism
Arachidonic acid metabolism
Drug metabolism - other enzymes
Ether lipid metabolism
Folate biosynthesis
Glycerophospholipid metabolism
Protein phosphatase and associated proteins
Purine metabolism
Amino acid related enzymes
Ascorbate and aldarate metabolism
Fatty acid degradation
Fructose and mannose metabolism

Lost after neobodonids (N13)
Protein kinases
Purine metabolism
Pyrimidine metabolism
Glycosyltransferases
Other types of O-glycan biosynthesis
Peptidases
Glycosaminoglycan binding proteins
Glyoxylate and dicarboxylate metabolism
Histidine metabolism
Nicotinate and nicotinamide metabolism
Protein phosphatase and associated proteins
Pyruvate metabolism
Tryptophan metabolism
Tyrosine metabolism
Arginine and proline metabolism
beta-Alanine metabolism
Butanoate metabolism
Cysteine and methionine metabolism
Glycerophospholipid metabolism
Glycosaminoglycan degradation

Lost in kinetoplastids (N15)
Peptidases
Protein phosphatase and associated prot.
Protein kinases
Purine metabolism
Glycosyltransferases
Histidine metabolism
Butanoate metabolism
Glycine, serine and threonine metabolism
Lysine biosynthesis
Thiamine metabolism
Alanine, aspartate and glutamate metabolism
Nicotinate and nicotinamide metabolism
Phenylalanine, tyrosine and tryptophan biosynthesis
Pyrimidine metabolism
Starch and sucrose metabolism
Sulfur metabolism
Valine, leucine and isoleucine biosynthesis
Valine, leucine and isoleucine degradation
Fatty acid degradation
Folate biosynthesis

Lost after prokinetoplastids (N14)
Protein kinases
Peptidases
Protein phosphatase and associated proteins
Purine metabolism
Amino sugar and nucleotide sugar metabolism
Folate biosynthesis
Fructose and mannose metabolism
Glycerophospholipid metabolism
Glycine, serine and threonine metabolism
Pentose and glucuronate interconversions
Pyrimidine metabolism
Alanine, aspartate and glutamate metabolism
Amino acid related enzymes
Cysteine and methionine metabolism
Glycosylphosphatidylinositol(GPI)-anchor biosynthesis
Glycosyltransferases
Inositol phosphate metabolism
Lipid biosynthesis proteins
Riboflavin metabolism

Figure 11: Gains and losses of metabolic functions in Euglenozoa based on 20 conserved proteins. Nodes with 100% Bootstrap support and posterior probability of 1.0 marked with black circles. Key top right indicates colour-coding of groups. Scale bars indicate denote number of substitutions per site. Open Access Permissions from the Butenko, (2020), 'Evolution of Metabolic Capabilities and Molecular Features of Diplonemids. Kinetoplastids, and Euglenids' distributed by BMC Biology licences provided via http://creativecommons.org/licenses/by/4.0/. No changes have been made to the original source.

There is a common trend with a reduction of AA biosynthesis in heterotrophs. All kinetoplastids appear to lack biosynthetic mechanisms for lysine, isoleucine, valine, and the aromatic AAs (phenylalanine, histidine, tyrosine, and tryptophan) and are therefore auxotrophic. However, free-living prokinetoplastids, like diplonemids and euglenids, still possess all proteins involved with the shikimate pathway leading to the production of aromatic AAs and as such, diplonemids and euglenids can synthesise all twenty AAs (Butenko, et al., 2020). Similar patterns are repeated when comparing kinetoplastids to diplonemids and euglenids including a lack of enzymes for modular pathways in kinetoplastids that have been recorded in the sister-clades (Butenko, et al., 2020).

Diplonemids have one of the largest observed mitochondrial DNA recorded in an organelle at an estimated 250 mega base-pairs(Mbp) (Lukeš, et al., 2018; Butenko, et al., 2021). Composed of more than 80 covalently closed, non-catenated, 6-7 kilo base-pairs (Kbp) long, circular chromosomes with a small unique region dubbed a 'cassette' (Figure 12). The remainder of the circular chromosome (Kaur, et al., 2020) is comprised of repetitive sequences termed the 'constant region', essentially identical across chromosomes of the same class size. A single gene fragment, known as the 'module', is enclosed within the cassette which are subsequently spliced and reformed to produce mature forms of RNA (Kaur, et al., 2020).
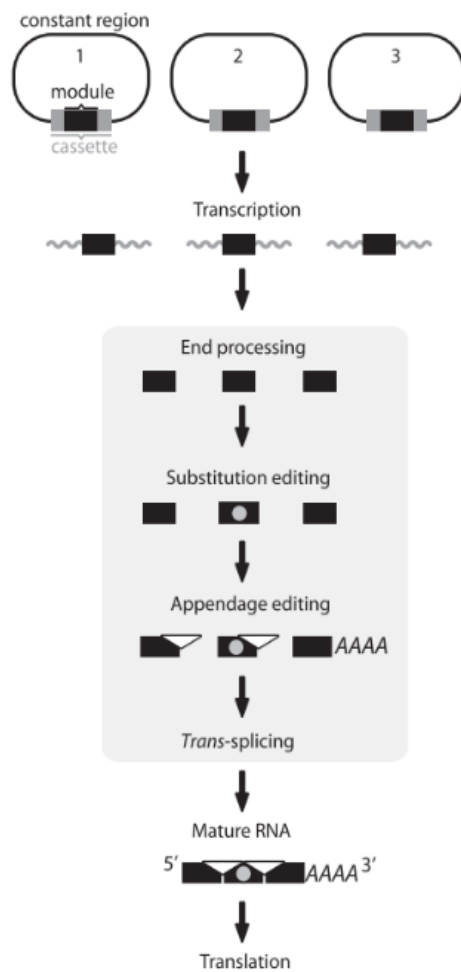
Figure 12: Mitochondrial gene expression editing in diplonemids. Similar mechanisms to the RNA editing occurring within kinetoplastids. This model depicts a module with a unique region, or cassette on chromosomes. The modules are transcribed separately from the promoter from the constant region. The subsequent post-transcriptional processes (indicated within the grey box) occur within diplonemid mitochondrion. The 3' and 5' non-coding regions are removed from the primary transcript and certain modules undergo substitution RNA editing and/or appendage RNA editing (3' nucleotide additions). The module transcript that prefaces the 3' end is poly-adenylated for mRNA and mitochondrial large subunit (mtLSU) rRNA or poly-uridylated in mitochondrial SSU (mtSSU) rRNA. The modules are then trans-spliced together forming mature RNA (mRNA/rRNA). Amended with permissions for use from Kaur (2020), 'Gene fragmentation and RNA editing without borders: eccentric mitochondrial genomes of diplonemids'. Oxford University Press, UK.

Diplonemids have unknown, fragmented mitochondrial DNA (mtDNA) called 'modules' coding for a protein that can be fragmented several times. Cytochrome *c* oxidase subunit 1 (COX1), for example, is fragmented into nine pieces and fragmentation patterns are generally conserved among the diplonemids (at the time of work published by Yabuki, et al., (2016)). Kaur, et al., (2020) revised this finding, uncovering putative 'mini-modules' (>2bp), although these cannot be unambiguously distinguished from appendage RNA editing in DNA-RNA inspection alone; embedded within other, larger, modules which indicate multiple gene transcriptions within a single module as observed within hemistasiids. It is thought, in part, that this ensures correct trans-splicing of the sequence, ensuring specificity.

RNA editing is a means to review an organism's transcripts, producing functional proteins from otherwise non-functioning fragments. In hemistasiids, as with the diplonemids, there are two forms of editing detected within the mitochondrion.. One is G-to-A substitution and the other, A-appendage to internal modules.

A-appendage editing is crucial for the maturation of the COX3 transcript in dinoflagellate organisms. In trypanosomatids, the 3' tails are a mixture of A+U and are generated in two steps involving U-insertion/deletion followed by the addition to the 3' A-tail. After this, the A-tail is elongated to 200-300 A+U heteropolymers through synthesis via the kinetoplast poly(A) polymerase I, marking the transcript for translation (Kaur, et al., 2020). RNA-editing terminal uridylyl transferase 1 form a complex with 3' exonucleases which are involved with the formation of mRNA 3' tails and possibly for rRNA or gRNA uridylation.

1.6 'Meta'omics', Metabarcoding and the Capabilities to Effectively and Accurately Identify Protist Colonies

Dependant on the source material, two approaches can be taken to DNA or RNA sequencing: direct on individual taxa (i.e. from an axenic culture or isolated single cell) or sequencing the entire DNA pool of a complex community (Burki, et al., 2021). Over the last decade, with the establishment of new tools both eukaryotic and prokaryotic microbes in surface and deep-sea waters have been revealed thereby building upon the already expansive eToL.

As previously discussed, phylogenetic studies rely on being able to identify taxa to place them unanimously in the correct position on a phylogenetic ToL. Here I discuss how this leads to the creation of new datasets (or build-upon pre-existing datasets) by measuring expression of an organism's genes in various conditions and provide information regarding gene regulation and infer the functions of previously unannotated genes. DNA metabarcoding (principal focus towards taxonomy of species in a sample), metagenomics (characterization of the genomes—DNA, present in an environmental sample), and metatranscriptomics (characterization of transcripts—RNA, present in an environmental sample) are three separate branches of biology which use either DNA/RNA to identify microbial organisms.

Metatranscriptomics

Metatranscriptomics is the study of gene expression within a collection of microbes (mRNA transcripts) of environmental samples which is used to study taxonomic and functional diversity of bacteria and archaea (Santofererra, et al., 2020; Burki, et al., 2021; Scott, et al., 2021; Obiol, et al., 2020), but more recently are seeing use with eukaryotic taxa. Metatranscriptomics can be coupled with transcriptomics, which identifies single-taxa functional transcripts and pathways under certain stresses as studied in species of *Arabidopsis* (Lowe, et al., 2017). Metatranscriptomics have multiple techniques associated with the acquisition and expansion of mRNA, from environmental communities. These include complementary DNA (cDNA) micro-arrays, oligonucleotide arrays, stable isotope probing (SIP), and multi-dimensional scaling (MDS) (Aguiar-Pulido, et al., 2016; Quackenbush, 2001). Using high-throughput, NGS to capture all available sequences. Microarrays require priori knowledge of the organism of interest and measure the abundance of a defined set of transcripts via the hybridization to an array of complementary probes (oligomers) (Hu and

Polyak, 2006). High-density arrays were used until the late 2000s covering large volumes of gene sets in models. In this time, it has since been succeeded by fluorescence techniques, correlating fluorescence intensity with sensitivity and measurement accuracy for low abundance transcripts (Lowe, et al., 2017). Serial analysis of gene expression (SAGE), and cap analysis of gene expression (CAGE) succeeded array sequencing removing barriers of priori knowledge (Hu & Polyak, 2006). SAGE uses cDNA which is then cut into 11bp oligonucleotides which can then be used for low-throughput sequencing methods creating a reference sequence. CAGE methods use tags from the 5' end of an mRNA transcript. Therefore, the transcriptional start site of genes can be identified when the tags are aligned against a reference genome (Lowe, et al., 2017).

RNA-Sequencing (RNA Seq.) refers to the sequencing of transcript cDNAs, influenced by the development of high-throughput, NGS technologies. Single, short, fragmented nucleotide sequences generated from an RNA transcript produce an expressed sequence tag (EST). >45 million ESTs from >1,400 different species of eukaryotes which are used to either complement existing genome projects or as a low-cost alternative for gene discovery (Parkinson & Blaxter, 2009). One such technique of RNA Seq. is NanoPore sequencing (see materials and methods chapter) that can detect modified bases otherwise masked when sequencing cDNA and eliminates amplification steps that contribute to bias (Lowe, et al., 2017). RNA-Seq. link sequence abundance expression patterns by aligning transcript sequences to a reference genome or *de novo* aligned to one another if no reference is available.

Metagenomics

Metagenomics, like metatranscriptomics, employ NGS approaches, using micro-array technologies studying genomic (gDNA) from mixed communities of organisms within environmental samples but do not require prior PCR amplification or cell cultivation. One such technique that can be used is the 'shotgun' approach which is the untargeted sequencing of all microbial genomes present in an environmental sample (Quince, et al., 2017). Illumina platforms are predominant with shotgun metagenomics owing to its wide availability, high outputs (≤1.5 tera bases (Tb) per run), and high accuracy (error rate of 0.1-1%) but are limited to the potential experimentation biases and complexity of computational analyses and their interpretations.

Targeted approaches used in other methods, use the 18S/ITS short hypervariable regions of conserved genes, outlined previously within this thesis, amplicons for sequencing. Universal primers for V4 and V9 regions of the 18S rRNA are not truly universal and each variable region will be more efficient for some eukaryotic clades over others (Hadziavdic, et al., 2014; Flegontova, et al., 2018; d'Avila-Levy, et al., 2015). V4 primers work poorly for excavates, due to high variability in length across major eukaryotic clades (Flegontova, et al., 2018) but will be more effective for other clades. For example, COX1 for animals; two large subunits (LSUs) of chloroplast rubisco and maturase K for plants; 16S for bacteria, and ITS1 for fungi (d'Avila-Levy, et al., 2015).

Metabarcoding

Metabarcoding is a technique that identifies taxa in an environmental sample directly using one or multiple DNA/RNA (environmental: eDNA/eRNA) markers without the need for microscopic observation or cultivation and had aided in diversifying eToL lineages (Burki, et al., 2021). eDNA can even identify species with seemingly non-distinguishable morphology, or extinct, preserved DNA in sediments (Burki, et al., 2021). This technique uses the accumulation of datasets and 'barcodes' taxa allowing for rapid identification. Advances in NGS have made this technique more viable, revealing the otherwise restricted distribution of lineages in a community.

Metabarcoding is not without its limitations, however. Variability of multi-copy rDNA operons between different taxa and cell size means it can only provide semiquantitative information after correction which requires priori knowledge. Also, errors in DNA amplification or sequencing amongst other practical errors can contribute to an overestimation of diversity. Identifiers for metabarcoding differ of each taxon, previously discussed in this subsection with regards to universal primers. Protists tend to use the 18S rRNA V4/V9 regions, but can use ITS amongst other variable regions, generating OTUs (Maslov, et al., 2012) referring to groups of closely-related taxa, clustering these together based on similarity thresholds, often 97-99% for protists (Obiol, et al., 2020; Burki, et al., 2021). One observation using metabarcoding is that communities are nearly always composed of a small number of abundant lineages accompanied by many rare ones. However, these rare lineages suggest additional processes relating to large population sizes and high dispersal abilities of microbes (Santoferrara, et al., 2020).

Flegontova (2020) used metabarcoding techniques on the V8 18S rDNA to calculate eukaryotic diversity. They found that 70% of the kinetoplastid OTUs and 98% of reads from the Tata Oceans planktonic samples belonged to the Neobodonida group which indicate that kinetoplastids, amongst other protists, were diverse across oceanic samples. Following up with 57 marine planktonic samples, from three different oceanic regions, revealed that three kinetoplastid taxa dominated their findings (Figure 13) these being *Neobodo*, an *unknown Metakinetoplastid*, and *Rhynchomonas* (Flegontova, et al., 2020).

Coupling both metabarcoding and microscopy techniques, such as FISH and electron microscopy (EM), provide a deeper knowledge of interactions within communities. Typically, the reads from metabarcoding are limited to phylogenetic information, complicating taxonomic identification which is overcome by pairwise similarity searches against a reference database. However, by using high-throughput methods, such as PacBio sequencing, longer read sequences can be fully transcribed from an environment including the 18S and 28S rRNA gene (Santoferrara, et al., 2020). Currently, both metatranscriptomics and metagenomics are limited by the scarcity of reference protistan genomes and transcriptomes which has led to the advancement in metabarcoding (Santoferrara, et al., 2020; Massana, et al., 2020).

taxonomic group	Arctic relative abundance	Arctic relative richness	Adriatic relative abundance	Adriatic relative richness	Cariaco relative abundance	Cariaco relative richness	Tara relative abundance	Tara relative richness	combined dataset relative abundance	combined dataset relative richness
Alveolata	36.71%	19.63%	17.75%	18.63%	33.57%	18.39%	12.15%	14.35%	13.26%	14.48%
Amoebozoa	0.09%	0.33%	0.00%	0.11%	0.01%	0.23%	0.25%	0.18%	0.24%	0.18%
Ancyromonadida	0.06%	0.09%	0.00%	0.04%	0.00%	0.08%	0.00%	0.01%	0.00%	0.01%
Archaeplastida	0.31%	2.21%	0.36%	1.55%	0.12%	1.54%	0.63%	0.79%	0.61%	0.81%
Cryptista	0.15%	0.51%	0.80%	0.65%	0.17%	0.58%	0.38%	0.25%	0.38%	0.26%
Diplonemea	0.94%	7.36%	3.63%	5.75%	1.08%	3.17%	2.36%	17.52%	2.32%	17.34%
Haptista	1.23%	1.61%	1.85%	2.95%	0.40%	1.57%	0.64%	0.43%	0.68%	0.45%
Kinetoplastea	0.03%	0.39%	0.05%	0.27%	0.01%	0.24%	0.15%	0.14%	0.14%	0.14%
Obazoa	7.05%	5.04%	29.14%	4.55%	9.05%	4.35%	38.03%	11.41%	36.59%	11.27%
Picozoa	0.51%	0.35%	0.12%	0.16%	0.03%	0.13%	0.24%	0.14%	0.25%	0.14%
Rhizaria	11.22%	4.68%	9.61%	4.45%	29.31%	7.37%	16.19%	6.38%	15.96%	6.34%
Stramenopiles	7.58%	9.87%	3.56%	7.60%	3.05%	8.35%	5.21%	4.01%	5.26%	4.15%
Telonemia	0.99%	0.48%	1.06%	0.65%	0.31%	0.48%	0.18%	0.10%	0.22%	0.11%
unknown Eukaryota	31.74%	34.00%	29.20%	34.57%	21.19%	36.75%	13.38%	12.01%	14.35%	12.55%
other eu- and prokaryotes	1.41%	13.46%	2.86%	18.08%	1.70%	16.77%	10.22%	32.29%	9.74%	31.77%
Eupelagonemidae	99.69%	96.63%	98.19%	93.77%	99.88%	98.65%	99.61%	98.81%	99.58%	98.81%
DSPDII	0.30%	2.90%	1.24%	5.86%	0.00%	0.53%	0.37%	1.11%	0.39%	1.10%
Hemistasia	0.00%	0.07%	0.00%	0.00%	0.00%	0.26%	0.00%	0.00%	0.00%	0.00%
Diplonema	0.01%	0.34%	0.57%	0.37%	0.11%	0.53%	0.00%	0.02%	0.01%	0.03%
Flectonema	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Rhynchopus	0.00%	0.07%	0.00%	0.00%	0.00%	0.00%	0.02%	0.05%	0.02%	0.05%
unknown Prokinetoplastina	0.04%	3.85%	0.00%	0.00%	0.20%	3.45%	0.09%	8.65%	0.09%	8.46%
Ichthyobodo	0.00%	0.00%	0.03%	2.70%	0.00%	0.00%	0.00%	0.38%	0.00%	0.37%
Perkinsela	0.02%	1.28%	0.00%	0.00%	1.41%	3.45%	0.10%	4.32%	0.10%	4.23%
unknown Metakinetoplastina	36.95%	61.54%	85.12%	64.86%	93.17%	72.41%	26.43%	46.05%	26.84%	46.32%
Azumiobodo	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.04%	1.32%	0.04%	1.29%
Dimastigella	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.19%	0.00%	0.18%
Neobodo	23.44%	15.38%	1.42%	10.81%	0.00%	0.00%	54.91%	18.80%	54.40%	18.75%
Rhynchobodo	0.02%	2.56%	0.00%	0.00%	0.00%	0.00%	0.04%	2.63%	0.04%	2.57%
Rhynchomonas	39.06%	6.41%	12.95%	16.22%	4.42%	10.34%	16.64%	6.95%	16.77%	6.80%
Parabodo	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.75%	0.00%	0.74%
Procryptobia	0.11%	1.28%	0.00%	0.00%	0.00%	0.00%	0.03%	0.56%	0.03%	0.55%
Eubodonida	0.04%	3.85%	0.00%	0.00%	0.80%	10.34%	1.67%	6.02%	1.65%	5.88%
Trypanosomatida	0.32%	3.85%	0.48%	5.41%	0.00%	0.00%	0.03%	3.38%	0.04%	3.86%
	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

Figure 13: Breakdown of the relative abundance (number of individuals per species) of each sample, and the relative richness (number of species in an area) broken down into three sections. Main eukaryotic lineages observed in the top section of the table. The middle section is the breakdown of the diplonemids whilst the bottom shows the breakdown of kinetoplastids. Adapted with permission from Olga Flegontova et al., John Wiley & Sons, 2020. Vol. 22 Issue 9, page 18.

1.7 Different modes of motility in flagellated protists and actin-related processes

Many organisms rely on active, directional cell movement for migration, cell feeding or reproduction (Viswanadha, et al., 2017). Protists have multiple methods of traversing their environments, the main two however are swimming through liquid (mediated by flagella or cilia beating) and crawling across/between solid surfaces (using pseudopodia). Each of which have a role of being able to support the organism's movement through their environments. (Fritz-Laylin, et al., 2017a; Pollard, et al., 2008; Laybourn-Perry, et al., 2019; Prostak, et al., 2021).

Flagellated motility

The flagellum is a microtubule-based organelle stemmed from a blepharoplast, an organelle formed from a centriole that serves as the nucleation site for the growth of axoneme microtubules and is the site for new flagellum formation. In kinetoplastids, this is the site where non 'free-living' kinetoplasts adjoin. Flagella are capable of several other functions outside of cell motility, such as feeding and sensing (Figure 14). Structurally the axoneme of the flagellum is comprised of an array of microtubules arranged in a '9+2' configuration; nine microtubule pairs surrounding a single, central pair (Fritz-Laylin, et al., 2017a; Hammond, et al., 2021).

Intriguingly, some cilia display focal adhesion, a structural complex that links intracellular actin bundles and an extracellular substrate together (Fritz-Laylin, et al., 2017b) and other forms of non-motile cilia include rotating nodal cilia (Figure 15, rotary), which have dynein arms but lack both a central pair and radial spokes. This limits the type of ciliary beat possible, with the purpose of the organelle to move fluid in order to generate left-right asymmetry across, for example, a developing animal embryo. Most animal cell types are also able to assemble an immotile primary cilium, which has the function of acting as an antenna to detect environmental cues and thus to initial intracellular signal transduction cascades (Pollard, et al., 2008).

Figure 14: Examples of eukaryotic behaviours/functions and associated structures and function. Some structures have multiple functions dependent upon the species studied and life cycle stage depicting a few examples of modes of motility between microbial eukaryotes. Amended with permissions from Keeling (2019), 'Combining morphology, behaviour, and genomics to understand the evolution and ecology of microbial eukaryotes', Published by the Royal Society.

Both flagellum and cilia (the two terms refer to operationally the same organelle) are assembled and maintained by intraflagellar transport (IFT), an axoneme-associated, bidirectional transport mechanism that is essential for the assembly, maintenance, and length control of cilia and well represented across eukaryotes. This IFT complex consists of three primary components: IFT-A/B, and a Bardet-Biedl Syndrome (BBSome) complex which aids with flagellar signalling molecules, transporting proteins, and can be exploited for gliding motility (Hammond, et al., 2021).
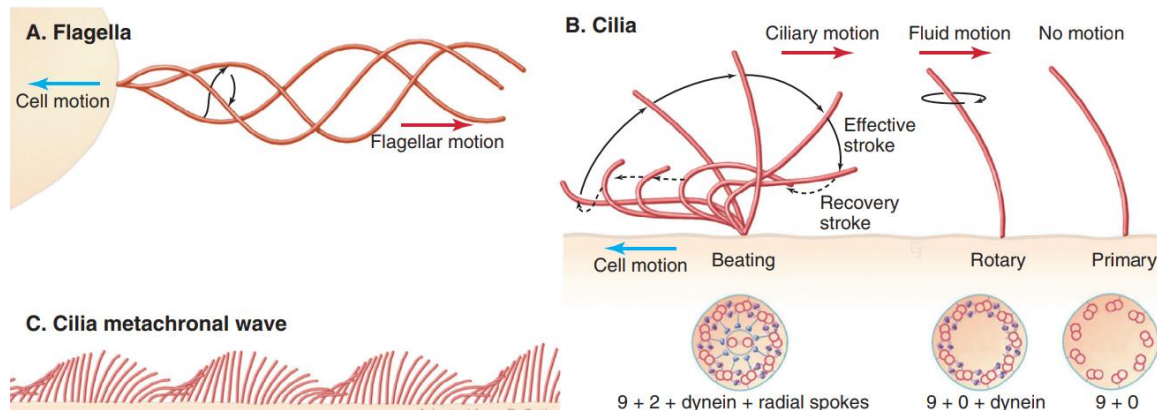
Figure 15: Beating patterns of cilia and flagella. A) one of the modes of flagellar motion generated by flagellated cells. Whipping motion causes the cell to be propelled forward. B) cilia cell motion dependant on the type of axoneme configuration which gives different potential beating patterns. Beating cilium allow for cell movement whilst rotary allow the dispersion of fluids to move around the cell. Primary cilia are capable of being used as antennae, coordinating cellular signalling processes. C) coordinated beating of epithelial cilia. Reproduced with permissions from Pollard (2008), Ch. 38 Cellular Motility sec. 9 which has been modified from a drawing by P. Satir from the Albert Einstein College of Medicine, NY.

Pseudopodia formation genes

In contrast, with regards to cell crawling, pseudopodia have two major types of motility functions described as contraction-hydraulic type and two-way flow type. 'Contraction-hydraulic' motility provides an area in the cytosol (ectoplasm) which is denser than the endoplasm (granular). The ectoplasm contracts on the endoplasm that causes forward momentum from the posterior resulting in a bulge (pseudopodium) and conversion happens at these sites to reform the endo/ectoplasmic concentrations within the cell (Kaneshiro, 1995). The two-way method is simplified to cytoplasmic flow, mediated by the assembly of polymerized actin networks or detaching the cytoplasm from the cytoskeleton in opposite directions on opposite sides of the pseudopodium generating momentum via polarity (Pollard, et al., 2008; Fritz-Laylin, et al., 2017b; Prostak, et al., 2021). For pseudopod movement actin and accessory proteins are essential to propel the cell forward. Adenosine diphosphate (ADF)/cofilins, associated with reorganizing the actin cytoskeleton, bind and sever aged ADF-actin filaments away from the leading edge allowing the branched networks to be assembled and disassembled in seconds and thud propel the cell forward (Pollard, et al., 2008).

As previously mentioned in this MSc, pseudopod formation is regulated byArp2/3, (WASP), and SCAR/WAVE proteins. Arp2/3 complex, actin, formins, and their regulators (WASP & SCAR/WAVE) are highly conserved across eukaryotes that exhibit pseudopod-based motility but have not been observed or recorded in all eukaryotic taxa (Velle, et al., 2020; Prostak, et al., 2021). Fritz-Laylin (2017b) studies noted that plants and multicellular fungi not known to form pseudopod also lacked SCAR/WAVE genes. However, zoospores from fungi were observed to crawl across surfaces using actin-filled Arp2/3-dependant pseudopods indicating this was likely lost at some point during multicellularity.

WASP and SCAR/WAVE are widely conserved Arp2/3 complex activators that respond to different signal cascades. SCAR/WAVE (which form a larger regulatory complex of four other proteins: PIR121, Nap1, abi, and HSPC300) drives actin polymerisation and plays a major role in the formation of protrusions used for cell motility (Davidson & Insall, 2013; Pollitt & Insall, 2009) whilst the WASP interaction is less clear outside of activation of Arp2/3 complexes. However, cells lacking WASP and SCAR/WAVE also lost either or both Arp2/3 complexes (Fritz-Laylin, et al., 2017b; Prostak, et al., 2021).

Arp2/3 complexes organise filaments into branched networks and are coupled and regulated by adenosine triphosphate (ATP) and class 1 or class 2 nucleation-promotion factors (NPFs); class 1 promote a conformational change activating the Arp2/3 complex and enable branching of a new filament whilst class 2 bind to the Arp2/3 complex and actin filaments stabilizing the Arp2/3-mediated branches, and actin (Goley & Welch, 2006).

Actin is one of the most highly expressed proteins in eukaryotic cells with clear roles in motility and phagocytosis (Velle, et al., 2020) and are hypothesised to have evolved before the origin of eukaryotes. They are initiated by the Arp2/3 complex, assembling branched actin networks, and formins which nucleate and elongate linear networks. Actin-mediated cell motility commonly use the same four mechanisms: Contraction of actin-myosin networks, microtubule motor movement, reverse assembly of actin filaments or microtubules (Pollard, et al., 2008).

Multiple modes of motility

Although motile cells predominantly only have one form of functional appendage for propulsion through their environments, this is not always the case. The amoeba, *Naegleria fowleri*, has two observed motile forms. The genome encodes >20 actins, a likely indication of the importance of actin to amoeboid life. Transcription levels of most actin nucleators and NPFs were elevated in amoeba relative to flagellates (Velle, et al., 2020). However, one of the two Arp2/3 paralogs, one of the two WAVE, two WASH (an ortholog to WASP/SCAR) complex subunits, and six formin-family proteins were either expressed at similar levels in both life stage morphologies or preferentially expressed in flagellates (Velle, et al., 2020; Ryder, et al., 2013). I explore these properties in depth in the results section on a wide variety of other taxa.

Velle (2020) performed experiments on inhibiting actin nucleation pathways in *Naegleria* to observe the impact of stressors on cells. The results of using CK-666 to inhibit Arp2/3 complex pathways and SMIFH2 to inhibit formin activity (and some myosins) showed that neither inhibitor caused global disruptions in F-actin, a key cytoskeletal component in dendritic filopodia (Kim, 2009), and SMIFH2 did not reliably correlate with any morphological phenotypes they analysed. However, cells treated with CK-666 presented robust actin cortex and actin-rich spikes that resembled filopodia. Their results suggested that the Arp2/3 complex was important for forming actin puncta and in its absence, cells preferentially build filopodia-like structures that may allocate more actin to the cell cortex (Velle, et al., 2020).

1.8 Cellular responses to environmental stressors

A stressor is defined as an environmental factor affecting cellular homeostasis; examples could include the release of a chemical or biological agent—toxins by humans, or a change in osmotic gradients- either hyper/hypotonic, that, if not corrected, leads to cell impairment or death. Some protists have shown the ability to change morphology under specific stress conditions, for example *Naegleria*, typically a crawling amoeba, can differentiate into a swimming flagellate (Velle, et al., 2020), a process that involves microtubule re-assembly to form a functioning flagellum.

Parasitic protist morphological changes

Differentiation of *Leishmania* from promastigote to amastigote form is induced by temperature and pH changes. Vonlaufen, et al., (2008) showed, by altering axenic culture temperature and pH from 26°C to 37°C and pH7 to pH5.5 respectively and incubating with 5% CO_2 (thereby mimicking their host's internal environments) led to the expression of heat-shock proteins (Hsps) and amastigote-specific genes; also, subsequently seen with parasites including *Giardia lamblia*.

Under confinement, certain species of protists can adapt to their surroundings by means of changing morphology. For example, *Trichomonas vaginalis* can adhere to host epithelial cells by changing from a pear-shaped flagellate to an amoeboid (by altering/rearranging the assembly of new cytoskeletal polymers or actin filaments/microtubules) thereby increasing their contact surface with epithelial cells (Henriquez, et al., 2021). However, specific triggers that induce morphological change can vary, sometimes specific environmental stressor conditions must be met, but also can occur due to protein 'moonlighting' events.

Free-living protists morphological changes

Salpingoeca rosetta has been observed, in confinement experiments (Brunet, et al., 2021), to retract its flagellum and active α-motility, to ultimately escape and overcome confinement. A hypothesis that this ability to differentiate between a single free-living, motile, eukaryotic cell into a crawling, pseudo-amoeboid cell must have evolved from a common ancestor, yet this has not been investigated in depth. *N. fowleri, S. rosetta,* and *T. vaginalis* can alternate between a flagellate *and* an amoeboid form presenting the hypothesis that a protozoan ancestor of animals may already contain the genetic prerequisites for this flagellar-to-amoeboid (and *vice versa*) switch.

Fritz-Laylin (2010), originally stated that protozoans, which can alternate between flagellates and amoeboid forms are far removed from animals on the ToL. However, the close evolutionary relationship between choanoflagellates and animals and comparisons between the two groups at different life stages support the hypothesis that the amoeboid cell types of animals had evolved from ancestral flagellate cells after the establishment of multicellularity (Brunet, et al., 2021). Close outgroups to choanoflagellates and animals that can produce amoeboid cells or alternate between the two forms indicate that this ability to switch predates the divergence of both choanoflagellate and animal lineages linking biochemical studies of animals and protists' cellular structures and conserved molecules (*i.e.*, locomotion-involved cell protrusions) which are consistent with possible pre-metazoan origins (Brunet, et al., 2021).

Amoeboflagellate phenotypes have recently been described in several species in key phylogenetic positions including fungi (Karpov, et al., 2019), opisthokonts (apusomonads (Cavalier-Smith, & Chao, 2010), and breviates (Minge, et al., 2009)). Given that choanoflagellates and metazoans are closely-related sister groups, this implies their machinery are orthologous (*i.e.,* having genes similar in nucleotide sequences that originated from a common ancestor between groups), providing evidence for a pre-metazoan origin of blebbing/pseudopod and crawling mechanisms (Brunet, et al., 2021). However genomic evidence is lacking, in part because proteins involved with crawling motility can also fulfil crawling-independent mechanisms.

Choanoflagellate genomes encode predicted regulators for pseudopod formation (Arp2/3, SCAR/WAVE and WASP) two upstream activators of actin assembly-- although they have not been observed to form blebs. However, these proteins have been proposed to aid in phagocytic cup formation that, structurally, are like pseudopods but involved in feeding not motility (Brunet, et al., 2021; Fritz-Laylin, et al., 2017b).

As previously stated above; in my MSc research I set out to achieve two broad objectives: to characterize novel protists taxa from a series of environmental samples with the aid of confocal microscopy, and Basic Local Alignment Search Tool- Nucleotide (BLASTn) bioinformatics, a nucleotide BLAST tool which compares one or more nucleotide sequences to a reference sequence for a model species; and to look for possible cryptic signatures, namely those involved with pseudopod formation genes, of different modes of cell motility in evolutionarily little-studied eukaryotes. For this objective, I shall be utilising BLASTp bioinformatic analyses, a protein BLAST tool, which identifies readable, transcribable genes capable of producing proteins to aid in the identification of pseudopod genes.

Chapter 2

Material and Methods

2.1 Cell sampling locations and growth medium preparation

The aquatic samples used in this thesis were isolated from the campus lake and surrounding ponds of the east and west campuses at the University of York. HE-L3 and HE-L4 taken from the East campus Lake on campus, HE-R8 taken from the East campus Reed Bed, and HE-DP11, the East campus Deposit Pond (all exact locations can be found in the appendix). The unknown isolate, known in the laboratory as "Dark Stuff", was obtained from a rock pool at Embo Beach in north-east Scotland. Geographical location and reference table for each sample can be found in the appendix of this thesis.

York samples shared the same geographical location, and thus, shared growth media composed of: 50ml filtered York Lake water (thereby establishing an environmental climate) and 0.025g of yeast extract as a source of nutrition.

Dark Stuff was cultured in a synthetic medium containing per 250ml of deionised water, 0.12g of yeast extract and 8.25g of aquarium systems instant ocean sea salt. All growth media were stored, sterile at 5°C.

2.2 Cell culture

Each culture flask was kept as 10ml cultures, grown at room temperature and observed every other day under x10 magnification on a light microscope paying attention to yield (or abundance) of phagotrophic protist cells and bacterial prey. Once a desirable yield was established, each flask of cells were then sub-passaged 1:5 dilution into fresh media. Three flasks were set-up per sub-passage. New cultures were grown for a further three days before biomass was collected by centrifugation for analysis. An additional flask for each culture was maintained by passage at 1:10 dilution until no further protists could be detected by microscopy.

2.3 Genomic DNA extraction, purification, and storage

DNA extraction followed standard protocols for cultured eukaryotic cells. Cells were first collected by centrifugation at 6000 x g for 10 minutes, discarding supernatant until all cells were collected from the sub-passage flasks into a 1.5ml Eppendorf tube.

Cells were then resuspended in 500μl of 1xPBS (phosphate buffered saline) and centrifuged at 6,000 x g for 3 minutes to remove and remaining media. The supernatant was discarded, and the cell pellet was then resuspended in 500μl of genomic DNA (gDNA) extraction buffer which comprised of (50mM of 10% (10mM) hydroxymethyl aminomethane (Tris) to a final pH of 8.5 (made using 60.57g of Tris Base, 250ml of deionized H_2O, and using a strong base, such as sodium hydroxide (NaOH), to monitor the final pH, 100mM NaCl, and 1% (1mM) of ethylenediaminetetraacetic acid (EDTA)). The addition of 50μl 10% sodium dodecyl sulphate (SDS) to the cell-gDNA buffer mixture promoted cell lysis and the Eppendorf tube was gently inverted. 10μl of proteinase K was then added, to remove contaminants from the previous steps.

Eppendorf tubes were then incubated at 56°C for 2 hours after which 500μl of (1:1) phenol-chloroform (pH 7.5) was added and inverted to mix. This separates the cell sample into three phases; an aqueous phase, which contains gDNA; an interphase; and lastly an organic phase that contains waste products, such as: proteins, RNA, and lipids.

Eppendorf tubes were then centrifuged at 10,000 x g for 2 minutes at room temp to separate the aqueous and organic phases. The aqueous phase (approximately 500μl) was transferred into a sterile Eppendorf tube and gDNA precipitated by the addition of 1mL 100% ethanol (EtOH). Precipitated DNA was spooled using a glass pipette loop and washed by dipping six times in 70% EtOH before transfer to an Eppendorf tube containing 50μl of elution buffer (EB) (from the ThermoFisher PCR purification kit, EB consists of 10% (10mM) of Tris-HCl) to which RNAses (ribonucleases) were added in the form of 5μl of 'mini-prep' resuspension buffer containing RNAse.

The gDNA was incubated for 1 hour at 37°C (or 2 hours at room temperature) before being stored at either -20°C (in a diluted aliquot of 1:50μl-- 1μl of gDNA to 49μl of ultrapure water) or 4°C (undiluted).

2.4 PCR

For PCR amplification 12.5µl master mix solution ([ThermoFisher DreamTaq.](#)) consisting of DreamTaq. DNA Polymerase, 4mM $MgCl_2$, 2X DreamTaq. Green buffer, and deoxynucleotide triphosphates (dNTPs) was used. The addition of 1µl of both forward and reverse primers (each at 10µM concentration), 8.5µl of milliQ (ultrapure) water, and 2µl of extracted gDNA (from 1:50 dilution aliquots) was added to the master mix solution and lightly mixed. Thermocycler programs were amended from Carr (2008), a denaturation step at 94°C for 2 minutes, was followed by 30 cycles of 30 second denaturing at 94°C, various annealing temperatures (ranging from 50°C-56°C) for 30 seconds, and a 1-minute elongation at 72°C, with a final elongation step at 72°C for 10 minutes.

2.5 Primers

Universal primers to amplify 28S, Hsp83, M167, and LSU1-4 DNA were used. 28S (specific to eukaryotes) and LSU1-4 (catalyse peptide bond formation) correspond to the large subunits of eukaryotic cytoplasmic ribosomes (Gregory, et al., 2019). Hsp83 is well established in different kinetoplastids including: *Leishmania, T. cruzi*, and *T. bruci* (Folgueira & Requena, 2007). M167 is a gene associated with chloroplast, potentially showing photosynthetic capabilities (e.g. some taxa of euglenids) within our samples (Song, et al., 2019; Yamaguchi et al., 2012). Primer oligonucleotides for LSU and M167 used in this thesis are outlined in Table 4.

Forward Primer	Sequence	Reverse Primer	Sequence
LSUF1	5'-ACC CGC TGA AYT AA GCA TAT-3'	LSUR1	5'-GCT ATC CTG AGG GAA ACT TCG G-3'
LSUF1B	5'-GCG TTC RAA GWB TCG ATG-3'	LSUR2	5'-AGC CAA TCC TTW TCC CGA AGT TAC-3'
LSUF2	5'-CCG AAG TTT CCC TCA GGA TAG C-3'	LSUR3	5'-CCG CCC CAG YCA AAC TCC C-3'
LSUF3	5'-CCG CAK CAG GTC TCC AA-3'	LSUR4	5'-MRG GCT KAA TCT CAR YRG ATC G-3'
LSUF4	5'-GGG AAA GAA GAC CCT GTT GAG-3'		
M167	5'-CGT CTT TTT TTA GGA GGT CT-3'		

Table 4: Universal primer sequences used for amplification of Dark Stuff, HE-DP11, HE-L3, HE-L4, and HE-R8 gDNA. Special nucleotides are notes as follows: Y= pyrimidine, R=purine, W= adenine/thymine, B= guanine/thymine/cytosine, K= guanine/thymine. F regard to forward primers whilst R refers to reverse primers in cells.

2.6 Agarose gel electrophoresis

1% w/v agarose gels were prepared by adding 1g agarose to 100ml 1xTAE (Tris-acetate-EDTA) (Made by diluting 10X UltraPure^{TM} TAE Buffer (ThermoFisher Scientific) and diluting to a final concentration of 1X or 1:10 dilution) buffer before heating in a conventional microwave for two minutes on medium-high heat. Once cooled to handling temperature, 1µl of SYBR safe, to visualize DNA under ultraviolet (UV), was added. The liquid agar was then set in a cast using a well-comb creating wells in the gel. Once set and comb removed, the gel was then transferred to a BioRad electrophoresis chamber, filled to level with 1xTAE buffer. 3µl of each PCR reaction was run by agarose gel electrophoresis alongside 5µl of a 1kb reference ladder (New England Biolabs) for 40 minutes at 100V.

2.7 Preparation of L-agar plates and L-broth

Lysogeny broth (LB) agar plates were prepared as per manufacturer's instructions (Fisher Scientific). This involves the addition of 8g of LB miller powder to 200ml of deionized H_2O. Following autoclaving at 126°C (Prestige Medical) and, upon cooling to ~50°C, ampicillin (50µl/ml) was added to the liquid agar. LB-amp agar was then poured into petri-dishes over an open-flame to ensure sterility, allowed to set and then stored at 4°C. LB-broth was again prepared according to manufacturer's instructions, which involved adding 5g of LB miller powder to 200ml of deionized H_2O, before autoclaving and the addition of ampicillin (50µl/ml).

2.8 pGEM® T-easy cloning of PCR amplicons

PCR amplicons from each environmental sample, were subject to purification using a PCR clean-up kit (ThermoFisher Scientific) which consisted of: binding buffer, concentrated wash buffer, EB (10mM Tris-HCl at a pH of 8.5), and GeneJET Purification Columns, and involved using 1:1 w/v binding buffers, paying attention to pH colour changes and adjusting accordingly with either acidic or basic compounds to produce a yellow colour. Centrifuging at 12,000 x g for 30-60s, discarding the flow-through, then adding 700µl of wash buffer, centrifuging as above, and discarding flow through. Repeating once again without the addition of wash buffer to remove any residual wash buffer.

50µl of EB was added to the purification column and centrifuged, as above, into a microcentrifuge tube collecting the purified DNA and storing at -20°C ready for ligation. The purified DNA was then ligated into pGEM® T-easy vectors (Figure 17, Promega) following the manufacturers' instructions.

Ligation reactions comprised of the addition of 1µl pGEM® T-easy vector to 3µl of the PCR amplicon, 5µl of X2 rapid ligation buffer (a buffer used in junction with T4 ligase for efficient ligation of sticky-ended DNA) (MyBio), 0.5µl of T4 ligase, and 0.5µl of MilliQ water. Ligation reactions were left at room temperature for a minimum of 4h before being stored at 4°C overnight.
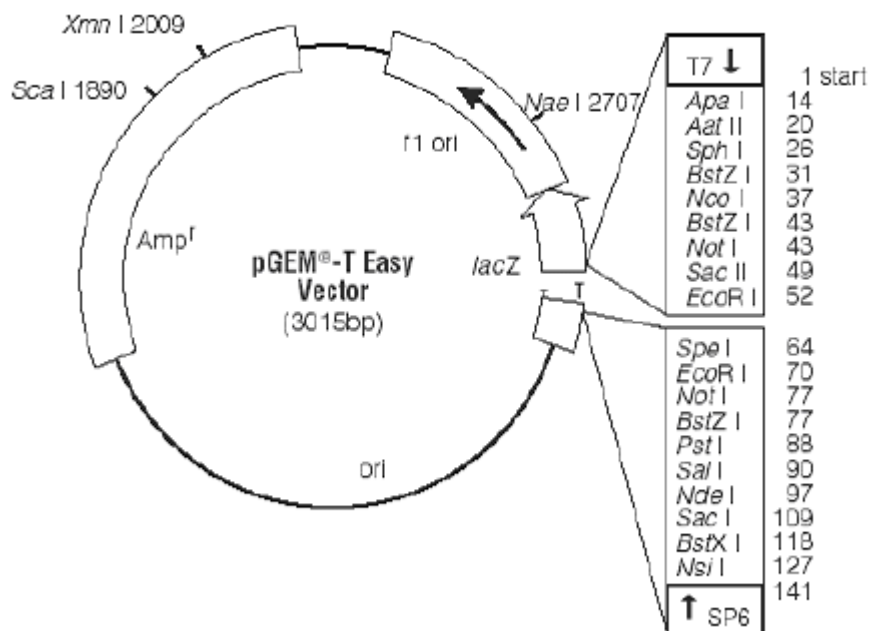
Figure 16: pGEM® T-easy (Promega) vector map. pGEM® T-easy vector map showing the multiple cloning sites and *Eco*RI restriction sites subsequently used to confirm cloning success.

5µl of each ligation reaction was then added to 50µl of XL-1 blue *E. coli* competent cells and incubated on ice for 30 minutes. Cells were heat-shock at 42°C for 1-minute., before being placed on ice for a further 2 minutes. The entire reaction was spread onto LB-ampicillin (LB-amp.) plates and incubated overnight at 37°C. The next day, five colonies from each transformation plate were transferred to 15ml Falcon tubes containing 3ml LB-amp. Falcon tubes were incubated overnight at 37°C, shaking at 180rpm.

2.9 Plasmid mini-prep and restriction digestion

Transformed *E. coli* cells were collected by centrifugation; cultures were transferred to 1.5ml Eppendorf tubes and centrifuged at 8000x g for 2 minutes at room temperature. Supernatant was discarded and plasmid DNA purified using a GeneJET Plasmid Miniprep Kit ([Thermo Scientific](#)) as per manufacturer's instructions. Which involves adding resuspension (250µl), lysis (250µl), and neutralization (350µl) solutions to the pelleted cells and centrifuging at 12,000 x g for 5 minutes. Supernatant is then added to a spin column and centrifuged, as above, for 1 minute. 500µl of the wash solution is added and centrifuged twice, discarding flow-through each time, and spun once again without wash buffer to remove any residual wash buffer. The spin column is then transferred to a fresh Eppendorf tube and incubated for 2 minutes with the addition of 50µl of EB and centrifuged as in previous steps for 2 minutes, collecting the flow through, which contains plasmid DNA.

In order to confirm successful cloning of our PCR inserts into the pGEM® T-easy vector a series of restriction digests were performed. 3µl of plasmid DNA was added to a reaction mix containing 1µl (10µM) *Eco*RI enzyme, 5µl of MilliQ water, and 1µl of 10X digestion buffer. Alternatively, HE-R8 and HE-DP11 samples were digested with 0.5µl (5µM) of both *Nco*I and *Nde*I, again in a total reaction volume of 10µl. Restriction digestions were incubated for 3-4h at 37°C and the presence or absence of insert was confirmed by agarose gel electrophoresis using 1% w/v agarose gels.

To calculate DNA concentrations a Nanodrop spectrophotometer was used. This calculated ng/µl of plasmid DNA in each mini-prep. For sequencing, mini-prep DNA was diluted to a final volume of 15µl at 50-100ng/µl DNA concentration. Samples were sequenced off-site using an M13 reverse primer.

2.10 DNA sequence analysis by reference to online datasets

Returned sequences were manually checked by eye and clipped were required. Searching for the multiple cloning sites flanking restriction sites, *Eco*RI (5'-GAATTC), *Nco*I (5'-CCATGG) and *Nde*I (CATATG) revealed the nucleotide sequence that could be taken forward for BLASTn analysis.

Nucleotide BLASTn (nr/nt) ([NCBI](#)) were carried out using the NCBI datasets to provide likelihood-based statistical data. Measurements for identifying species follow the both S (similarity) and E-value scores returned from the online BLASTn datasets. E-values returned show the probability, due to chance, that another alignment with a similarity greater than the S score. Therefore, a lower E-value corresponds to a lower chance there is another sequence similar to the input query DNA sequence.

2.11 RNA extraction and purification

For Dark Stuff RNA extraction, the mammalian cultured cells protocol was followed from the GeneJET RNA Purification Kit ([Thermo Scientific](#)) as per manufacturer's instructions with cells initially pelleted by centrifugation at 1,500 x g for 5 minutes. This included washing the pelleted cells in excess of 1xPBS, centrifuging as above and removing supernatant, resuspending with 300µl of DTT (DL-Dithiothreitol) treated lysis buffer (20µl of DTT for every 1ml of lysis buffer used or 1:50 ratio) and mixing gently. Adding 360µl of EtOH (>96%) mixed with a pipette and then transferred to a transfer column. This was then centrifuged for 1 minute at 12,000 x g where 700µl of wash buffer 1 was added and centrifuged at 12,000 x g for 1 minute, discarding flow-through. 600µl of wash buffer 2 was then added and centrifuged as above for 1 minute and repeated a second time with a lower volume (250µl) of wash buffer 2. From here it was transferred to a sterile Eppendorf tube, where two rounds of the addition of 100µl of RNAse free H_2O and centrifuged for 12,000 x g. Quantity and quality of RNA was determined using the Nanodorp spectrophotometer before storage at either -80°C as an ethanol precipitate or at -20°C in its 1:50 (1µl to 49µl of ultrapure water) diluted form.

2.12 Cell observational studies using DAPI staining

Cells were pelleted by centrifugation at 1,800rpm for 10 minutes and the supernatant removed. Cells were washed in 1ml 1xPBS, before re-centrifuging at 1,800rpm for 10 minutes. 100µl of 1xPBS was added to resuspend the cell pellet before adding to a hydrophobic pen created well on a microscope slide. In order for cells to adhere to the glass, slides were incubated at room temperature for 1 hour before excess liquid was removed. Cells were then fixed by filling the well with 50µl of paraformaldehyde, which cross-links bonds between molecules, and incubated at room temperature for 10 minutes.

Paraformaldehyde was removed from the well, and slides are transferred to a 100% ice-cold methanol-filled coupling jar stored at -20°C for 10 minutes. In order to rehydrate cells slides were transferred to a second coupling jar containing 1xPBS, incubated for 10 minutes at room temperature.

DAPI-containing (an AT-rich favouring immunofluorescent stain (Lukếs, et al., 2018)) mounting media was used to mount the slides a cover slip before securing in place using nail varnish. Slides were observed under x100 magnification on a Zeiss Axiovert light-microscope in tandem with ZenPro software.

2.13 SARS-Cov-2 disruptions to outstanding work

Due to SARS-CoV-2 disruptions leading to the closure of labs and York sample cell lines dying as a result, the original aims of the thesis were amended to reflect the work that could be done with the surviving samples of Dark Stuff which primarily focussed on protein-protein BLASTp analysis of Arp2/3 complex, WASP, and SCAR/WAVE genes, referred here as pseudopod formation genes, of kinetoplastids and a variety of other taxa from fungi and algae covering both free-living and parasitic lineages.

2.14 Protein BLASTp bioinformatics on *Bodo saltans* and a variety of other lineages

Bioinformatic analysis focusses on proteins involved in pseudopod formation with protein BLASTp from several different databases (Table 6) using query sequences of *Homo sapiens* Arp2, Arp3, and WASP and *S. rosetta* WASP and SCAR used to search for ortholog sequences in 30 different taxa, including different model datasets of species (such as *Galdieria sulphuraria*) to observe for inconsistences between mapped models. Each of the 30 taxa were chosen to cover a wide-range of different species from different groups of the eToL.

Each query sequence was subsequently put through protein BLASTp, which involved inputting the *H. sapiens* and *S. rosetta* AA sequence queries against the reference genomes of each individual taxa and model (Table 6) and returned sequences screened for E-values >1e-04. Similarity thresholds were not put in place, however, close attention to identities, gaps, and positives percentages were considered for hypothetical protein returns that displayed within an acceptable range for protein length and E-values. Hypothetical proteins went through CLC sequence alignment (https://www.qiagen.com/us) against both *H. sapiens* and *S. rosetta* queries to aid in identification of uncertainties between KOGG descriptions and annotated gene references.

Each recorded BLASTp returned upon screening from all *H. sapiens* and *S. rosetta* queries were collected and categorized into; numbers of actin isoforms, formins, and detection of the presence of Arp2, Arp3 (together forming the Arp2/3 complex), WASP, and SCAR/WAVE.

Organism	Database	Website
Bodo saltans	TritrypDB	https://tritrypdb.org/tritrypdb/app
Ectocarpus variabilis	Uniprot	https://www.uniprot.org/blast
Ectocarpus siliculosus		
Cyanidioschyzon merolae	Czam	http://czon.jp/blast/blast.html
Chondus crispus	JGI Phytocosm	https://phycocosm.jgi.doe.gov/phycocosm/home
Galdieria phlegrea		
Galdieria sulphuraria (Azora/MS1/MtSh/SAG21.92/YNP5578.1)		
Gracilariopsis chorda		
Porphyra umbilicalis		
Pyropia yezoensis		
Rozella allomycis		
Micromonas commoda		
Micromonas pusilla		
Chlamydomonas reinhardtii (external model/filtered)		
Cyanophora paradoxa		
Glaucocystis nostochinearum		
Spizellomyces punctatus	FungiDB	https://fungidb.org/fungidb/app
Microsporidium daphniae	MicrosporidiaDB	https://microsporidiadb.org/micro/app
Encephalitozoon cuniculi		
Naegleria grubi	AmoebaDB	https://amoebadb.org/amoeba/app
Gregarina niphandrodes	CryptoDB	https://cryptodb.org/cryptodb/app
Chromera velia		
Vitrella brassicaformis		
Ostreococcus tauri	NCBI	https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastHome
Ostreococcus lucimarinus		

Table 6: Databases used in protein BLASTp work on homolog proteins for *H. sapiens* and *S. rosetta* queries.

Chapter 3

Results

3.1 PCR of four University of York isolates and Dark Stuff using universal primers

Flagellated cells were successfully cultured from four samples collected from three University of York sites HE-L3, and HE-L4, HE-R8, and HE-DP11, and our unknown isolate, Dark Stuff. Genomic DNA of sufficient quality was generated to enable PCR using sets of universal primers to amplify 28S, Hsp83, LSU1, LSU2, LSU3, LSU4 and M167 (Figure 17). Products were successfully amplified using primer sets LSU1 to LSU4 in all four samples. 28S was successfully amplified in both HE-R8 and DP11 (indicated with a white arrow), and Hsp83 in HE-L4.
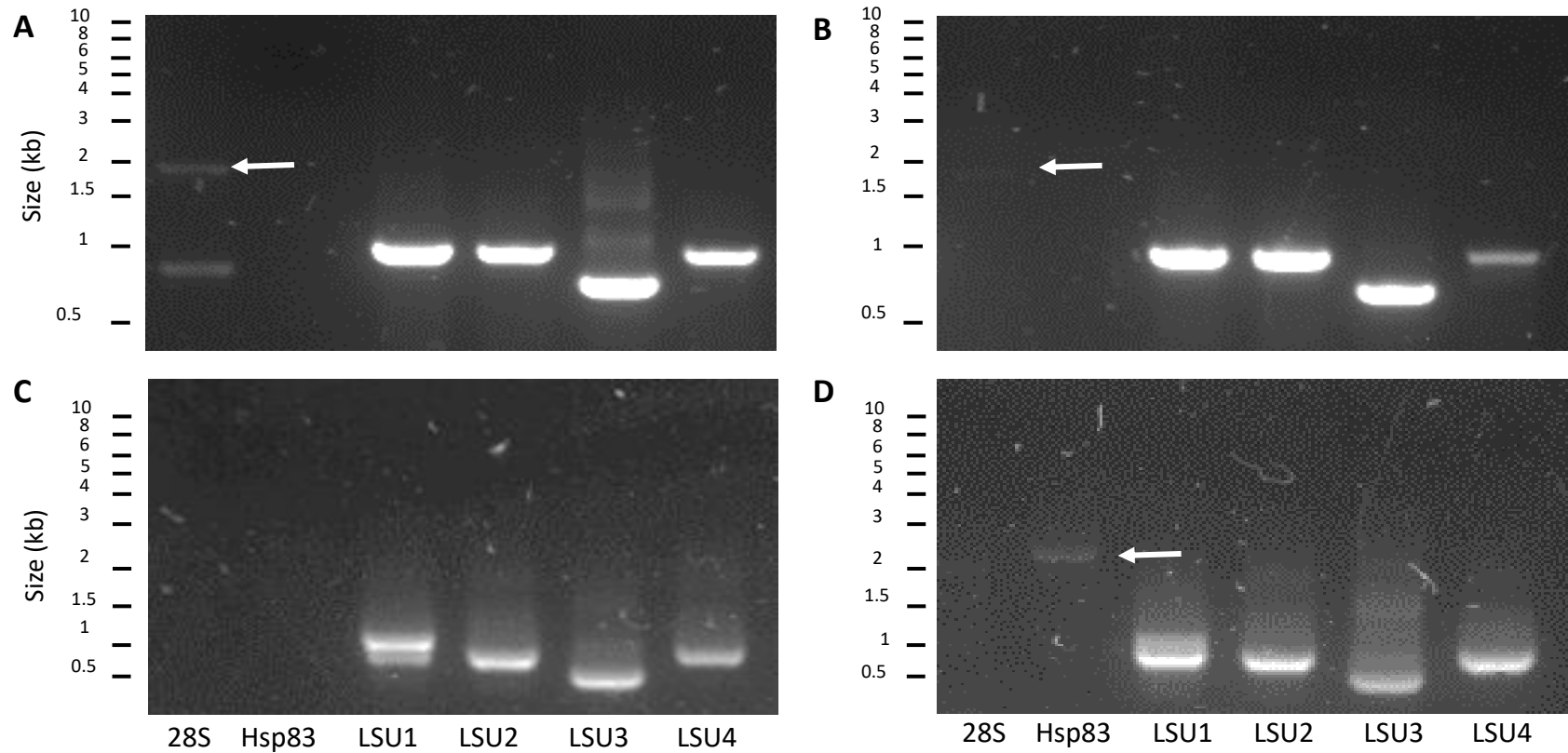
Figure 17. PCR amplification of four University of York isolates using universal primers. PCR products following genomic DNA isolation from four samples originating from three University of York sites (A, HE-R8; B, HE-DP11; C, HE-L3; D, HE-L4) and amplified using universal primer sets for 28S, Hsp83, LSU1, LSU2, LSU3 and LSU4. White arrows indicate the successful amplification of 28S in HE-R8 and HE-DP11 and Hsp83 in HE-L4 respectively.

Following PCR using genomic DNA isolated from University of York samples, genomic DNA from our unknown sample, Dark Stuff, was also isolated and subjected to PCR using primer sets for 28S, Hsp83, M167/8 and LSU1 through to LSU4 (Figure 18). Products were successfully amplified using primer sets LSU1, LSU3 and 28S.

In order to generate DNA of sufficient quality and quantity for sequencing, PCR amplicons were transformed into pGEM® T-easy. A summary of those PCR reactions taken forward for pGEM® T-easy cloning is given in Table 7.
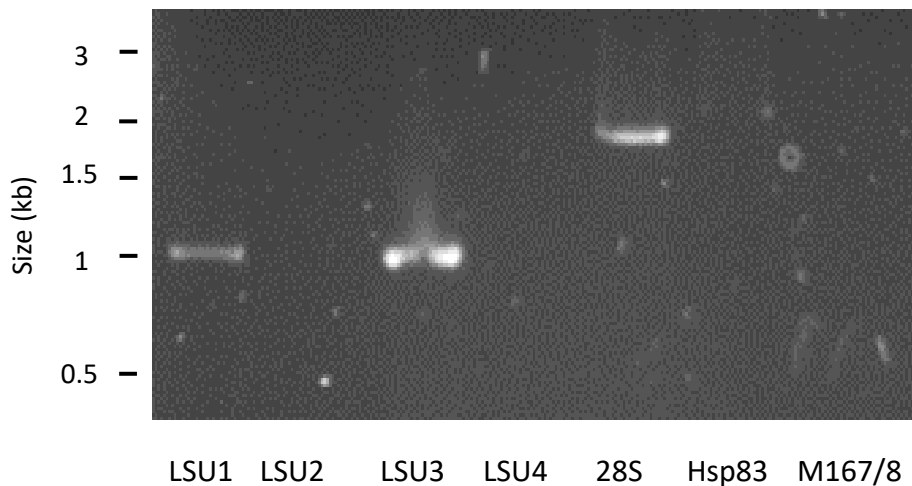
Figure 18. PCR amplification of Dark Stuff using universal primers. PCR products following genomic DNA isolated from Dark Stuff and amplified using universal primer sets for LSU1, LSU2, LSU3, LSU4, 28S, Hsp83 and M167/8.

Sample	Universal Primer Set						
	28S	Hsp83	M167	LSU1	LSU2	LSU3	LSU4
Dark Stuff	✓	✓		✓		✓	
HE-R8	✓			✓	✓	✓	✓
HE-DP11	✓			✓	✓	✓	✓
HE-L3				✓	✓	✓	✓
HE-L4		✓		✓	✓		

Table 7. A summary of Dark Stuff, HE-R8, HE-DP11, HE-L3 and HE-L4 PCR amplicons using sets of universal primers that were subsequently cloned into pGEM® T-easy. Successful PCR reactions with each universal primer set for Dark Stuff and isolates collected from The University of York are indicated with a tick; attempts were subsequently made to clone these into pGEM® T-easy.

3.2 Plasmid cloning and PCR restriction digests

Following cloning into pGEM® T-easy for each of the PCR products indicated in Table 7, successful insertion into the vector was confirmed using a series of *Eco*RI restriction digests (Figures 19-23). Following *Eco*RI digests of HE-R8 LSU2, HE-DP11 LSU2 and LSU3 and HE-L4 HSP83 plasmids, products are of a smaller than expected size when compared to their respective PCR product. This could be indicative of a cloning mishap or the presence of internal *Eco*RI site(s), both of which can be confirmed by subsequent sequence analysis. Table 8 indicate which pGEM® T-easy PCR products were correctly inserted and sent offsite for sequencing.
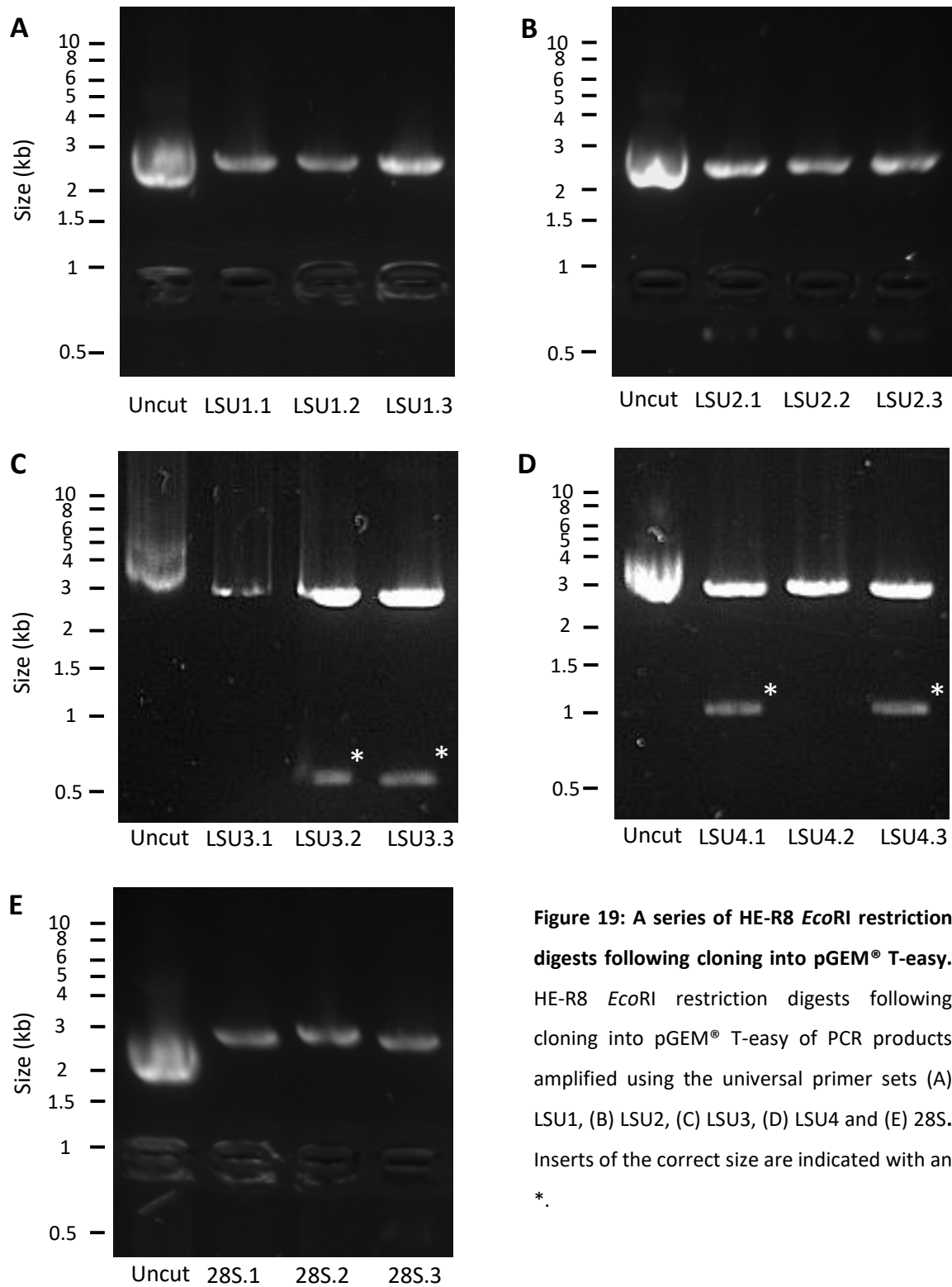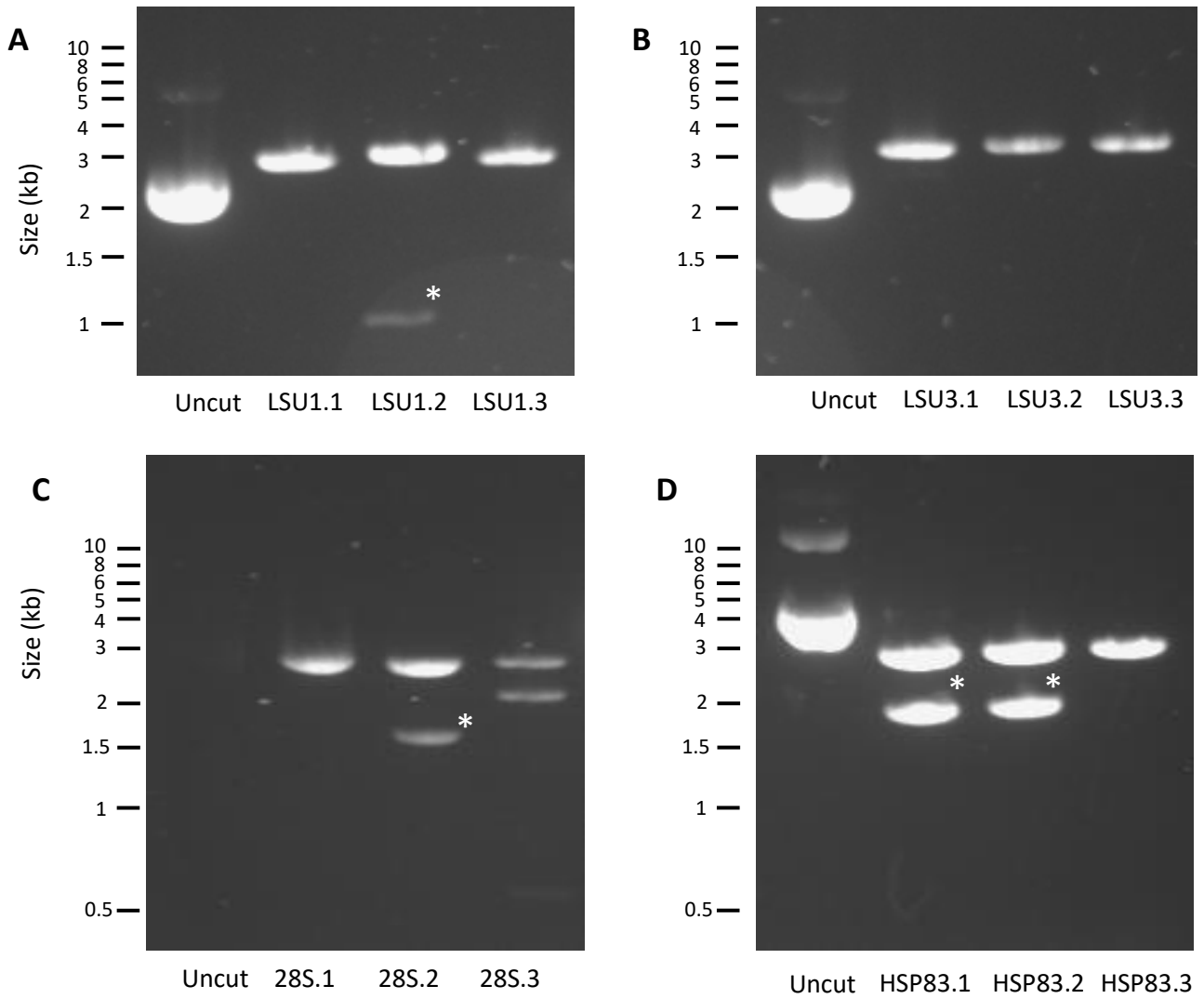
Figure 19: A series of HE-R8 *Eco*RI restriction digests following cloning into pGEM® T-easy. HE-R8 *Eco*RI restriction digests following cloning into pGEM® T-easy of PCR products amplified using the universal primer sets (A) LSU1, (B) LSU2, (C) LSU3, (D) LSU4 and (E) 28S. Inserts of the correct size are indicated with an *.
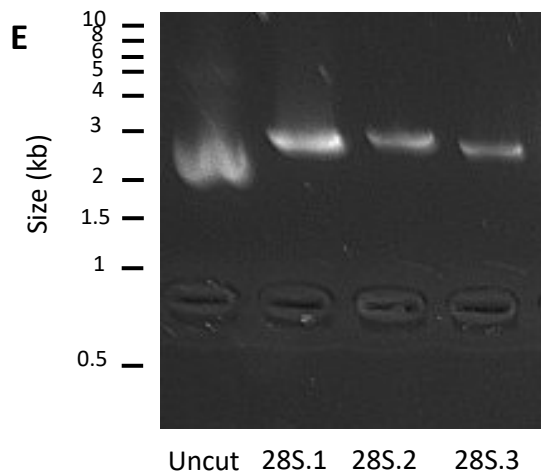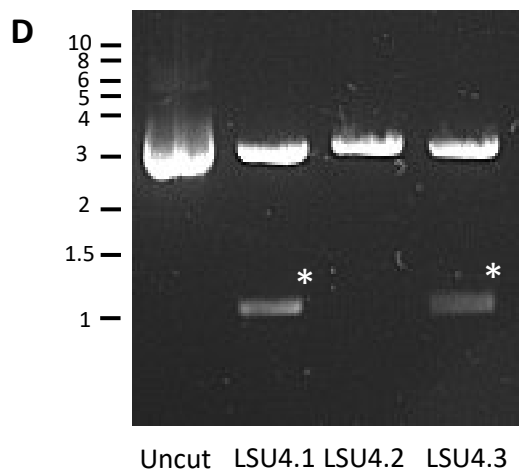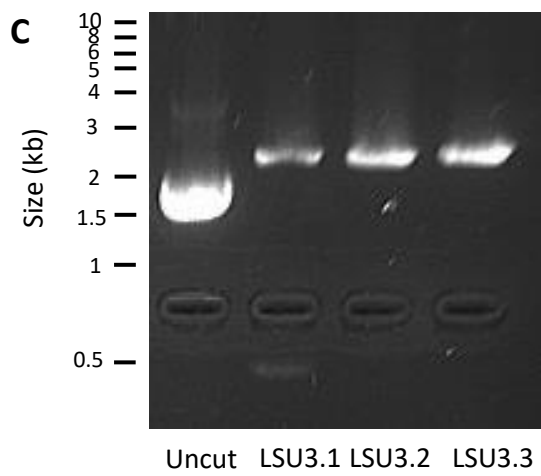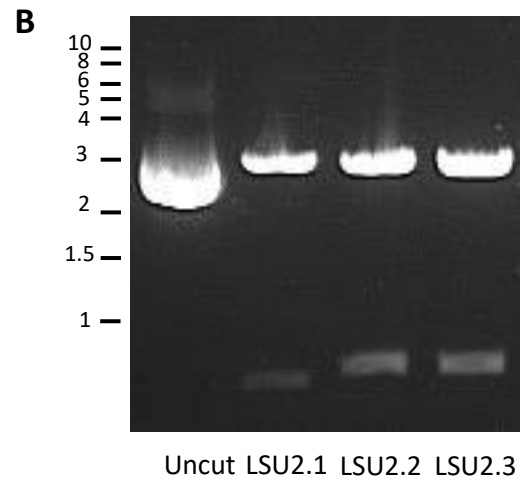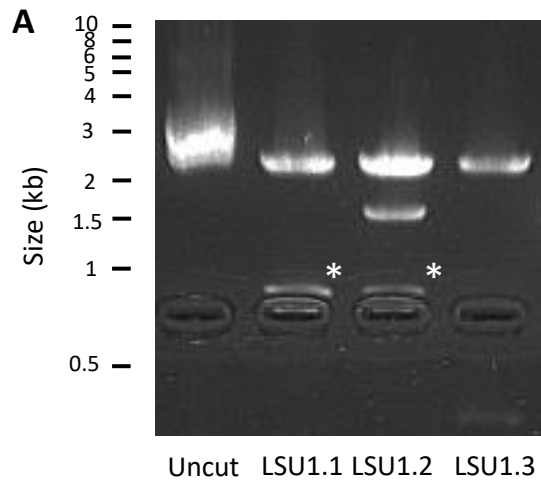
Figure 20: A series of HE-DP11 *Eco*RI restriction digests following cloning into pGEM® T-easy. HE-R8 *Eco*RI restriction digests following cloning into pGEM® T-easy of PCR products amplified using the universal primer sets (A) LSU1, (B) LSU2, (C) LSU3, (D) LSU4 and (E) 28S. Inserts of the correct size are indicated with an *.
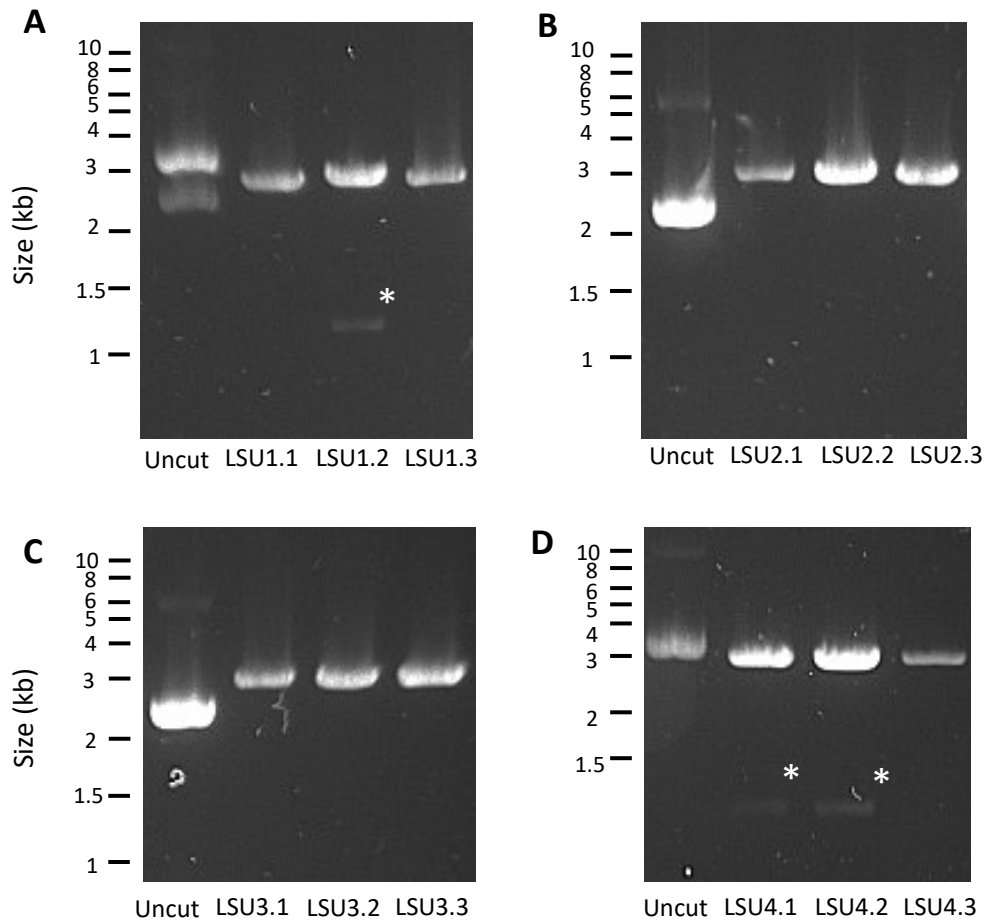
Figure 21: A series of HE-L3 *Eco*RI restriction digests following cloning into pGEM® T-easy. HE-L3 *Eco*RI restriction digests following cloning into pGEM® T-easy of PCR products amplified using the universal primer sets (A) LSU1, (B) LSU2, (C) LSU3 and (D) LSU4. Inserts of the correct size are indicated with an *.

A

Uncut LSU1.1 LSU1.2 LSU1.3

B

Uncut LSU2.1 LSU2.2 LSU2.3

C

Uncut HSP83.1 HSP83.2 HSP83.3
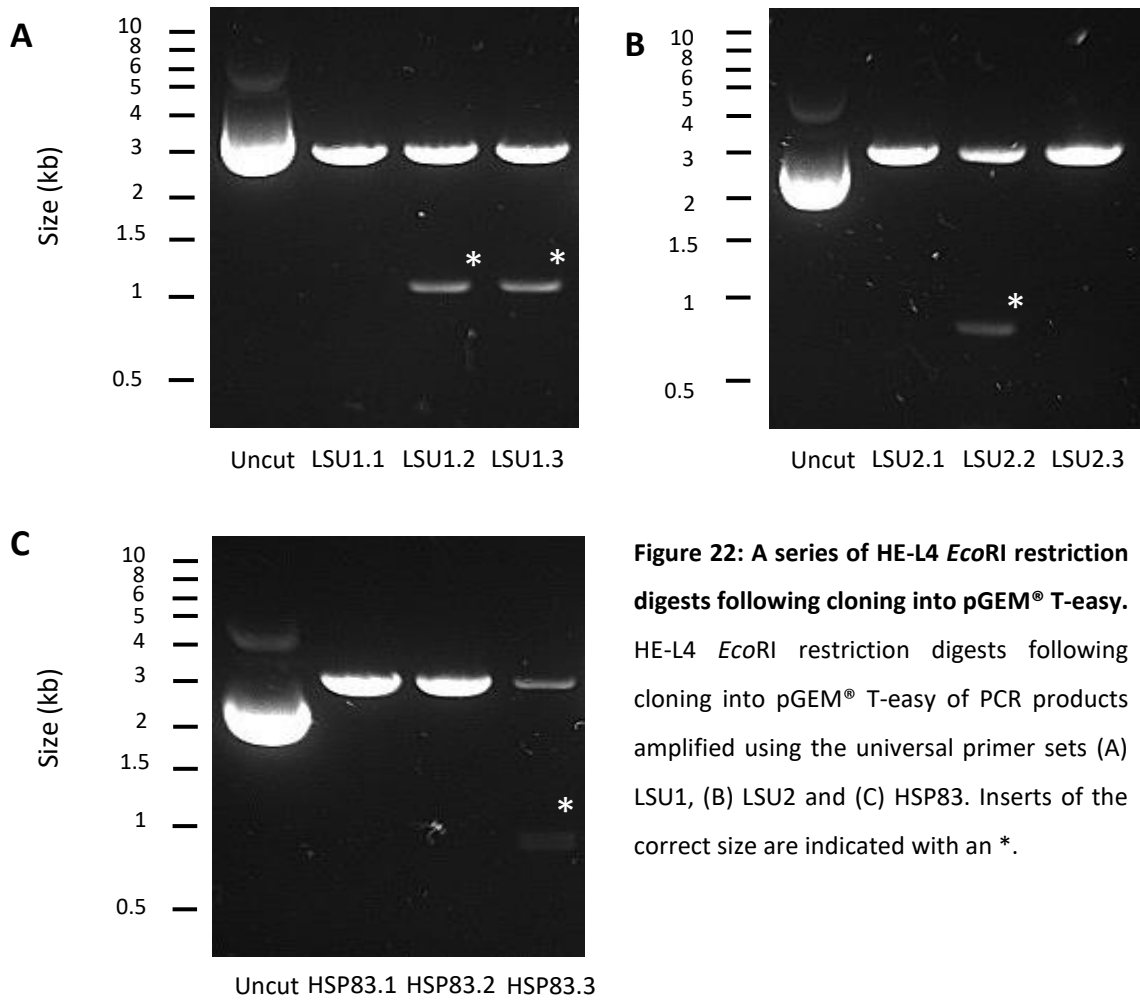
Figure 22: A series of HE-L4 *Eco*RI restriction digests following cloning into pGEM® T-easy. HE-L4 *Eco*RI restriction digests following cloning into pGEM® T-easy of PCR products amplified using the universal primer sets (A) LSU1, (B) LSU2 and (C) HSP83. Inserts of the correct size are indicated with an *.

74

Figure 23: A series of Dark Stuff *Eco*RI restriction digests following cloning into pGEM® T-easy. Dark Stuff *Eco*RI restriction digests following cloning into pGEM® T-easy of PCR products amplified using the universal primer sets (A) LSU1, (B) LSU3, (C) 28S and (D) HSP83. Inserts of the correct size are indicated with an *.

	Plasmid					
Sample	28S	Hsp83	LSU1	LSU2	LSU3	LSU4
Dark Stuff	✓	✓	✓			
HE-R8	✓			✓	✓	✓
HE-DP11			✓	✓	✓	✓
HE-L3			✓			✓
HE-L4		✓	✓	✓		

Table 8: Summary table indicating plasmids sent for external sequencing. Successful cloning of an insert was determined through restriction digests, see section 3.2.

3.3 Light microscopy of *Bodo saltans*, Dark Stuff and two University of York isolates

Differential interference contrast (DIC) and DAPI fluorescence images of *B. saltans*, Dark Stuff and two University of York isolates (HE-R8 and HE-DP11) were captured by confocal (*B. saltans* and Dark Stuff) and wide-field (HE-R8 and HE-DP11) epifluorescence microscopy respectively (Figures 24-27). Nuclear and kinetoplast DNA, where present, are indicated.
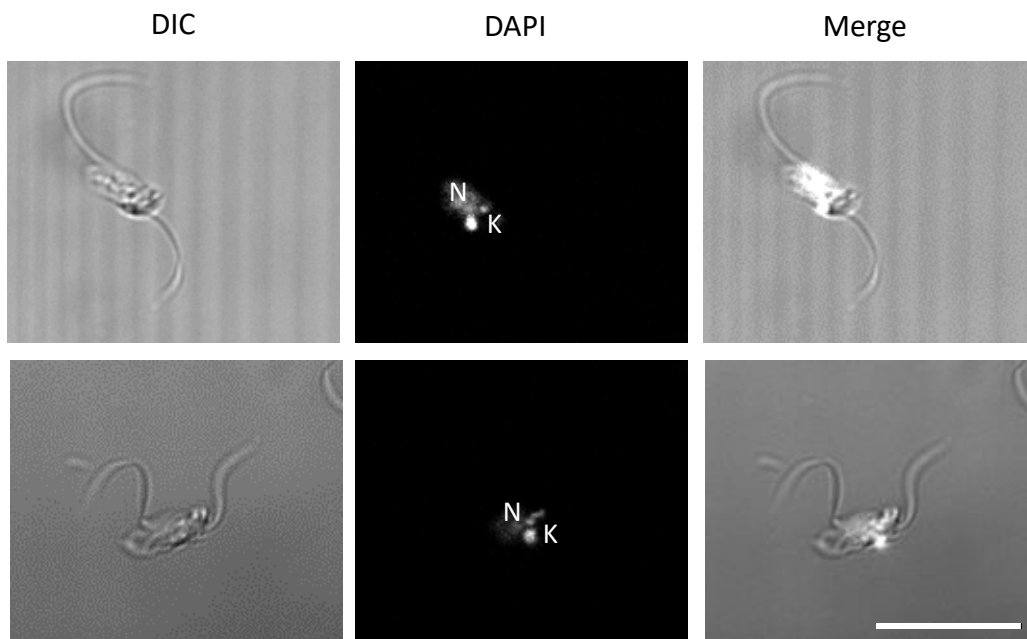
Figure 24: Fluorescence and DIC Microscopy of *Bodo Saltans* cells. DIC and DAPI (nucleic acid) fluorescence microscopy of two *B. saltans* cells, alongside merged images. N, nuclear DNA; K, kinetoplast DNA. Scale bar indicates 10µm.
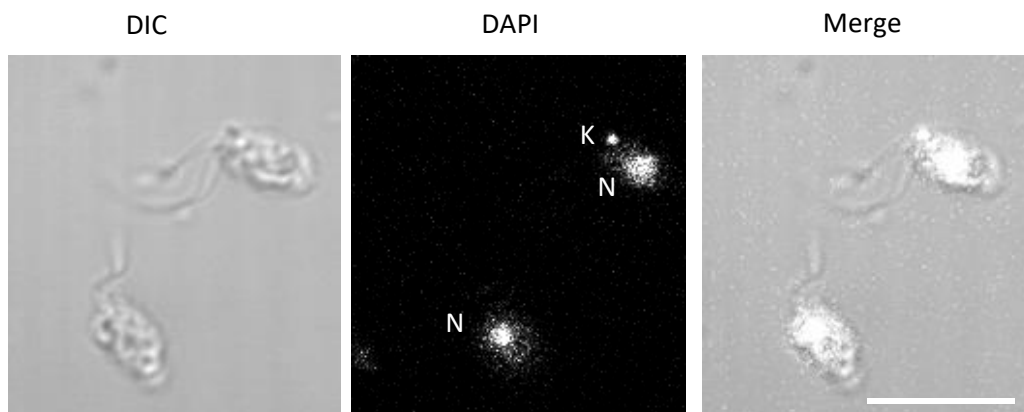
Figure 25: Fluorescence and DIC Microscopy of Dark Stuff cells. DIC and DAPI (nucleic acid) fluorescence microscopy of Dark Stuff cells, alongside merged images. N, nuclear DNA; K, kinetoplast DNA. Scale bar indicates 10µm.

Figure 26: Fluorescence and DIC Microscopy of a HE-R8 cell. DIC and DAPI (nucleic acid) fluorescence microscopy of a HW-R8 cell, alongside merged images. N, nuclear DNA. Scale bar indicates 10µm.
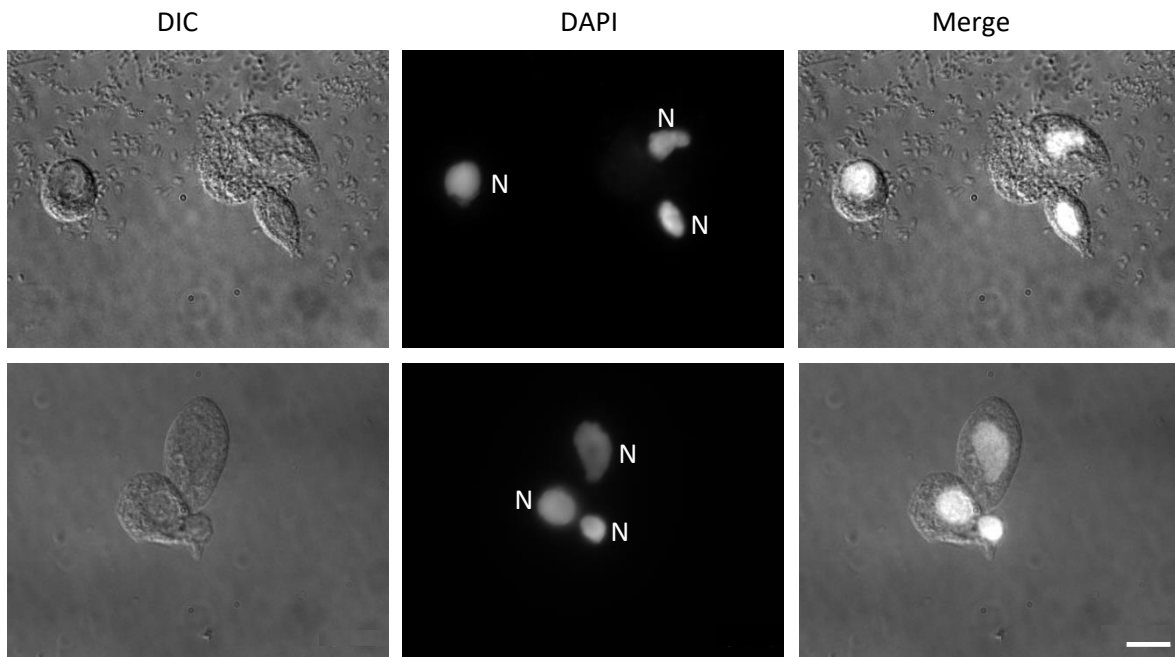
Figure 27: Fluorescence and DIC Microscopy of HE-DP11 cells. DIC and DAPI (nucleic acid) fluorescence microscopy of a HW-DP11 cells, alongside merged images. N, nuclear DNA. Scale bar indicates 10µm.

3.4 Bioinformatic analysis of returned University of York isolates and Dark Stuff sequences

Returned raw sequences (see appendix) were analysed by eye to ensure correct insertion of an insert and clipped before BLASTn analysis, Figure 28 shows, in FASTA format, subsequent clipping of poor-quality nucleotides. When using an M13 reverse sequencing primer, *Eco*RI (5'-GAATTC) and *Spe*I (5'-ACTAGT) restriction site are upstream of the multiple cloning site, where our cloned insert should be found (see example in Figures 28 & 30). Internal *Eco*RI restriction sites, previously suggested following *Eco*RI restriction digests of plasmids used for sequencing (Figure 29), are also present in; HW-DP11 LSU3 and LSU4, HE-R8 LSU3 and LSU4, and Dark Stuff 28S returned sequences. In other examples (indicated in Table 9) usable sequences were not returned.

>Dark_Stuff_Hsp83.1_M13rev29_clipped

AAGCTTTTAGGTGAACTATAGAATACTCAAGCTATGCATCCAACGCGTTGGGAGCTCTCCCATATGGTCGACCT
GCAGGCGGCCGC**GAATTC***ACTAGT*GATTCAGCTGATGTCCCTGATCATCAATACGTTTTACAGCAACAAGGAA
ATCTTCTTGCGTGAGATCATCTCCAACGCCTCCGATGCCCTCGACAAGATCCGCTACCAGAGCTTGACCGACAA
GGACGTCCTCCGTGACGAGCCCAGCCTCAAGATCCAGTTGATCCCCAACAAGGCCAACAAGACCTTGACCATC
CGTGATACTGGTATTGGTATGACCAAGAACGATATGGTGAACAACCTCGGTACCATCGCCCGCTCCGGCACCA
AGGCGTTCATGGAGGCGATCGAGTCTGGTGGCGACATCAGCATGATCGGTCAGTTCGGTGTTGGTTTCTACTC
TGCCTACCTCGTTGCCGACAAGGTCACCGTGATCTCCAAGCACAACGACGATGAAGCGCACATCTGGGAGTCC
TCTGCTGGCGGTACATTCACCGTTTCTAGCGTCGACGCCTCTACTGTGACTCGCGGTACCGAGATCATCCTCAG
CATGAAGGAAGATCAGCAGGAGTACCTCGAGGAGCGCCGCATCAAGGACCTCGTGAAGAAGCACAGCGAGT
TCATTGGCTACGACATCGAGCTCCAGGTTGAGAAGACCACCGAGAAGGAGGTCACCGACGACGAGGCCGAG
GAGGAGAAGAAGGACGAGGACGAGCCCAAGGTCGAGGAGGTCGACGAGAAGAAGGAGAAGAAGACCAAG
AAGGTCAAGGAGGTCTCCACCGAGTTCGAGATCCAGAACAAGAACAAGCCCCTCTGGACCCGCGACCCCAGG
ACGTCACAAGGAGGAGTACGCCTCCTTCTACAGGCGATCTCCACGACTGGGAGGACACCTTGCTGCAGGCAC
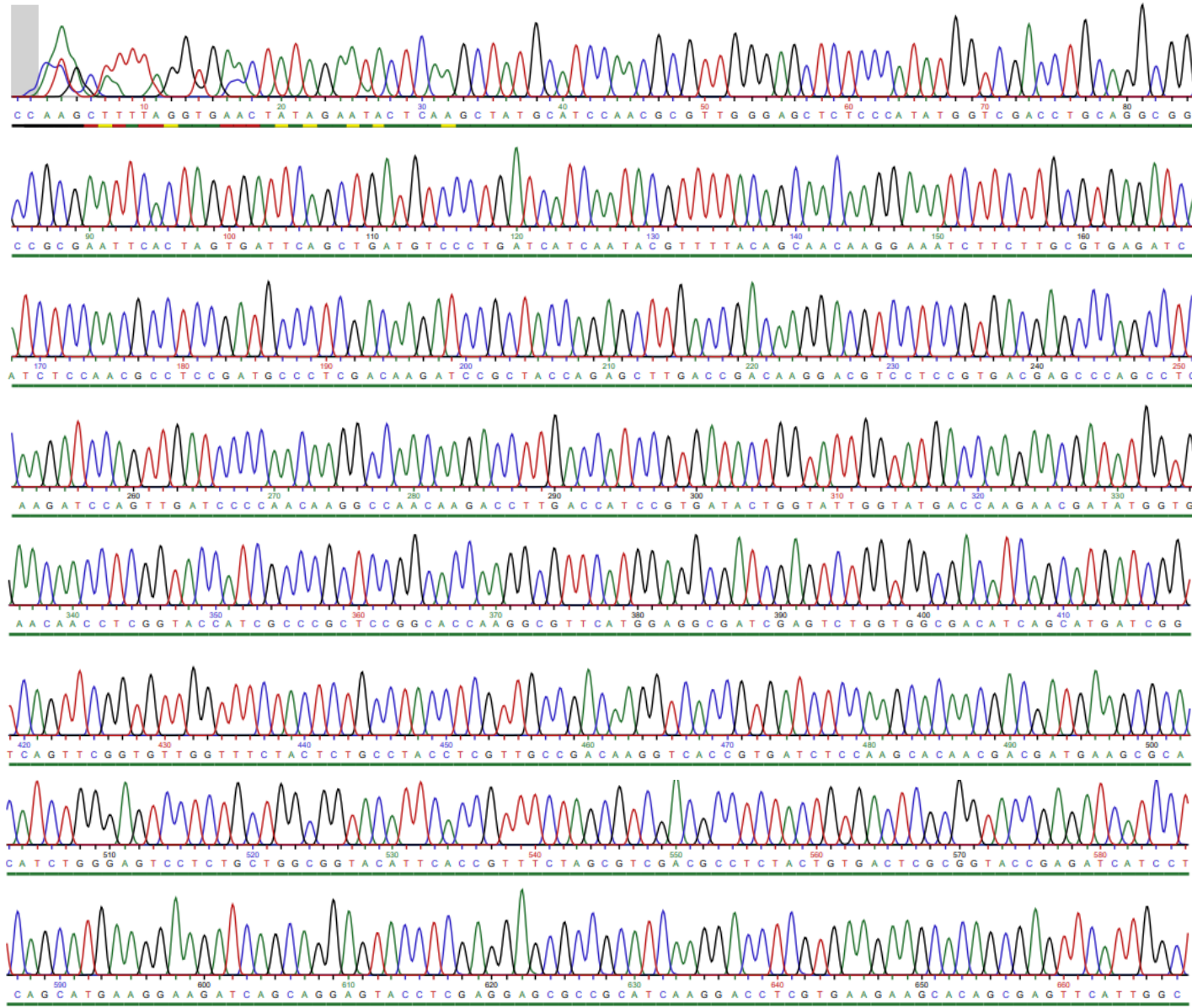TTCTCGTTGAGGGCAGTTGG

Figure 28: Example of clipped sequence returned following sequencing of pGEM® T-easy plasmid with a universal M13 reverse primer. FASTA formatted sequence returned following amplification of HSP83 from Dark Stuff genomic DNA, cloning into pGEM T-easy and sequencing using M13 reverse primer. The multiple cloning site of pGEM®T-easy is flanked by a 5'- *Eco*RI (5'-GAATTC) and *Spe*I (5'-ACTAGT) restriction site, highlighted in bold and italicized respectively. Sequence taken forward for BLASTn analysis is underlined.

>DP11_LSU3.1_M13_REV_UNCLIPPED

CCCCGAACTTTTAGGTGACACTATAGAATACTCAAGCTATGCATCCAACGCGTTGGGAGCTCTCCCATATGGTC
GACCTGCAGGCGGCCGC**GAATTC**<u>ACTAGT</u>GATTCCGCCCCAGTCAAACTCCCCCCCTAACCATGTCTTTCGCAA
AAATCAGACAATAAATGTCCTTAACTTTAGAACTGGTTTGCACCAGTCTTTTCTTCACGAAATAAGCAAAATGA
CTACGTGGGTAGTGGTTTTACAGGCTCGATTTCTCTCCCACCTAGGCTATACCCCACGTGCCATTTCACAAAGT
TAGACTAGAGTCAAGCTCAACAGGGTCTTCTTTCCCCGCTGATTATTCTAAGCCCGTTCCCTTAGCTGTGGGTT
CGCTAGATAGTAGATAGGGACAGTGGTAATCTCATTAATCCATTCATGCGCGTCACTAATTAGATGACGAGGC
ATTTGGCTACCTTAAGAGAGTCATAGTTACTCCCGCCGTTTACCCGCGCTTGGTTGAATTGCGTCACTTTGACA
TTCAGAGCACTGGGCAGAAATCACATTGTGTCATCACCTGTTGAGGCCGTCACAATGCTTTGTTTTTATTAAAC
AGTCGGATTACCTTTGTCCGCTTCAGTTCTGAGTTGATCGTTAATTGTATAAAGACGACCGAGGTCTACCATAT
*GAATTC*TTCGGTCGCAAGTCTATCAGCATGTCGCCACACTAACAAACAAGCTTGGATATCATCACATGGCCT
TTATACCCGATCCTCAGAGCCAATCCTTATCCCGAAGTTACGGATCCAATTTGCCGACTTCCCTTATCTACATTG
TTCTATGGACCAGAGGCTGCTAACCTTGGAGACCTGCTGCGGAATC***GAATTC***CCGCGGCCG<u>*CCATGG*</u>CGGCC
GGGAGCATGCGACGTCGGGCCCATTCGCCCTATAGTGAGTCGTATTACAATTCACTGGCCGTCGTTTTACAAC
GTCGTGACTGGGAAAACCCTGGCGTTACCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCGCTGGCGAAT
AACGAGAAGGCCCGCCCGATCGCCCTTCCCACAATTGCGCACCCGAAGGGGAAAGGACCGCCTGTACGGCC

Figure 29: Example clipped sequence, with an internal *Eco*RI restriction site, returned following sequencing of pGEM® T-easy plasmid with the universal M13 reverse primer. FASTA formatted sequence returned following amplification of LSU3 from HE-DP11 genomic DNA, cloning into pGEM® T-easy and sequencing using M13 reverse primer. The multiple cloning site of pGEM® T-easy is flanked by a 5'- *Eco*RI (5'-GAATTC) and *Spe*I (5'-ACTAGT) restriction sites, highlighted in bold and underlined respectively and a 3'- *Eco*RI (5'-GAATTC) and *Nco*I (5'-CCATGG) restriction site, highlighted in bold italics and underlined italics respectively. The internal *Eco*RI site highlighted in bold text and flanked by *s.

A

CCAAGCTTTTAGGTGAACTATAGAATACTCAAGCTATGCATCCAACGCGTTGGGAGCTCTCCCATATGGTCGACCTGCAGGCGG

CCGCGAATTCACTAGTGATTCAGCTGATGTCCCTGATCATCAATACGTTTTACAGCAACAAGGAAATCTTCTTGCGTGAGATC

ATCTCCAACGCCTCCGATGCCCTCGACAAGATCCGCTACCAGAGCTTGACCGACAAGGACGTCCTCCGTGACGAGCCCAGCCTC

AAGATCCAGTTGATCCCCAACAAGGCCAACAAGACCTTGACCATCCGTGATACTGGTATTGGTATGACCAAGAACGATATGGTG

AACAACCTCGGTACCATCGCCCGCTCCGGCACCAAGGCGTTCATGGAGGCGATCGAGTCTGGTGGCGACATCAGCATGATCGG

TCAGTTCGGTGTTGGTTTCTACTCTGCCTACCTCGTTGCCGACAAGGTCACCGTGATCTCCAAGCACAACGACGATGAAGCGCA

CATCTGGGAGTCCTCTGCTGGCGGTACATTCACCGTTTCTAGCGTCGACGCCTCTACTGTGACTCGCGGTACCGAGATCATCCT

CAGCATGAAGGAAGATCAGCAGGAGTACCTCGAGGAGCGCCGCATCAAGGACCTCGTGAAGAAGCACAGCGAGTTCATTGGC

Figure 30: Example chromatogram showing clipped sequence prior to BLASTn analysis. Part of a chromatogram returned following amplification of HSP83 from Dark Stuff genomic DNA, cloning into pGEM® T-easy and sequencing using M13 reverse primer. Sequences were initially scanned by eye and usable 'good' sequence identified by chromatogram peaks being evenly spaced, with similar orders of magnitude and minimal noise (or presence of baseline peaks). Chromatogram sequence highlighted in grey indicates areas of low quality, subsequently clipped prior to BLASTn analysis.

Sample	Universal Primer Combination	Usable Sequence Returned	Blastn top hits	E-value	Probable Taxonomic Lineage	Taxonomic Group
HE-DP	LSU1	Y	*Tetrahymena sp.* KGE15 partial 28S rRNA gene, strain KGE15	0.00E+00	*Tetrahymena pyriformis*	Cilophora
	LSU2	N				
	LSU3	Y	*Tetrahymena pyriformis* gene for 26S LSU rRNA	0.00E+00		
	LSU4	Y	*Tetrahymena pyriformis* gene for 26S LSU rRNA	0.00E+00		
HE-R8	LSU2	Y	*Tetrahymena rostrata* strain TR01 18S rRNA gene	0.00E+00	*Tetrahymena rostrata*	Cilophora
	LSU3	Y	*Tetrahymena rostrata* strain TR01 18S rRNA gene	0.00E+00		
	LSU4	N				
	28S	N				
HE-L3	LSU1	Y	*Plagiopyla sp.* QZ-2012 28S ribosomal RNA gene	0.00E+00	*Plagiopyla sp.*	Ciliates
	LSU4	N				
HE-L4	LSU1	Y	*Plagiopyla sp.* QZ-2012 28S rRNA gene, partial sequence	0.00E+00	*Plagiopyla sp.*	Ciliates
	LSU2	Y	*Plagiopyla sp.* QZ-2012 28S rRNA gene, partial sequence	0.00E+00		
	HSP83	Y	*Rhynchobodo* ATCC50359 Hsp90 gene, partial cds	0.00E+00		
Dark Stuff	LSU1	Y	*Neobodo saliens* HFCC11 28S rRNA gene	0.00E+00	*Neobodo saliens Rhynchobodo*	Kinetoplastids
	28S	N				
	HSP83	Y	*Rhynchobodo* ATCC50359 Hsp90 gene, partial cds	0.00E+00		

Table 9: Summary of BLASTn analysis. Usable sequences returned were used as queries for BLASTn analysis with top hits and associated E-value noted.

3.5 Bioinformatics on motility-related genes, Arp2/3, formins, WASP, and SCAR/WAVE

Motility-related genes *(H. sapiens* Arp2/3 complexes, formins, WASP, and *S. rosetta* WASP and SCAR/WAVE) were used as query sequences in BLASTp (protein-protein BLASTp) analysis using protein sequence datasets originating from several different sources summarized in Table 6. KOGG descriptions and annotations were used, and hypothetical proteins were compared against *H. sapiens* and *S. rosetta* queries using CLC sequence alignment software taking into consideration similarities and alternative AAs. In some instances, different models for some species were used as a comparison between databases and annotated KOGG sequences. Table 10 shows a breakdown of the number of actin isoforms and actin-related proteins (which includes Arp2, Arp3, Arp4, Arp6, and Arp10) observed between species and models. Higher numbers of actin isoforms correlate to higher gene copy numbers. Table 11 shows the breakdown of the presence of formins, Arp2, and Arp3 of which is indicated as either being positive or negative for the genes. A common trend observed is that species retained both Arp2/3 complex genes, except for *Encephalitozoon cuniculi* GB-M1 model which, unusually, only returned one copy of Arp2. Figure 31 shows the breakdown of returned sequences from each investigated species for both *H. sapiens* and *S. rosetta* sets of queries WASP, and SCAR/WAVE, and the presence of Arp2/3 complex genes. This also compared the multiple models used for some of the investigated species to highlight any inconsistencies between models.

Species	Model	Number of actins/ (actin-related)
Bodo saltans		16 (2)
Ectocarpus variabilis		N
Ectocarpus siliculosus		9 (3)
Cyanidioschyzon merolae		5 (2)
Chondrus crispus	*Stackhouse*	6 (4)
Galdieria phlegrea	*Soos*	9 (3)
Galdieria sulphuraria (Azora/MS1/MtSh/SAG21.92/YNP5578.1)	*Azora*	8 (2)
	MS1	9 (2)
	MtSh	10 (2)
	SAG21.92	6 (2)
	YNP5578.1	7 (1)
Gracilariopsis chorda		9 (2)
Porphyra umbilicalis	*isolate 4086291*	7 (3)
Pyropia yezoensis		10 (3)
Rozella allomycis		6 (4)
Micromonas commoda		4 (4)
Micromonas pusilla		4 (3)
Chlamydomonas reinhardtii	*External*	9 (5)
	Filtered	7 (4)
Cyanophora paradoxa	*CCMP329*	9 (8)
Glaucocystis nostochinearum		2
Spizellomyces punctatus		3
Mitosporidium daphniae		3 (1)
Encephalitozoon cuniculi	*GB-M1*	4(1)
	EC1	2
	EC2	3
	EC3	2
	EcunIII	2
Naegleria gruberi		26 (9)
Gregarina niphandrodes		6
Chromera velia		20 (6)
Vitrella brassicaformis		7 (4)
Ostreococcus tauri		9 (8)
Ostreococcus lucimarinus		5 (5 hypothetical proteins)

Table 10: protein BLASTp returns for actin isoforms. Models refer to available database references which returned differing numbers of actins and actin-related proteins. Actin-related proteins are in parenthesis, of which include Arp2/3, but also Arp4p, Arp6p, and Arp10p or Act3p as per KOGG notation. *O. lucimarinus* returns were hypothetical and thus underwent extra sequence alignment to conclude origins.

Species	Model	Formins	Arp2	Arp3
Bodo saltans		N	Y	Y
Ectocarpus variabilis		N	N	N
Ectocarpus siliculosus		Y (4)	Y	Y
Cyanidioschyzon merolae		N	N	N
Chondrus crispus	*Stackhouse*	N	Y	Y
Galdieria phlegrea	*Soos*	N	Y	Y
Galdieria sulphuraria (Azora/MS1/MtSh/SAG21.92/YNP5578.1)	*Azora*	N	Y	Y
	MS1	N	Y	Y
	MtSh	N	Y	Y
	SAG21.92	N	Y	Y
	YNP5578.1	N	Y	Y
Gracilariopsis chorda		N	N	N
Porphyra umbilicalis	*isolate 4086291*	N	Y	Y
Pyropia yezoensis		N	Y	Y
Rozella allomycis		N	Y	Y
Micromonas commoda		N	Y	Y
Micromonas pusilla		N	Y	Y
Chlamydomonas reinhardtii	*External*	N	N	N
	Filtered	N	N	N
Cyanophora paradoxa	*CCMP329*	N	Y	Y
Glaucocystis nostochinearum		N	N	N
Spizellomyces punctatus		N	Y	Y
Mitosporidium daphniae		N	Y	Y
Encephalitozoon cuniculi	*GB-M1*	N	Y	N
	EC1	N	N	N
	EC2	N	N	N
	EC3	N	N	N
	EcunIII	N	N	N
Naegleria gruberi		N	Y	Y
Gregarina niphandrodes		N	N	N
Chromera velia		N	Y	Y
Vitrella brassicaformis		N	Y	Y
Ostreococcus tauri		N	Y	Y
Ostreococcus lucimarinus		N	Y	Y

Table 11: Protein BLASTp returns for formin and Arp2/3 complexes from *H. sapiens* queries. Species returns for formins and Arp2/3 complexes. *E. siliculosus* was the only species to return formins under KOGG annotation of which there were four different genes to note. Most species retained Arp2/3 complex genes as is expected for a conserved gene with only a few exceptions across sister-species. N= negative, Y= positive.
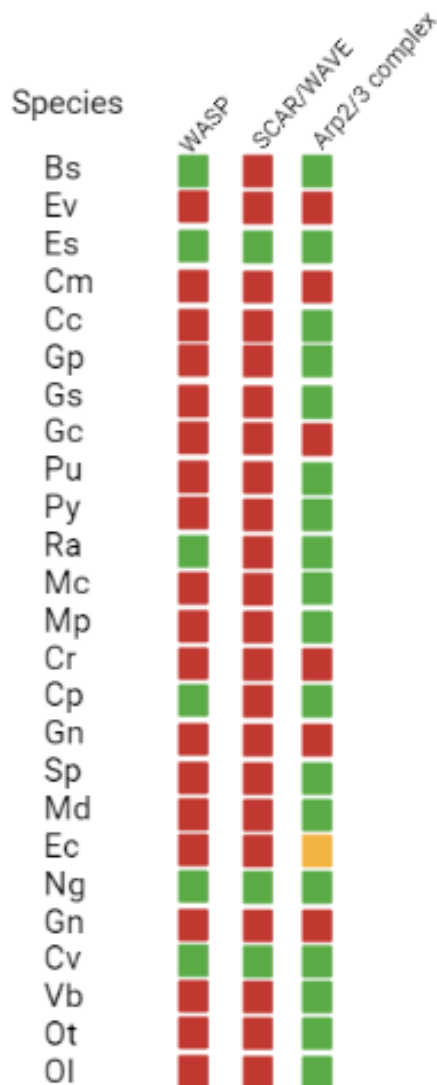
Figure 31: Annotation of observed Arp2/3 complexes, WASP, and SCAR/WAVE gene returns from protein BLASTp on investigates species. Species names are abbreviated to genus and species initials. *Bd= Bodo saltans, Ev= Ectocarpus variabilis, Es= Ectocarpus siliculosus, Cm= Cyanidioschyzon merolae, Cc= Chondrus crispus, Gp= Galdieria phlegrea, Gs= Galdieria sulphuraria* (and includes models Azora, MS1, MtSh, SAG21-92, and YNP5578.1), *Gc= Gracilariopsis chorda, Pu= Porphyra umbilicalis, Py= Pyropia yezoensis, Ra= Rozella allomycis, Mc= Micromonas commode, Mp= Micromonas pusilla, Cr= Chlamydomonas reinhardtii* (including both external and filtered models), *Cp= Cyanophora paradoxa, Gn= Glaucocystis nostochinearum, Sp= Spizellomyces punctatus, Md= Mitosporidium daphniae, Ec= Encephalitozoon cuniculi* (and includes models GB-M1, ECI, ECII, ECIII, and EcunIII), *Ng= Naegleria gruberi, Gn= Gregarina niphandrodes, Cv= Chromera velia, Vb= Vitrella brassicaformis, Ot= Ostreococcus tauri, and Ol= Ostreococcus lucimarinus.* Coloured boxes denote positive or negative returns upon further sequence alignment for hypothetical proteins unlisted with KOGG descriptions. Green boxes indicate positive for genes. Red boxes indicate negative for genes. Yellow indicates uncertainty, Ec returned only one positive for Arp2 in one model but negative in the rest. All Ec models lacked Arp3.

86

Chapter 4

Discussion

4.1 Isolation and Identification of protists originating from Embo in Scotland and a number of locations at the University of York

Unknown isolates

The cloning and sequencing of 28S, Hsp83 and LSU1-4 genes in Dark Stuff and isolates collected from the University of York led to us being able to tentatively identify the ciliates and, in the case of Dark Stuff, a kinetoplastid present in our samples; further corroborated by light microscopy data. Following BLASTn analysis, Dark Stuff sequences returned *Rhynchobodo* and *Neobodo* as top hits (Figure 32) (when amplifying Hsp83 and LSU1 respectively), both of which are kinetoplastid genera, and both being among the twenty most common 'zooflagellates' (by older systems of classification) found in freshwater, but that can also tolerate saltwater (Morgan-Smith, et al., 2013; von der Heyden, & Cavalier-Smith, 2005). However, geographical distribution maps do not show data that suggests *N. saliens* has been located in the U.K. (https://www.gbif.org/species/11214023). *Rhynchobodo*, in contrast, has been sighted in the U.K. yet the geographical distribution map used does not indicate specifically where and other attempts at finding other maps were unsuccessful (https://www.gbif.org/species/7667356) (Figure 32).

Von der Heyden (2005) identified kinetoplastids, *Neobodo* and *B. saltans* (Metakinetoplastina) and *Ichthyobodo* (Prokinetoplastina) in Danish lake sediments thereby suggesting that Dark Stuff, being found in Scotland, typically of a cooler climate, may indeed be identified correctly. Light-microscopy, combining DIC and DAPI images, also identified a kinetoplast in Dark Stuff , a characteristic unique to the kinetoplastids (Figure 25). DIC microscopy also identified the presence of a flagellum, although, due to limited resolution, we cannot determine if this is a single-flagellum or two flagella of differing lengths, which morphologically, is characteristic of both *Neobodo* and *Rhynchobodo*.

Although the findings in Chapter 3 can provide inferences as to the identification of Dark Stuff; EM, and full genome sequencing, potentially against a known reference genome, would ultimately be needed to precisely identify our cultured kinetoplastid, identifying ultrastructural components such as the presence of a kinetoplast, length of the PFR, basal body pairs, and the flagellar pocket (Wheeler, et al., 2013). The Darwin Tree of Life project (a British and Ireland-focused genome sequencing project with the aim of sequencing all species native to the UK) may provide such a reference genome, or conversely our newly sequenced genomes could, ultimately, form part of this extensive dataset.
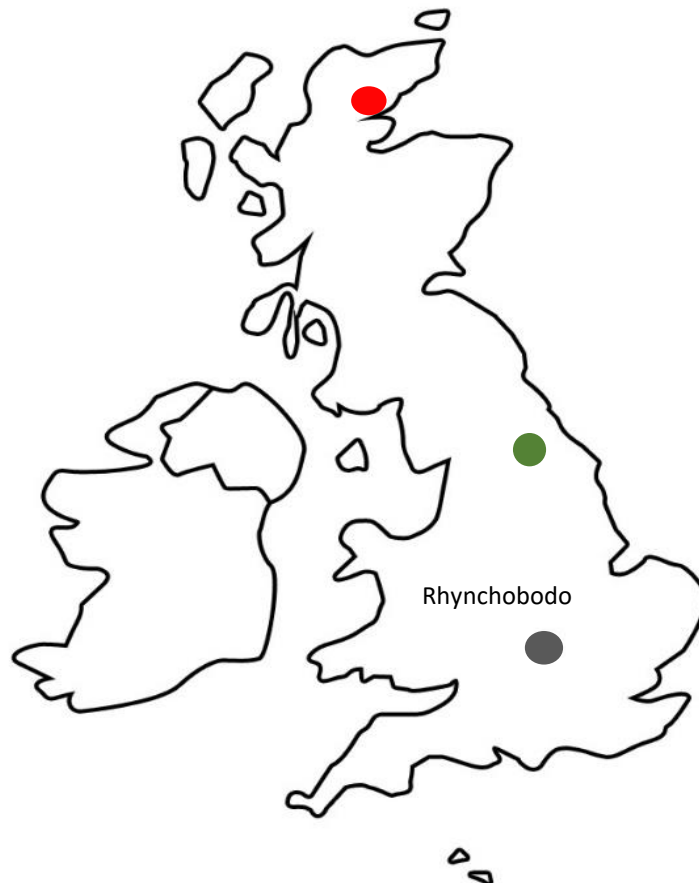
Figure 32: Geographical distribution map for *Rhynchobodo* and *Neobodo saliens* recorded in the U.K. *Rhynchobodo* has been sighted in the U.K. but geography maps do not precisely share where it was sighted, indicated with a grey circle. *Neobodo saliens* alternatively has not been observed in the U.K. and as such is not represented in this figure. Embo beach general location is indicate with a red circle, and the University of York general location is indicated with a green circle.

In contrast, all York samples returned sequences that allow us to broadly identify our cultured cells as ciliates. Although difficult to distinguish their flagella/cilia by light microscopy, these large-bodied species were observed, by eye, to propel themselves through their media. Additionally, cells lacked a kinetoplast, as observed in Dark Stuff (Figures 26 & 27). Ciliates, as a whole, are a large group of complex unicellular organisms with more than 8,000 described species divided into 11 classes (as of Zhang, et al., 2012) that are cosmopolitan in fresh-water and marine environments and range from predators of flagellates to phototrophic members.

Following BLASTn analysis, HE-DP11 and HE-R8 returned top hits indicating the presence of *Tetrahymena*. Light- microscopy and sample origin also potentially confirmed the presence of this ciliate in our samples, with these large cells, with large nuclei (Doerder & Brunk, 2012), isolated from a fresh-water source. Using geographical maps (https://eol.org/pages/4691) (Figure 33), 31 sites have been reported in the U.K. to have *Tetrahymenidae* species, of which four locations, in the North of England around Cumbria (so, broadly, latitudinally equivalent to our York sites), including the species *T. pyriformis*, and *T. rostrata*, both of which were the top hits following HE-DP11 and HE-R8 BLASTn analysis.

Despite the absence of light microscopy images for cells originating from HE-L3 and HE-L4, following BLASTn analyses, top hits indicate both these species (sampled from the same lake in York) as *Plagiopyla* species of the sub-phylum Ciliophora, known to be cosmopolitan in origin, typically found in both fresh-water and marine environments world-wide (Zhang, et al., 2012). Again, these large ciliate cells, when observed by eye under a light-microscope, propelled themselves through their media using multiple flagella/cilia. Although *Plagiopyla* species known geographical distribution remains fairly limited (with most observed locations being in New Zealand), there is, however, one record for *Plagiopyla nasuta* at a location in Cumbria (Plagiopyla Stein 1860 - Encyclopedia of Life (eol.org)) (Stein, 1860), in Figure 33.
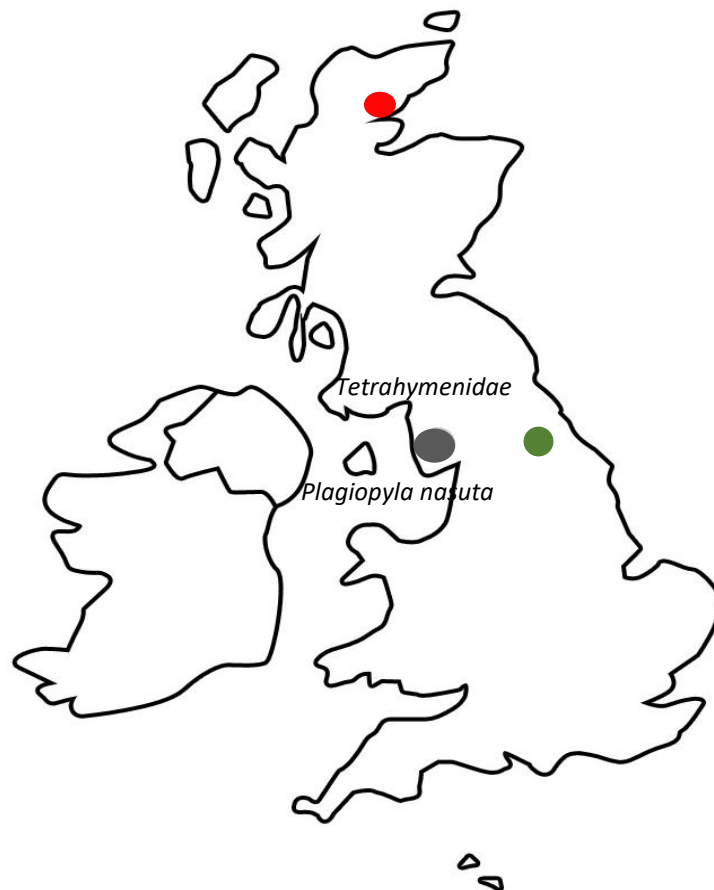
Figure 33: Geographical distribution map for *Tetrahymena pyriformis*, and *T. rostrata* and *Plagiopyla sp.* recorded in the U.K. Both *Tetrahymena* species and *Plagiopyla sp.* have been reported in Cumbria (Stein, 1860), latitudinally equivalent to the York sites. Both indicated with a shared grey circle. Embo beach general location is indicated with a red circle, and the University of York general location is indicated with a green circle.

As with Dark Stuff, each of the York samples would need extensive EM analysis. For Tetrahymena, the presence of a "crescent" micronucleus (MIC) during early conjugation (both an asexual and sexual reorganization process (Suganuma, & Yamamoto, 1992) and macronucleus (MAC) (Ruehle, et al., 2016). For *Plagiopyla*, striated band length, structure of buccal cavity (essentially the mouth of the cell), position and length of cytoproct (an exocytotic vesicle that appears when solid particles are ready to be expelled), and the presence of MIC and MAC (Nitla, et al.,2019). As well as whole genome sequencing, ideally against a reference genome, to ultimately identify our cultured ciliate cells.

4.2 An extensive bioinformatic analysis to probe the hypothesis that switching between flagellate and amoeboid (and vice-versa) cell forms is attributable to a defined set of 'motility genes'

Bioinformatics (BLASTp analyses) on 34 different taxa, revealed, in some species, genes associated with α-motility and pseudopod formation (Arp2/3, formin, WASP, and SCAR/WAVE). Documentation of modes of motility was recorded in Table 12 to compare the findings from this analysis within current literature.

BLASTp analyses

In order to probe further the hypothesis that transformation from a flagellate (capable of swimming or gliding) to an amoeboid (α-motility) and vice versa, is a confinement-induced stress-response mechanism, that may or may not have been lost across eukaryotic taxa (Brunet, et al., 2021), and whether or not all actin-related proteins are required for this switch, protein query sequences for *H. sapiens* Arp2, Arp3, and WASP and *S. rosetta* WASP and SCAR/WAVE were used to identify potential orthologous sequences in a number of different taxa of various lineages (including fungi, kinetoplastids, and algae). However, not all proteins were annotated in all datasets and therefore some degree of interpretation using CLC sequence alignment software was needed to interpret results (Tables 10 &11, and Figure 31).

Bodo saltans

Interestingly *B. saltans* appears to have retained both WASP and Arp2/3 complexes and is second only to *N. gruberi* in terms of actin isomer number but, to date, has only been observed to have flagellated motility (Gomaa, et al., 2022). However, current literature does not confirm if the presence of both WASP and SCAR/WAVE is necessary, or if this morphological switch can be observed with only one of these proteins alongside Arp2/3 complexes. Future work observing, under static confinement, flagellate-to-amoeboid switching in a variety of cells and taxa would be helpful in answering this question and further bioinformatic probing of other genomes, across the eToL, would provide further inferences as to the conservation of these complexes and highlight the potential for flagellate-to-amoeboid (and vice-versa) switching in other organisms as a response to environmental stress.

Ectocarpus

E. siliculosus and its sister-species, *E. variabilis*, appear to be polar-opposites to one another in terms of conservation of pseudopod-formation proteins found within their respective genomes. *E. siliculosus* has retained genes for actin (six isomers), formins (four returned sequences (interestingly *E. variabilis* is the only species probed here to apparently possess formins), Arp2/3 complexes, WASP, and SCAR/WAVE whilst *E. variabilis* appears to lack actins altogether. However, neither species have been observed to have a flagellate form (Jia, et al., 2017; MacKay, & Gallant 1991). This prompts the discussion, given that *E. siliculosus* has two differing forms of growth (apical- occurring at the tip that increases length, and intercalary- below the tip that grows branches increasing surface area), if these genes are in some way related to intercalary growth within branching, or if actin-related proteins have potential moonlighting properties that have yet to be identified (see section 4.3).

Red algae

C. merolae is a red alga (family *Cyanidiaceae*) adapted to high sulphur, acidic, hot spring environments (De Luca, et al., 1978) that has an extremely simplistic cell composition of only a few major organelles (chloroplast, mitochondrion, and nucleus). Despite a streamlined genome three actin isomers were apparent following our BLASTp analysis*.*

C. crispus is another red alga (family *Gigantinaceae*) which changes colour dependent upon the depth it is inhabiting and is typically found along Atlantic coasts (Collen, et al., 2014). The Stackhouse model returned sequences for Arp2/3 complex proteins, and two actin isomers were identified, which would perhaps be expected as they grow as large branching networks.

G. phlegrea, like *C. merolae* as both are part of the *Cyanidiaceae* family, is noted for having great metabolic versatility (di Cicco, et al., 2021) for its resistance and ability to thrive in unfavourable wastewater conditions. Like the other red alga analysed, Arp2/3 complex sequences and, in this case, six actin isomers were returned following BLASTp analysis.

G. sulphuraria is the sister species to *G. phlegrea,* again of the family *Cyanidiaceae* and is broadly known for its metabolic capabilities including photosynthesis and heterotrophy and, as with *C. merolae,* is an acidophile capable of growing in low pH environments of between pH0 and pH4 (Schönknecht, et al., 2013). Of the five models investigated in Tables 10, returned sequences only differed based on actin isomer number, with the presence of Arp2/3

complex proteins in all (Figure 31) and all immobile and possessing branching filaments (Table 12).

G. chorda (*Cyanidiaceae*) is another close relative of red alga species used in this analysis and is known for its extensive medical benefits in eastern countries, showing potential neuroprotective effects in cultured hippocampal neurons (Mohibbullah, et al., 2016; Lin, 2008). Unlike other members of the *Cyanidiaceae*, actin-related proteins (Arp2/3) were not found (as was the case with *E. variabilis)*, but nine actin isomers were returned following analysis. This is perhaps indicative that the loss of Arp2/3 complex proteins may have been a relatively recent, in evolutionary terms, event.

P. umbilicalis is a marine red alga that has been noted to be present in shorelines around the U.K. (https://www.marlin.ac.uk/species/detail/1463). As with other red algae, *P. umbilicalis* forms branching filaments (Miranda, et al., 2013) and therefore lacks WASP, and SCAR/WAVE orthologues. Brawley, et al., 2017 appears to confirm observations here, as it was noted that *P. umbilicalis* has four different actin isomers (three belonging to hypothetical actin-related proteins) and Arp4, associated with chromatin remodelling. However, they did not find any Arp2/3 complex genes, which are apparent in this analysis, perhaps, in part, due to hypothetical proteins being mis-interpreted. However, Arp2/3 complexes appear typically conserved in red algae spores and are apparent in all red algae species looked at in this analysis.

P. yezoensis, another red alga (family *Bangiophyceae*), is a common rhodophyte found in intertidal zones that can survive severe water loss through the regulation of ferredoxin-mediated transport chains (Yu, et al., 2018). Interestingly, the ferredoxin protein is thought to have moonlighting functions in other parasitic taxa, potentially highlighting a precedent for other expressed proteins performing moonlighting roles in a similar fashion (section 4.3). As with the other red algae, this species lacks motility and retains Arp2/3 complex proteins and seven actin isomers.

Parasitic lineages

R. allomycis was the first member of the phylum Cryptomycota (a fungal lineage) to have its genome sequenced and it is an obligate parasite of the Blastocladiomycotan fungus *Allomycis* (Letcher, et al., 2017)*.* Immature resting sporangia grow, as naked protoplasm, inside the

hyphae of the host slime mold. The parasite requires the host to produce the cell wall of the zoosporangium, which then cleaves into numerous zoospores with a single flagellum. This fungus therefore possesses two, or three, modes of motility through its life cycle from immature to mature sporangia (James, et al., 2006). *R. allomycis* has been shown to produce and retract its flagellum, thereby switching to an amoeboid form, through expression of WASP and Arp2/3 genes, also with two actin isoforms. This apparently shows a lack of one of the Arp2/3 activators (SCAR/WAVE) does not necessarily inhibit pseudopod formation.

M. daphniae (also referred to as *Microsporidia daphniae* in other literature) belongs to the phylum Microsporidia and is a parasite to water fleas of the genus *Daphnia* and emerged from the early fungal lineage Rozellomycota that replicate *en masse*, via binary fission, within host tissues – completely occupying the host's haemocoel (a cavity within invertebrates) (Simakova, et al., 2018; Corsaro, et al., 2016). Characterized by extreme morphological and genomic adaptations to intracellular parasitism; including flagellum loss, mtDNA reduction, a much-compacted genome and plasticity (Vávra, & Lukeš, 2013) and possessing polar filaments (a syringe-like hollow tube structure arranged into the spore) that injects spore content into the host cell (Franzen, 2004). This parasitic lifestyle is reflected in our analysis, as only Arp2/3 complexes and two or three actin isoforms appear present, perhaps unsurprising as this obligate parasitic lifestyle would not be expected to require motility-related proteins for pseudopodia or pseudopodia-like projection formation (Corsaro, et al., 2016).

E. cuniculi, also of the phylum Microsporidia, and like *M. daphnia,* have morphological and genomic adaptations related to their parasitic lifestyles, lacking both peroxisomes and mitochondria (Katinka, et al., 2001). They also have not been observed to form pseudopodia and lack flagellum (Magalhães et al., 2022). Interestingly, models in this investigation only returned actin isomers, without WASP, and SCAR/WAVE. However, the GB-M1 (JGI Phytocosm- Peyretaillade, et al., 2009) model returned one Arp2 complex protein, but other models (ECI-III and EcunIII) from NCBI lacked this gene. However, JGI Phytocosm represents a more recent model for this Microsporidia species and as such is likely to better represent *Encephalitozoon*.

G. niphandrodes (phylum Apicomplexa) is a parasite of the beetle, *Tenebrio molitor*. Larvae *T. molitor*, infected with *G. niphandrodes*, show signs of nutritional deficiency highlighting parasitic properties this species has within its host (Rodriguez, et al., 2007). As is the case with other parasitic lineages, a lack of motility-associates genes is represented within *G. niphandrodes'* genome, with only six actin isoforms identified.

C. velia (family *Chromeraceae*), is phylogenetically related to Apicomplexa but, as it is photosynthetic, is a good organism link between botanical protists (algae) and zoological protists (protozoa) (Moore, et al., 2008). Although characterized as immotile coccoid cells, they have been documented by Oborník, et al., (2011) to form flagella. The *C. velia* genome encodes all motility-associated proteins, and a large number of actin isomers (14); therefore, potentially indicating this morphological change (from an immotile coccoid cell to a flagellate) is not only necessary for pseudopod formation, but that as the immotile coccoid cells do not form branching filaments (such as in algae), some form of motility must be possible.

Micromonas

M. commoda, and its sister species, *M. pusilla* are both picophytoplankton, with these *Micromonas* strains originally described as a single species, *M. pusilla* (Knight-Jones, & Walne, 1951), but more recently it has become accepted that *M. pusilla* comprises of a number of cryptic species (Šlapeta, et al., 2006; Worden 2006). Both species are fast swimmers and contain Arp2/3 complex genes with very few actin isomers. Additionally, no WASP and SCAR/WAVE genes were found, thereby perhaps indicating an inability of these picophytoplanktons to switch to amoeboid forms

Green and blue-green algae

C. reinhardtii (family *Chlamydomonadaceae*), is a single-celled, bi-flagellated green alga with an eyespot that senses light. It is a found world-wide in both fresh-water and soil environments (Ueki, et al., 2016). Both Chlamydomonas models analyses here returned only actin isomers; no motility-related proteins were found. This is perhaps unsurprising as these species have not been documented to change morphology through any life-cycle stage.

V. brassicaformis (family *Vitrellaceae*) is a free-living, photosynthetic marine green algae, and in conjunction with *C. velia* (*V. brassicaformis* and *C. velia* are the only two, currently described members of the phylum Chromerida), is phylogenetically the closest organism to the Apicomplexa (a group containing many human and animal parasites). Although closely related to *C. velia, V. brassicaformis* differs significantly in morphology, life-cycle, and accessory photosynthetic pigmentation (Oborník, et al., 2012). *V. brassicaformis* form flagella, however they have also been observed, at points in their life-cycle, to possess immotile autospores and bi-flagellate motile zoospores. Our analysis revealed only Arp2/3 complex proteins and a small number of actin isomers, perhaps unsurprising as the full complement of motility proteins would perhaps not be expected in an organism that isn't capable of pseudopod formation (Oborník, et al., 2012).

G. nostochinearum, of the family *Glaucocystaceae*, is an endophytic blue-green algae, which have been observed to change colour from green to yellow with age (Echlin, 1967), with younger cultures possessing more endocyanelles (a glaucophyte chloroplast only containing chlorophyll α). Echlin (1967) described older cultures as having two flagellar-like structures between the host cell wall and cell membrane, but was unable to detect, via EM, a pair of central fibrils characteristics of algal flagella. Furthermore Schnepf, et al., 1966 detected the presence of flagellar bases but appeared unable to find whole flagella. Here, *G. nostochinearum* lacked any motility-related proteins with only two isoforms of actin, potentially indicating this species uses its flagellum in a more passive way, potentially for sensing or feeding mechanisms as proposed by Keeling (2019).

C. paradoxa, also of the family *Glaucocystaceae*, is a fresh-water species of a Glaucophyte (which differ from red and green algae (Palmer, et al., 2004)) with photosynthetic capabilities (cyanelles—plastids with photosynthetic pigments) and may be of basal Archaeplastida origin (Keeling, 2004). This species is a flagellated, free-living cell and our analysis indicates it has the potential to switch morphologies (presence of Arp2/3 complexes, actin isoforms, and WASP). However, this could be due, in part, to its evolutionary history, being a branch that split away from other red and green algae that have since lost these motility-related proteins (as with *C. reinhardtii*).

Spizellomyces punctatus

S. punctatus (family *Spizellomycetaceae*) is a chytrid fungal species found in soils and, like terrestrial fungi, is a saprotrophic feeder of decaying matter. During its life-cycle it is capable of producing both uniflagellated and amoeboid zoospores (Gill, & Fast, 2006). Interestingly, this species possesses the ability to post-transcriptionally edit its RNA, potentially owing to its ability to produce flagellated and amoeboid zoosporia as our analysis shows *S. punctatus* only retains Arp2/3 complex genes and a limited number of actin isoforms. However, our findings differ from that of Prostak (2021), who appeared to show WASP and SCAR/WAVE *S. punctatus* orthologues, finding that were perhaps the more expected, given *S. punctatus'* ability to switch morphologies. It is possible we used an outdated FungiDB reference genome or that those WASP and SCAR/WAVE proteins previously identified were called as hypothetical, unannotated proteins in our analysis.

Naegleria gruberi

N. gruberi (family *Vahlkampfiidae*), found in both fresh-water and soil environments, is a free-living organism that able to transition freely between flagellate and amoeboid forms, and *vice versa,* and can also exist as a cyst when under periods of environmental stress (Carosi, et al., 1977). Elicited morphology is dependent upon several factors including reproductive stages (amoeba), nutrient availability (starving leading to the formation of flagella), and temperature (cold temps causing cyst formation) (Visvesvara, et al., 2007). This cell form variety is reflected in our analysis as Arp2/3 complexes, many actin isomers (>17), WASP, and SCAR/WAVE all appear present, supporting the hypothesis that these genes are related to stress-induced, motility-related morphological change.

Ostreococcus

O. tauri and its sister-species *O. lucimarinus* both belong to the family *Bathycoccaceae* and are free-living, green algae with very simple ultrastructures and compact genomes (indicated by a lack of WASP and SCAR/WAVE proteins). They lack a cell wall, contain only a single chloroplast, and do not possess any kind of flagellum (Subirana, et al., 2013). It is perhaps unsurprising therefore, due to their simplicity, that they appear to not possess any of our motility-related proteins.

4.3 Moonlighting proteins, a complex secondary feature of well-characterized proteins, may also be responsible for switching modes of motility

'Moonlighting proteins' refer to proteins with a characterized function 'by-day', but opt for another, unrelated function, independent from the initial characterized function 'by-night' (Jeffery, 1999). This also means that inactivation of one function does not affect the secondary function and vice versa. Many proteins capable of 'moonlighting' are primarily enzymes, whilst others can be receptors, ion channels, or chaperones. These events come though gene fusion and splice variants and are not found in one particular lineage (Rodríguez-Saavedra, et al., 2021), but in organisms ranging from multicellular eukaryotes (e.g *Spinacia oleracea* and *H. sapiens*) to single-celled prokaryotes such as *Staphylococcus aureus* and single-celled eukaryotes with free-living (*Entamoeba histolytica*) and parasitic lifestyles (*Leishmania infantum* and *T. vaginalis* (Ginger, 2014)).

In *L. infantum*, peroxiredoxins have a classical role in the detoxification of reactive oxygen species (the '2-Cys' class) however also have a 'moonlighting' chaperone role also associated with peroxidase activity, essential for virulence of this parasite. Mouse models with mitochondrial peroxiredoxin-null genotypes showed that *L. infantum* virulence was restored through expression of a cysteine-lacking, active-site mitochondrial peroxiredoxin (removing peroxidase activity), thereby indicating the moonlighting properties of *L. infantum* peroxiredoxin (Castro, et al., 2011). In *T. vaginalis*, candidate adherins (α- and β- subunits of succinyl-CoA synthetase, malic enzyme, and pyruvate: ferredoxin oxidoreductase (Meza-Cervantez, et al., 2011; Alderete, et al., 1995)) are found in the hydrogenosomes (mitochondrial-related organelles lacking carbon-compound utilizing capabilities) in abundance, utilizing pyruvate in a different manner with regard to typical mitochondrial based methods. However, failure to demonstrate dual localization and a clear pathway by which the hydrogenosomes attach to the cell surface (also given the organelle is double-membrane bound) raises scepticism of these enzymes being putative adherins (Hirt, et al., 2007; Ginger, 2014). Additionally, *Toxoplasma gondii* has shown aldolase, an enzyme commonly used to break-down specific complex sugars, possesses protein moonlighting functions by connecting surface-localized transmembrane adherins with actin filaments beneath the parasitic cell-surface, commonly used for α-motility of apicomplexan parasites

across cell surfaces and cell invasion (Jewett & Sibley, 2003; Sibley, 2010). As proposed by the *T. vaginalis* studies (Meza-Cervantez, et al., 2011; Alderete, et al., 1995) moonlighting proteins can be responsible for motility switching, therefore it could also be possible for the pseudopod formation genes to elicit a morphological switch, much in the same way as the secondary function of candidate adherins in *T. vaginalis* in species that have not shown a change in morphology, yet still retain all or some pseudopod formation genes.

Species	Movement			Citation
	Immobile/branching filaments	Flagellated/gliding	α/crawling	
Bodo saltans		✓		Gomaa, et al., 2022
Ectocarpus variabilis	✓		✓	MacKay & Gallant, 1991
Ectocarpus siliculosus	✓		✓	Jia, et al., 2017
Cyanidioschyzon merolae	✓			Ichinose & Iwane, 2021
Chondrus crispus	✓			Gutierrez & Fernandez, 1992
Galdieria phlegrea	✓			Čížková et al., 2019
Galdieria sulphuraria	✓			Čížková et al., 2019
Gracilariopsis chorda	✓			Lin, 2008
Porphyra umbilicalis	✓			Miranda, et al., 2013
Pyropia yezoensis	✓			Fukui, et al., 2014
Rozella allomycis	✓	✓		James, et al., 2006
Micromonas commoda		✓		Simon, et al., 2017
Micromonas pusilla		✓		Zingone, et al., 2006
Chlamydomonas reinhardtii		✓		Pröschold, et al., 2005
Cyanophora paradoxa		✓		Lino & Hashimoto, 2003
Glaucocystis nostochinearum		✓		Echlin 1967
Spizellomyces punctatus	✓	✓	✓	Gill & Fast, 2006
Mitosporidium daphniae	✓			Corsaro, et al., 2016
Encephalitozoon cuniculi	✓			Magalhães et al., 2022
Naegleria gruberi	✓	✓	✓	Forrester, et al., 1967
Gregarina niphandrodes	✓			Bessette & Williams 2022
Chromera velia	✓	✓		Oborník, et al., 2011
Vitrella brassicaformis		✓		Oborník et al., 2012
Ostreococcus tauri	✓			Subirana et al., 2013
Ostreococcus lucimarinus	✓			Subirana et al., 2013

Table 12: Differing modes of motility as described in literature. Types of motilities described in literature for each species revised in BLASTp investigation. Some species have life-cycle dependant motility (*e.g.,* zoospores of fungi species). ✓ = confirmation of mode of motility as some stage of the cell life-cycle.

4.4 Disruptions to lab-based experiments due to SARS-CoV-2

Lab closure over a period of time, due to the Covid pandemic, led to samples collected from the University of York being lost. As a result of this, RNA extractions and sequencing was unable to be taken any further forward and therefore the focus and aims of the project changed in order to try and identify, bioinformatically, pseudopod forming genes (Arp2/3, WASP, and SCAR/WAVE) in a range of different taxa. This analysis was coupled with comprehensive literature searches to link any findings with morphological characteristics and thereby identify potential organisms capable of switching modes of motility in response to environmental stress. Additionally, static confinement experiments, using live samples of Dark Stuff (that was maintained through the Covid pandemic) and light-microscopy were planned. However, training to use imaging equipment was limited due, in part, to rules put in place limiting the number of people allowed to be in close proximity to one another in the imaging suite.

4.5 Future work

Although our analyses do provide inferences as the identification of Dark Stuff, comprehensive EM that can focus on intracellular detail and characteristics, coupled with extensive literature searching, could provide a more robust morphological identification. In addition, and as mentioned previously, full-genome sequencing, potentially informed by the Darwin Tree of Life project would also comprehensively characterise and identify our cultured kinetoplastid. More broadly, future work could also encompass a more metagenomic approach, whereby communities are sampled and analysed, as opposed to single-celled cultures.

The bioinformatic identification of a number of species that, apparently, possess motility-related genes and therefore have the potential to switch from flagellate to amoeboid forms (and vice-versa) could be experimentally probed through cell confinement experiments (Brunet, et al., 2021). Cell confinement studies using *B. saltans*, which lives as a flagellated free-living kinetoplastid, and *C. velia*, a free-living photosynthetic organism (both of which returned all motility-related genes and therefore conceivably has the potential to switch motility mode), could be used in confinement experiments to see if morphological changes are indeed possible. Additional confinement experiments with, for example, *R. allomycis* which only retain some pseudopod formation genes, could ascertain the influence of other environmental conditions (light, nutrition deficits, temperature, and pH differences) on the morphological switching between amoeboid and flagellated or immotile forms. Finally, confinement experiments with *E. variabilis* (that interestingly did not appear to possess any of our motility genes) could provide inferences as to the potential for other moonlighting proteins to be involved in cell form and motility.

Bibliography

Adl, S. M., Leander, B. S., Simpson, A. G. B., et al., 2017. Diversity, Nomenclature, and Taxonomy of Protists, *Systematic Biology*, 56(4), pp.684-689. DOI:10.1080/10635150701494127

Aguiar-Pulido, V., Huang, W., Suarez-Ulloa, V., et al., 2016. Metagenomics, metatranscriptomics, and Metabolomics Approaches for Microbiome Analysis. *Evolutionary Bioinformatics Online*, 12(12s1), pp.5-16. PMID: 27199545, https://doi.org/10.4137/EBO.S36436

Albuquerque, P., Mendes, M. V., Santos, C. L., et al., 2009. DNA signature-based approaches for bacterial detection and identification. *Science of the Total Environment*, *407*(12), pp.3641-3651. https://doi.org/10.1016/j.scitotenv.2008.10.054

Alderete, J. F., O'Brien, L. O., Arroyo, R., et al., 1995. Cloning and molecular characterization of two genes encoding adhesion proteins involved in Trichomonas vaginalis cytoadherence. *Molecular Microbiology*, 17(1), pp.69–83. ISSN:0950-382X, https://doi.org/10.1111/j.1365-2958.1995.mmi_17010069.x

Alves, A. A., Gabriel, H. B., Bezerra, M. J. R., et al., 2020. Control of Assembly of extra-axonemal structures: The paraflagellar rod of trypanosomes. *Journal of Cell Science*, 133(10), p. jcs242271. ISSN: 1477-9137, https://doi.org/10.1242/jcs.242271

Bear, J.E., Rawls, J.F. & Saxe, C.L., 1998. SCAR, a WASP-related protein, isolated as a suppressor of receptor defects in late Dictyostelium development. *J. Cell Biol.* 142(5), pp. 1325-1335. ISSN: 1540-8140, https://doi.org/10.1083/jcb.142.5.1325

Behjati, S., & Tarpey, P. S., 2013. What is Next Generation Sequencing?, *Archives of disease in childhood- Education & practice edition,* 98(6), pp.236-238. DOI 10.1136/archdischild-2013-304340

Berdjeb, L., Parada, A., Needham, D. M., et al., 2018. Short-term dynamics and interactions of marine protist communities during the spring–summer transition. *ISME Journal* 12(8), pp. 1907–1917. ISSN: 1751-7370, https://doi.org/10.1038/s41396-018-0097-x

Bessette, E., & Williams, B., 2022. Protists in the Insect Rearing Industry: Benign Passengers or Potential Risk? *Insects*. 2022; 13(5), p. 482. PMCID: PMC9144225, PMID: 35621816, https://doi.org/10.3390/insects13050482

Brawley, S. H., Blouin, N. A., Ficko-Blean, E., et al., 2017. Insights into the red algae and eukaryotic evolution from the genome of Porphyra umbilicalis (Bangiophyceae, Rhodophyta). *Proceedings of the National Academy of Sciences*, *114*(31), pp.E6361-E6370. https://doi.org/10.1073/pnas.1703088114

Brunet, T., Albert, M., Roman, W., et al., 2021. A flagellate-to-amoeboid switch in the closest living relatives of animals. *Elife*. 15(10), p. e61037. PMID: 33448265; PMCID: PMC7895527, doi: 10.7554/eLife.61037

Burki, F., Roger, A., Brown, M. & Simpson, A., 2020. The New Tree of Eukaryotes. *Trends in Ecology & Evolution*, 35(1), pp.43-55. PMID: 31606140, https://doi.org/10.1016/j.tree.2019.08.008

Burki, F., Sandin, M.M. & Jamy, M., 2021. Diversity and ecology of protists revealed by metabarcoding. *Current Biology*, 31(19), pp. R1267-R1280. PMID: 34637739, DOI: 10.1016/j.cub.2021.07.066

Butenko, A., Opperdoes, F.R., Flegontova, O., et al., 2020. Evolution of metabolic capabilities and molecular features of diplonemids, kinetoplastids, and Euglenids. *BMC Biology*, 18(1), pp. 1-28. PMCID: PMC7052976, PMID: 32122335, doi: 10.1186/s12915-020-0754-1

Butenko, A., Hammond, M., Field, M., et al., 2021. Reductionist Pathways for Parasitism in Euglenozoans? Expanded Datasets Provide New Insights. *Trends in Parasitology*, 37(2), pp.100-116. ISSN: 1471-5007, PMID: 33127331, DOI: 10.1016/j.pt.2020.10.001

Carosi, G., Scaglia, M., Filice, G., & Willaert, E., 1977. A comparative electron microscope study of axenically cultivated trophozoites of free-living amoebae of the genus Acanthamoeba and Naegleria with special reference to the species of N. gruben (Schardinger 1899), N. fowleri (Carter 1970) and N. jadini (Willaert et Le Ray 1973). *Arch. Protistenkunde*, *119*(3), pp.264-273.

Carr, M., Leadbeater, B. S. C., Hassan, R., Nelson, M., 2008. Molecular phylogeny of choanoflagellates, the sister group to metazoa. *Proceedings of the National Academy of Sciences*, 105(43), pp.16641–16646. https://doi.org/10.1073/pnas.0801667105

Cavalier-Smith, T., 2003. The excavate protozoan phyla Metamonada Grasse emend. (Anaeromonadea, Parabasalia, Carpediemonas, Eopharyngia) and Loukozoa emend. (Jakobea, Malawimonas): Their evolutionary affinities and new higher taxa. *International Journal of Systematic and Evolutionary Microbiology*, 53(6), pp.1741–1758. PMID: 14657102, DOI: 10.1099/ijs.0.02548-0

Cavalier-Smith, T., & Chao, E. E., 2010. Phylogeny and evolution of apusomonadida (protozoa: apusozoa): new genera and species. *Protist*, *161*(4), pp.549-576. https://doi.org/10.1016/j.protis.2010.04.002

Castro, H., Teixeira, F., Romao, S., et al., 2011. Leishmania mitochondrial peroxiredoxin plays a crucial peroxidase-unrelated role during infection: insight into its novel chaperone activity. *PLoS pathogens*, *7*(10), p. e1002325. PMID: 22046130, PMCID: PMC3203189, https://doi.org/10.1371/journal.ppat.1002325

Cerón-Romero, M.A., Fonseca, M.M., de Oliveira Martins, L., et al., 2021. Phylogenomic Analyses of 2,786 Genes in 158 Lineages Support a Root of The Eukaryotic Tree of Life Between Opisthokonts and All Other Lineages. *Genome Biology and Evolution*, bioRxiv evac119, https://doi.org/10.1093/gbe/evac119

Clayton, C., 2016. Gene expression in Kinetoplastids. *Current Opinion in Microbiology*, 32, pp.46-51. PMID: 27177350, DOI: 10.1016/j.mib.2016.04.018

Choi, J., & Park, J. S., 2020. Comparative analyses of the V4 and V9 regions of 18S rDNA for the extant eukaryotic community using the Illumina platform. *Sci Rep,* 10(1), pp. 1-11. PMID: 32300168, PMCID: PMC7162856, https://doi.org/10.1038/s41598-020-63561-z

Čížková, M., Vítová, M. and Zachleder, V., 2019. The red microalga Galdieria as a promising organism for applications in biotechnology. *Microalgae—From Physiology to Application, IntechOpen, London*, *1*, p.17. http://dx.doi.org/10.5772/intechopen.89810

Collen, J., Cornish, M. L., Craigie, J., et al., 2014. Chondrus crispus–a present and historical model organism for red seaweeds. In *Advances in Botanical Research* (Vol. 71, pp. 53-89). Academic Press. https://doi.org/10.1016/B978-0-12-408062-1.00003-2

Corliss, J. O., 2001. Protozoan cysts and spores, *eLS preprint,* DOI: 10.1038/npg.els.0001934

Corsaro, D., Michel, R., Walochnik, J., et al., 2016. Molecular identification of *Nucleophaga terricolae* sp. nov. (Rozellomycota), and new insights on the origin of the Microsporidia. *Parasitol Res* 115(8), pp. 3003–3011. PMID: 27075306**, https://doi.org/10.1007/s00436-016-5055-9

Damasceno, J. D., Marques, C. A., Black, J., et al., 2021. Read, write, Adapt: Challenges and opportunities during Kinetoplastid genome replication. *Trends in Genetics*, 37(1), pp.21–34. PMID: 32993968, PMCID: PMC9213392, DOI: 10.1016/j.tig.2020.09.002

Davidson, A. J., & Insall, R. H., 2013. Scar/Wave: A Complex Issue. *Communicative & Integrative Biology*, 6(6), p. 1509. PMID: 24753786, PMCID: PMC3984289, DOI: 10.4161/cib.27033

d'Avila-Levy, C. M., Boucinha, C., Kostygov, A., et al., 2015. Exploring the environmental diversity of kinetoplastid flagellates in the high-throughput DNA sequencing era*. Memórias do Instituto Oswaldo Cruz*, 110(8), pp.956–965. PMID: 26602872, PMCID: PMC4708014, DOI: 10.1590/0074-02760150253

De Luca, P., Taddei, R., & Varano, L., 1978. *Cyanidioschyzon merolae*: a new alga of thermal acidic environments. *Webbia*, *33*(1), pp.37-44. https://doi.org/10.1080/00837792.1978.10670110

de Souza, W., de Carvalho, T. U., & Barrias, E.S., 2010. Ultrastructure of *Trypanosoma cruzi* and its interaction with host cells. In *American Trypanosomiasis Chagas Disease*, (pp.393-432). https://doi.org/10.1016/B978-0-12-801029-7.00018-6

Dias, F., Vasconcellos, L. R., Romeiro, A., et al., 2014. Transovum transmission of trypanosomatid cysts in the Milkweed bug, *Oncopeltus fasciatus*. *PloS one*, *9*(9), p. e108746. PMCID: PMC4178184, PMID: 25259791, https://doi.org/10.1371/journal.pone.0108746

di Cicco, M. R., Palmieri, M., Altieri, S., et al., 2021. Cultivation of the Acidophilic Microalgae Galdieria phlegrea with Wastewater: Process Yields. *International Journal of Environmental Research and Public Health*, *18*(5), p.2291. https://doi.org/10.3390/ijerph18052291

Dobáková, E., Flegontov, P., Skalický, T., Lukeš, J., 2015. Unexpectedly streamlined mitochondrial genome of the euglenozoan *Euglena gracilis*. *Genome Biology and Evolution*,

7(12), pp.3358–3367. PMID: 26590215, PMCID: PMC4700960, https://doi.org/10.1093/gbe/evv229

Doerder, F. P., & Brunk, C., 2012. Natural populations and inbred strains of Tetrahymena. *Methods Cell Biol*. (Vol. 109, pp. 277-300) Academic Press. PMID: 22444148, doi: 10.1016/B978-0-12-385967-9.00009-8.

Echlin, P., 1967. The biology of *Glaucocystis nostochinearum*: I. The morphology and fine structure. *British Phycological Bulletin*, 3(2), pp.225-239. https://doi.org/10.1080/00071616700650091

Faktorová, D., Dobáková, E., Peña-Diaz, P., & Lukeš, J., 2016. From simple to supercomplex: mitochondrial genomes of euglenozoan protists. *F1000 Faculty Rev-392*, 5, p.392. PMCID: PMC4806707, PMID: 27018240, doi: 10.12688/f1000research.8040.2

Flegontova, O., Flegontov, P., Malviya, S., et al., 2018. Neobodonids are dominant kinetoplastids in the Global Ocean. *Environmental Microbiology*, 20(2), pp.878–889. ISSN: 1462-2920 https://doi.org/10.1111/1462-2920.14034

Flegontova, O., Flegontov, P., Londoño, P. A. C., et al., 2020. Environmental determinants of the distribution of planktonic diplonemids and kinetoplastids in the oceans. *Environmental Microbiology*, 22(9), pp.4014–4031. PMID: 32779301, DOI: 10.1111/1462-2920.15190

Folgueira, C., & Requena, J.M., 2007. A postgenomic view of the heat shock proteins in kinetoplastids. *FEMS Microbiology Reviews*, 31(4), pp.359–377. DOI:10.1111/j.1574-6976.2007.00069.x

Forrester, J., Gingell, D., & Korohoda, W., 1967, Electrophoretic Polarity exhibited by Amoeboid Cells of *Naegleria gruberi*. *Nature,* 215(5108), pp.1409–1410. PMID: 6055469, https://doi.org/10.1038/2151409a0

Franzen, C., 2004. Microsporidia: how can they invade other cells?. *Trends in parasitology*, 20(6), pp.275-279. https://doi.org/10.1016/j.pt.2004.04.009

Fritz-Laylin L. K., Prochnik S. E., Ginger M. L., et al., 2010. The genome of *Naegleria gruberi* illuminates early eukaryotic versatility. *Cell*, 140(5), pp. 631-42. PMID: 20211133, doi: 10.1016/j.cell.2010.01.032.

Fritz-Laylin, L. K., Lord, S. J. & Mullins, R. D., 2017a. Our evolving view of cell motility. *Cell Cycle*, 16(19), pp.1735–1736. PMID: 28820330, PMCID: PMC5628642, DOI: 10.1080/15384101.2017.1360655

Fritz-Laylin, L. K., Lord, S. J., & Mullins, R.D., 2017b. WASP and SCAR are evolutionarily conserved in actin-filled pseudopod-based motility. *Journal of Cell Biology*, 216(6), pp.1673–1688. PMID: 28473602, PMCID: PMC5461030, DOI: 10.1083/jcb.201701074

Fukui, Y., Abe, M., Kobayashi, M., et al., 2014. Isolation of *Hyphomonas* Strains that Induce Normal Morphogenesis in Protoplasts of the Marine Red Alga *Pyropia yezoensis* . *Microb Ecol,* 68(3), pp. 556–566. PMID: 24840921, https://doi.org/10.1007/s00248-014-0423-4

Gill, E. E., & Fast, N. M., 2006. Assessing the microsporidia-fungi relationship: combined phylogenetic analysis of eight genes. *Gene*, *375(1)*, pp.103-109. PMID: 16626896, https://doi.org/10.1016/j.gene.2006.02.023

Ginger, M. L., 2014. Protein moonlighting in parasitic protists. *Biochemical Society Transactions*, 42(6), pp.1734–1739. ISSN: 0300-5127, https://doi.org/10.1042/BST20140215

Goley, E., & Welch, M., 2006. The arp2/3 complex: An actin nucleator comes of age. *Nature Reviews Molecular Cell Biology*, 7(10), pp.713–726. PMID: 16990851, https://doi.org/10.1038/nrm2026

Gomaa, F., Li, Z. H., Beaudoin, D. J., et al., 2022. Crispr /cas9-induced disruption of *Bodo saltans* paraflagellar rod-2 gene reveals its importance for cell survival. *Environmental Microbiology*, 24(7), pp. 3051–3062. PMID: 35099107 https://doi.org/10.1111/1462-2920.15918

Gregory, B., Rahman, N., Bommakanti, A., et al., 2019. The small and large ribosomal subunits depend on each other for stability and accumulation. *Life science alliance*, 2(2), p. e201800150. PMID: 30837296, PMCID: PMC6402506, https://doi.org/10.26508/lsa.201800150

Gutierrez, L. M., & Fernandez, C., 1992. Water motion and morphology in *Chondrus crispus* (rhodophyta) 1. *Journal of Phycology*, 28(2), pp.156–162. https://doi.org/10.1111/j.0022-3646.1992.00156.x

Hadziavdic, K., Lekang, K., Lanzen, A., Jonassen, I., Thompson, E.M. and Troedsson, C., 2014. Characterization of the 18S rRNA gene for designing universal eukaryote specific primers. *PloS one*, 9(2), p.e87624. DOI 10.1371/journal.pone.0087624

Hamann, E., Gruber-Vodicka, H., Kleiner, M., et al., 2016. Environmental *Breviatea* Harbour Mutualistic arcobacter epibionts. *Nature*, 534(7606), pp. 254–258. PMID: 27279223, PMCID: PMC4900452, DOI: 10.1038/nature18297

Hammond, M., Zoltner, M., Garrigan, J., et al., 2021. The distinctive flagellar proteome of *Euglena gracilis* illuminates the complexities of protistan flagella adaptation. *New Phytologist*, 232(3), pp.1323–1336. PMID: 34292600, DOI: 10.1111/nph.17638

Harmer, J., Yurchenko, V., Nenarokova, A., et al., 2018. Farming, slaving and enslavement: histories of endosymbiosis during kinetoplastid evolution. *Parasitology*, 145(10), pp.1311-1323. PMID: 29895336, DOI: 10.1017/S0031182018000781

Hejazi, M. A., Barzegari, A., Gharajeh, N. H., Hejazi, M. S., 2010. Introduction of A novel 18S rDNA gene arrangement along with distinct its region in the Saline Water Microalga Dunaliella. *Saline Systems*, 6(1), pp. 1-11. PMID: 20377865, PMCID: PMC2867797, DOI: 10.1186/1746-1448-6-4

Henriquez, F.L., Mooney, R., Bandel, T., et al., 2021. Paradigms of protist/bacteria symbioses affecting human health: Acanthamoeba species and *Trichomonas vaginalis*. *Frontiers in Microbiology*, 7(11), p. 616213. PMID: 33488560, PMCID: PMC7817646, DOI: 10.3389/fmicb.2020.616213

Hirt, R.P., Noel, C. J., Sicheritz-Ponten, T., et al., 2007. Trichomonas vaginalis surface proteins: A view from the genome. *Trends in Parasitology*, 23(11), pp.540–547. https://doi.org/10.1016/j.pt.2007.08.020

Hoare, C., & Wallace, F., 1966. Developmental Stages of Trypanosomatid Flagellates: a New Terminology. *Nature* 212(5068), pp. 1385–1386. https://doi.org/10.1038/2121385a0

Hu, M., & Polyak, K., 2006. Serial analysis of gene expression. *Nature Protocols*, 1(4), pp.1743-1760. PMID: 17487157, DOI: 10.1038/nprot.2006.269

Hughes, A. L., & Piontkivska, H., 2003. Phylogeny of Trypanosomatidae and Bodonidae (Kinetoplastida) based on 18S rRNA: evidence for paraphyly of Trypanosoma and six other genera. *Molecular biology and evolution*, *20*(4), pp.644-652. https://doi.org/10.1093/molbev/msg062

Ichinose, T.M., & Iwane, A.H., 2021. Long-term live cell cycle imaging of single *Cyanidioschyzon merolae* cells. *Protoplasma* 258(3), pp. 651–660 . PMID: 33580410, PMCID: PMC8052221, DOI: 10.1007/s00709-020-01592-z

Ivanoff, S. S., 1933. Stewart's Wilt Disease of Corn, with Emphasis on the Life History of Phytomonas stewartia in Relation to Pathogenesis. *J. Agric. Res, 47,* pp.749-770

Jackson, A., Otto, T., Aslett, M., et al., 2016. Kinetoplastid Phylogenomics Reveals the Evolutionary Innovations Associated with the Origins of Parasitism. *Current Biology*, 26(2), pp.161-172. PMID: 26725202, PMCID: PMC4728078, DOI: 10.1016/j.cub.2015.11.055

James, T.Y., Kauff, F., Schoch, C.L., et al., 2006. Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature*, *443*(7113), pp.818-822. https://doi.org/10.1038/nature05110

Jia, F., Ben Amar, M., Billoud, B. and Charrier, B., 2017. Morphoelasticity in the development of brown alga Ectocarpus siliculosus: from cell rounding to branching. *Journal of the Royal Society Interface*, *14*(127), p.20160596. https://doi.org/10.1098/rsif.2016.0596

Jeffery C. J., 1999. Moonlighting Proteins. *Trends Biochem. Sci.* 24(1), pp.8–11. PMID: 10087914, DOI: 10.1016/s0968-0004(98)01335-8

Jewett, T. J., & Sibley, L. D., 2003. Aldolase forms a bridge between cell surface adhesins and the actin cytoskeleton in apicomplexan parasites. *Molecular cell*, *11*(4), pp.885–894. PMID: 12718875, https://doi.org/10.1016/s1097-2765(03)00113-8

Kaneshiro, E., 1995. Amoeboid Movement, Cilia, and Flagella. *Cell Physiology Source Book*, pp.611-637. https://doi.org/10.1016/B978-0-12-656970-4.50051-8

Karpov, S. A., Vishnyakov, A. E., Moreira, D., & López-García, P., 2019. The ultrastructure of Sanchytrium tribonematis (Sanchytriaceae, fungi incertae sedis) Confirms its close relationship to Amoeboradix. *Journal of Eukaryotic Microbiology*, *66*(6), pp.892-898. https://doi.org/10.1111/jeu.12740

Katinka, M. D., Duprat, S., Cornillot, E., et al., 2001. Genome sequence and gene compaction of the eukaryote parasite Encephalitozoon cuniculi. *Nature*, *414*(6862), pp.450-453. https://doi.org/10.1038/35106579.

Kaufer, A., Ellis, J., Stark, D., & Barratt, J., 2017. The evolution of trypanosomatid taxonomy. *Parasites & Vectors*, 10(1), p.287. https://doi.org/10.1186/s13071-017-2204-7

Kaufer, A., Stark, D., & Ellis, J., 2019. Evolutionary Insight into the Trypanosomatidae Using Alignment-Free Phylogenomics of the Kinetoplast. *Pathogens*, 8(3), p.157. PMID: 31540520, PMCID: PMC6789588, DOI: 10.3390/pathogens8030157

Kaur, B., Záhonová, K., Valach, M., et al., 2020. Gene fragmentation and RNA editing without borders: eccentric mitochondrial genomes of diplonemids. *Nucleic Acids Research*, 48(5), pp.2694-2708. PMID: 31919519, PMCID: PMC7049700, DOI: 10.1093/nar/gkz1215

Keeling, P. J., 2004. Diversity and evolutionary history of plastids and their hosts. *American journal of botany*, *91*(10), pp.1481-1493. https://doi.org/10.3732/ajb.91.10.1481

Keeling, P. J., Burger, G., Durnford, D. G., et al., 2005. The Tree of Eukaryotes. *Trends in Ecology & Evolution*, 20(12), pp.670–676. PMID: 16701456, DOI: 10.1016/j.tree.2005.09.005

Keeling, P., 2019. Combining morphology, behaviour and genomics to understand the evolution and ecology of microbial eukaryotes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 374(1786), p.20190085. ISSN: 1471- 2970, DOI:10.1098/rstb.2019.0085

Kim, E., 2009. Postsynaptic development: Neuronal molecular scaffolds. *Encyclopedia of Neuroscience*, pp.817–824. DOI: 10.1016/B978-008045046-9.00360-0.

Kin K., & Schaap P., 2021. Evolution of Multicellular Complexity in The Dictyostelid Social Amoebas. *Genes*, 12(4), p. 487. https://doi.org/10.3390/genes12040487

Knight-Jones, E. W., & Walne, P. R., 1951. *Chromulina pusilla* Butcher, a dominant member of the ultraplankton. *Nature*, *167*(4246), pp.445-446. https://doi.org/10.1038/167445a0

Kostygov, A. Y., Dobáková, E., Grybchuk-Leremenko, A., et al., 2016. Novel trypanosomatid-bacterium association: *Evolution of endosymbiosis in action*. mBio, 7(2), p. e01985. PMID: 26980834, PMCID: PMC4807368, DOI: 10.1128/mBio.01985-15

Kostygov, A. Y., Frolov, A. O., Malysheva, M. N., et al., 2020. Vickermania gen. nov., trypanosomatids that use two joined flagella to resist midgut peristaltic flow within the fly host. *BMC biology*, *18*(1), pp.1-16. PMID: 33267865, PMCID: PMC7712620, DOI: 10.1186/s12915-020-00916-y

Lara, E., Moreira, D., Vereshchaka, A., & López-García, P., 2009. Pan-oceanic distribution of new highly diverse clades of deep-sea diplonemids. *Environmental microbiology*, *11*(1), pp.47-55. https://doi.org/10.1111/j.1462-2920.2008.01737.x

Lax, G., Eglit, Y., Eme, L., et al., 2018. Hemimastigophora is a novel supra-kingdom-level lineage of eukaryotes. *Nature*, 564(7736), pp.410-414. PMID: **30429611**, DOI: 10.1038/s41586-018-0708-8

Letcher, P.M., Longcore, J.E., Quandt, C.A., et al., 2017. Morphological, molecular, and ultrastructural characterization of *Rozella rhizoclosmatii*, a new species in Cryptomycota. *Fungal biology*, *121*(1), pp.1-10. PMID: 28007212, https://doi.org/10.1016/j.funbio.2016.08.008

Leander, B.S., 2020. Predatory protists. *Current Biology*, *30*(10), pp.R510-R516.

Lin, S. M., 2008. Morphological and phylogenetic studies of *Gracilariopsis chiangii*, new species (Gracilariaceae, Rhodophyta), an alga presently known as *Gracilaria chorda* in Taiwan. *Raffles Bulletin of Zoology Supplement*, *19*, pp. 19-26. Corpus ID: 86244315

Lino, M., & Hashimoto, H., 2003. Intermediate features of cyanelle division of *Cyanophora paradoxa* (Glaucocystophyta) between cyanobacterial and plastid division 1. *Journal of Phycology*, *39*(3), pp.561-569. https://doi.org/10.1046/j.1529-8817.2003.02132.x

Lowe, R., Shirley, N., Bleackley, M., et al., 2017. Transcriptomics technologies. *PLOS Computational Biology*, 13(5), p. e1005457. https://doi.org/10.1371/journal.pcbi.1005457

Lukeš, J., Guilbride, D. L., Votýpka, J., et al., 2002. Kinetoplast DNA network: Evolution of an improbable structure. *Eukaryotic Cell*, 1(4), pp.495–502. PMID: 12455998, PMCID: PMC117999, DOI: 10.1128/EC.1.4.495-502.2002

Lukeš, J., Wheeler, R., Jirsová, D., et al., 2018. Massive mitochondrial DNA content in diplonemid and kinetoplastid protists. *IUBMB Life*, 70(12), pp.1267-1274. https://doi.org/10.1002/iub.1894

Machida, R., & Knowlton, N., 2012. PCR Primers for Metazoan Nuclear 18S and 28S Ribosomal DNA Sequences. *PLoS ONE*, 7(9), p. e46180. PMID: 23049971, PMCID: PMC3458000, DOI: 10.1371/journal.pone.0046180

MacKay, R. M., & Gallant, J. W., 1991. Beta-tubulins are encoded by at least four genes in the brown alga *Ectocarpus variabilis*. *Plant Mol Biol* 17(3), pp.487–492. PMID: 1883999, https://doi.org/10.1007/BF00040642

Maddison, D. R., & Schulz, K. S., & Maddison, W. P., (eds.) 2007. The Tree of Life Web Project. *Zootaxa, 1668(1)*, pp. 19-40. DOI:https://doi.org/10.11646/zootaxa.1668.1.4. Internet address: http://tolweb.org

Magalhães, T. R., Pinto, F. F., & Queiroga, F. L., 2022. A multidisciplinary review about Encephalitozoon cuniculi in a One Health perspective. *Parasitol Res*. pp. 1-17 https://doi.org/10.1007/s00436-022-07562-z

Mansur-Pontes, C. L., de Moraes, M. H., Lückemeyer, D. D., et al., 2021. Differential expression and activity of arginine kinase between the American trypanosomatids Trypanosoma rangeli and Trypanosoma cruzi. *Experimental Parasitology*, 230, p. 108159. https://doi.org/10.1016/j.exppara.2021.108159

Massana, R., Labarre, A., López-Escardó, D., et al., 2020. Gene expression during bacterivorous growth of a widespread marine heterotrophic flagellate. *The ISME Journal*, 15(1), pp. 154-167. PMID: 32920602, PMCID: PMC7852580, DOI: 10.1038/s41396-020-00770-4

Maslov, D. A., & Simpson, L., 2007. Strategies of Kinetoplastid cryptogene discovery and analysis. *Methods in Enzymology*, (Vol. 424, pp.127–139). Academic Press. PMID: 17662839, https://doi.org/10.1016/S0076-6879(07)24006-6

Maslov, D. A., Votýpka, J., Yurchenko, V., Lukeš, J., 2012. Diversity and phylogeny of insect trypanosomatids: All that is hidden shall be revealed. *Trends in Parasitology*, 29(1), pp. 43–52. PMID: 23246083, DOI: 10.1016/j.pt.2012.11.001

Matt, G. & Umen, J., 2016. Volvox: A simple algal model for embryogenesis, morphogenesis and cellular differentiation. *Developmental Biology*, 419(1), pp. 99–113. PMID: 27451296, PMCID: PMC5101179, DOI: 10.1016/j.ydbio.2016.07.014

McWatters, D. C., & Russell, A. G., 2017. *Euglena* Transcript Processing. *Euglena: Biochemistry, Cell and Molecular Biology*, 979, pp. 141-158. ISBN: 978-3-319-54908-8, https://doi.org/10.1007/978-3-319-54910-1_8

Medkour, H., Varloud, M., Davoust, B., & Mediannikov, O., 2020. New Molecular Approach for the Detection of Kinetoplastida Parasites of Medical and Veterinary Interest. *Microorganisms*, 8(3), p.356. PMID: 32131458, PMCID: PMC7143920, DOI: 10.3390/microorganisms8030356

Meza-Cervantez, P., González-Robles, A., Cárdenas-Guerra, R., et al., 2011. Pyruvate: Ferredoxin oxidoreductase (PFO) is a surface-associated cell-binding protein in trichomonas vaginalis and is involved in trichomonal adherence to host cells. *Microbiology*, 157(12), pp.3469–3482. DOI:10.1099/mic.0.053033-0

Michaeli, S., 2015. The response of trypanosomes and other eukaryotes to ER stress and the spliced leader RNA silencing (SLS) pathway in Trypanosoma brucei. *Critical Reviews in Biochemistry and Molecular Biology*, *50*(3), pp.256-267. https://doi.org/10.3109/10409238.2015.1042541

Midha, S., Rigden, D. J., Siozios, S., et al., 2021. *Bodo saltans* (Kinetoplastida) is dependent on a novel paracaedibacter-like endosymbiont that possesses multiple putative toxin-antitoxin systems. *The ISME Journal*, 15(6), pp. 1680–1694. PMID: 33452479, PMCID: PMC8163844, DOI: 10.1038/s41396-020-00879-6

Miki, H., Suetsugu, S., & Takenawa T., 1998. WAVE, a novel WASP-family protein involved in actin reorganization induced by Rac. *EMBO J.* 17(23), pp.6932-6941. PMCID: PMC1171041, PMID: 9843499, doi: 10.1093/emboj/17.23.6932

Minge, M. A., Silberman, J. D., Orr, R. J., et al., 2009. Evolutionary position of breviate amoebae and the primary eukaryote divergence. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1657), pp.597-604. https://doi.org/10.1098/rspb.2008.1358

Miranda, L. N., Hutchison, K., Grossman, A. R., & Brawley, S. H., 2013. Diversity and abundance of the bacterial community of the red macroalga *Porphyra umbilicalis*: Did bacterial farmers produce macroalgae? *PLoS ONE*, 8(3). p. e58269. https://doi.org/10.1371/journal.pone.0058269

Mohibbullah, M., Abdul Hannan, M., Park, I. S., Moon, I. S., & Hong, Y. K., 2016. The edible red seaweed Gracilariopsis chorda promotes axodendritic architectural complexity in hippocampal neurons. *Journal of medicinal food*, *19*(7), pp.638-644. https://doi.org/10.1089/jmf.2016.3694

Moore, R. B., Oborník, M., Janouškovec, J., et al., 2008. A photosynthetic alveolate closely related to apicomplexan parasites. *Nature*, *451*(7181), pp.959-963. https://doi.org/10.1038/nature06635

Morgan-Smith, D., Garrison, C. E., & Bochdansky, A. B., 2013. Mortality and survival of cultured surface-ocean flagellates under simulated deep-sea conditions. *Journal of experimental marine biology and ecology*, *445*, pp.13-20. https://doi.org/10.1016/j.jembe.2013.03.017

Mukherjee, I., Hodoki, Y. & Nakano, S.-ichi, 2015. Kinetoplastid flagellates overlooked by universal primers dominate in the oxygenated hypolimnion of Lake Biwa, Japan. *FEMS Microbiology Ecology*, 91(8). DOI:10.1093/femsec/fiv083

Mukherjee, I., Hodoki, Y., Okazaki, Y., et al., 2019. Widespread dominance of kinetoplastids and unexpected presence of diplonemids in Deep Freshwater Lakes. *Frontiers in Microbiology*, 10, p.2375. PMID: 31681232, PMCID: PMC6805782, DOI: 10.3389/fmicb.2019.02375

Neuenschwander, S.M., Salcher, M.M. & Pernthaler, J., 2015. Fluorescence in situ hybridization and sequential catalyzed reporter deposition (2C-FISH) for the flow cytometric sorting of freshwater ultramicrobacteria. *Frontiers in Microbiology*, 68(6), pp. 3094-3101. PMID: 12039771, PMCID: PMC123953, DOI: 10.1128/AEM.68.6.3094-3101.2002

Nitla, V., Serra, V., Fokin, S.I., Modeo, L., Verni, F., Sandeep, B.V., Kalavati, C. and Petroni, G., 2019. Critical revision of the family Plagiopylidae (Ciliophora: Plagiopylea), including the description of two novel species, Plagiopyla ramani and Plagiopyla narasimhamurtii, and redescription of Plagiopyla nasuta Stein, 1860 from India. *Zoological Journal of the Linnean Society*, *186*(1), pp.1-45. https://doi.org/10.1093/zoolinnean/zly041

Obiol, A., Giner, C. R., Sánchez, P., et al., 2020. A metagenomic assessment of microbial eukaryotic diversity in the Global Ocean. *Molecular Ecology Resources*, 20(3), pp.718–731. PMID: 32065492, DOI: 10.1111/1755-0998.13147

Oborník, M., Vancová, M., Lai, D. H., et al., 2011. Morphology and ultrastructure of multiple life cycle stages of the photosynthetic relative of Apicomplexa, *Chromera Velia*. *Protist*, 162(1), pp.115–130. PMID: 20643580, https://doi.org/10.1016/j.protis.2010.02.004

Oborník, M., Modrý, D., Lukeš, M., et al., 2012. Morphology, ultrastructure and life cycle of *Vitrella Brassicaformis* n. sp., N. gen., a novel chromerid from the Great Barrier Reef. *Protist*, 163(2), pp.306–323. https://doi.org/10.1016/j.protis.2011.09.001

Ogbadoyi, E., Robinson, D., & Gull, K., 2003. A High-Order Trans-Membrane Structural Linkage Is Responsible for Mitochondrial Genome Positioning and Segregation by Flagellar Basal

Bodies in Trypanosomes. *Molecular Biology of the Cell*, 14(5), pp.1769-1779. PMID: 12802053, PMCID: PMC165075, DOI: 10.1091/mbc.e02-08-0525

Opperdoes, F.R., Butenko, A., Flegontov, P., et al., 2016. Comparative metabolism of free-living *Bodo saltans* and parasitic trypanosomatids. *Journal of Eukaryotic Microbiology*, 63(5), pp.657–678. PMID: 27009761, DOI: 10.1111/jeu.12315

Padilla, D. K., & Savedo, M. M., 2013. A systematic Review of Phenotypic Plasticity in Marine Invertebrate and Plant Systems. *Advances in Marine Biology,* pp.67-94. DOI 10.1016/b978-0-12-410498-3.00002-1

Palmer, J. D., Soltis, D. E., & Chase, M. W., 2004. The plant tree of life: an overview and some points of view. *American journal of botany*, *91*(10), pp.1437-1445. https://doi.org/10.3732/ajb.91.10.1437

Parkinson J., & Blaxter M., 2009. Expressed sequence tags: an overview. *Methods Mol Biol.* 533, pp.1-12. PMID: 19277571, DOI: 10.1007/978-1-60327-136-3_1.

Pérez-Juárez, H., Serrano-Vázquez, A., Lara, E., et al., 2018. Population dynamics of amoeboid protists in a tropical desert: seasonal changes and effects of vegetation and soil conditions. *Acta Protozoologica*, *57*(4), pp.231-242. doi:10.4467/16890027AP.18.017.10093

Persson, A., 2001. Proliferation of Cryptic Protists and Germination of Resting Stages from Untreated Sediment Samples with Emphasis on Dinoflagellates. *Ophelia*, 55(3), pp.151-166. DOI 10.1018/00785326.2001.10409482

Peyretaillade, E., Gonçalves, O., Terrat, S., et al., 2009. Identification of transcriptional signals in *Encephalitozoon cuniculi* widespread among Microsporidia phylum: support for accurate structural genome annotation. *BMC genomics*, *10*, p. 607. PMCID: PMC2803860, PMID: 20003517, https://doi.org/10.1186/1471-2164-10-607

Pizarro, J., Hills, T., Senisterra, G., et al., 2013. Exploring the *Trypanosoma brucei* Hsp83 Potential as a Target for Structure Guided Drug Design. *PLoS Neglected Tropical Diseases*, 7(10), p. e2492. PMID: 24147171, PMCID: PMC3798429, DOI: 10.1371/journal.pntd.0002492

Pollard, T.D., Earnshaw, W. C., Lippincott-Schwartz, J., 2008. Ch 38 Cellular Motility. *Cell biology*. Philadelphia: Saunders/Elsevier. ISBN: 13: 9781416022558

Pollitt, A. Y. & Insall, R. H., 2009. Wasp and Scar/wave proteins: The drivers of Actin Assembly. *Journal of Cell Science*, 122(15), pp. 2575–2578. PMCID: PMC2954249, PMID: 19625501, doi: 10.1242/jcs.023879

Pröschold, T., Harris, E. H., & Coleman, A. W., 2005. Portrait of a Species: Chlamydomonas reinhardtii, *Genetics 170(4)*, pp. 1601–1610. PMID: 15956662, PMCID: PMC1449772, DOI: 10.1534/genetics.105.044503

Prostak, S. M., Robinson, K. A., Titus, M. A., & Fritz-Laylin, L. K., 2021. The actin networks of chytrid fungi reveal evolutionary loss of cytoskeletal complexity in the fungal kingdom. *Current Biology*, 31(6), pp. 1192-1205, e6. PMCID: PMC8812817, NIHMSID: NIHMS1662691, PMID: 33561386, doi: 10.1016/j.cub.2021.01.001

Quackenbush, J., 2001. Computational analysis of microarray data . *Nat Rev Genet* 2(6), pp.418–427. https://doi.org/10.1038/35076576

Quince, C., Walker, A., Simpson, J., et al., 2007. Shotgun metagenomics, from sampling to analysis. *Nat Biotechnol* 35(9), pp.833–844. PMID: 28898207 https://doi.org/10.1038/nbt.3935

Raghukumar, S. (2017). Fungi: Characteristics and Classification. In: Fungi in Coastal and Oceanic Marine Ecosystems. Springer, Cham. DOI: 10.1007/978-3-319-54304-8_1

Rodríguez-Saavedra, C., Morgado-Martínez, L. E., Burgos-Palacios, A., et al., 2021. Moonlighting proteins: The case of the hexokinases. *Frontiers in Molecular Biosciences*, 9(8), p. 701975. PMID: 34235183, PMCID: PMC8256278, DOI: 10.3389/fmolb.2021.701975

Rodriguez, Y., Omoto, C. K., & Gomulkiewicz, R., 2007. Individual and population effects of eugregarine, Gregarina niphandrodes (Eugregarinida: Gregarinidae), on Tenebrio molitor (Coleoptera: Tenebrionidae). *Environmental Entomology*, *36*(4), pp.689-693. https://doi.org/10.1093/ee/36.4.689

Ruehle, M. D., Orias, E. and Pearson, C. G., 2016. Tetrahymena as a unicellular model eukaryote: genetic and genomic tools. *Genetics*, *203*(2), pp.649-665. https://doi.org/10.1534/genetics.114.169748

Ryder, P. V., Vistein, R., Gokhale, A., et al., 2013. The wash complex, an endosomal arp2/3 activator, interacts with the Hermansky–Pudlak syndrome complex BLOC-1 and its cargo phosphatidylinositol-4-kinase type iiα. Molecular Biology of the Cell, 24(14), pp.2269–2284. PMID: 23676666, PMCID: PMC3708732, DOI: 10.1091/mbc.E13-02-0088

Sanders, R. W., 2011. Alternative Nutritional Strategies in Protists: Symposium Introduction and a Review of Freshwater Protists that Combine Photosynthesis and Heterotrophy. *Journal of Eukaryotic Microbiology,* 58(3), pp.181-184. DOI: 10.1111/j.1550-7408.2011.00543.x

Santoferrara, L., Burki, F., Filker, S., et al., 2020. Perspectives from Ten Years of Protist Studies by High-Throughput Metabarcoding. *Journal of Eukaryotic Microbiology*, 67(5), pp.612-622.

Schnepf, E., Koch, W., & Deichgräber, G., 1966. Zur Cytologie und taxonomischen Einordnung von Glaucocystis. *Archiv für Mikrobiologie*, *55*(2), pp.149-174. https://doi.org/10.1007/BF00418636

Schoenle, A., Hohlfeld, M., Hermanns, K., et al., 2021. High and specific diversity of protists in the deep-sea basins dominated by diplonemids, kinetoplastids, ciliates and foraminiferans. *Commun Biol,* 4(1), pp. 1-10. https://doi.org/10.1038/s42003-021-02012-5

Schönknecht, G., Chen, W. H., Ternes, C. M., et al, 2013. Gene transfer from bacteria and archaea facilitated evolution of an extremophilic eukaryote. *Science*, *339*(6124), pp.1207-1210. DOI: 10.1126/science.1231707

Scott, M., Woolums, A., Swiderski, C., et al., 2021. Multipopulational transcriptome analysis of post-weaned beef cattle at arrival further validates candidate biomarkers for predicting

clinical bovine respiratory disease. *Scientific Reports*, 11(1). doi: 10.1038/s41598-021-03355-z

Shahi, S.K., Freedman, S.N. & Mangalam, A.K., 2017. Gut microbiome in multiple sclerosis: The players involved and the roles they play. *Gut Microbes*, 8(6), pp.607–615. https://doi.org/10.1111/jeu.12813

Sibley, L. D., 2010. How Apicomplexan parasites move in and out of cells. *Current Opinion in Biotechnology*, 21(5), pp. 592–598. PMID: 20580218, PMCID: PMC2947570, DOI:10.1016/j.copbio.2010.05.009

Silva, F. M., Kostygov, A. Y., Spodareva, V. V., et al., 2018. The reduced genome of *Candidatus Kinetoplastibacterium sorsogonicusi*, the endosymbiont of *Kentomonas sorsogonicus* (Trypanosomatidae): loss of the haem-synthesis pathway, *Parasitology*. Cambridge University Press, 145(10), pp. 1287–1293. doi: 10.1017/S003118201800046X.

Singh, P. K., Bourque, G., Craig, N. L., et al., 2014. Mobile genetic elements and genome evolution 2014. *Mobile DNA* 18(5), p. 26. PMID: 30117500, PMCID: PMC4363357, DOI: 10.1186/1759-8753-5-26

Simakova, A. V., Tokarev, Y. S., & Issi, I. V., 2018. A new microsporidium Fibrillaspora daphniae gn sp. n. infecting Daphnia magna (Crustacea: Cladocera) in Siberia and its taxonomic placing within a new family Fibrillasporidae and new superfamily Tubulinosematoidea (Opisthosporidia: Microsporidia). *Parasitology research*, *117*(3), pp.759-766. https://doi.org/10.1007/s00436-018-5749-2

Simon, N., Foulon, E., Grulois, D., et al., 2017. Revision of the genus *Micromonas* Manton et Parke (Chlorophyta, Mamiellophyceae), of the type species *M. pusilla* (Butcher) Manton & Parke and of the species *M. commoda* van Baren, Bachy and Worden and description of two new species based on the genetic and phenotypic characterization of cultured isolates. *Protist*, *168*(5), pp.612-635. PMID: 29028580, https://doi.org/10.1016/j.protis.2017.09.002

Skalický, T., Dobáková, E., Wheeler, R., et al., 2017. Extensive flagellar remodelling during the complex life cycle of Paratrypanosoma, an early-branching trypanosomatid. *Proceedings of the National Academy of Sciences*, 114(44), pp. 11757-11762. PMID: 29078369, PMCID: PMC5676924, DOI: 10.1073/pnas.1712311114

Šlapeta, J., López-García, P., & Moreira, D., 2006. Global dispersal and ancient cryptic species in the smallest marine eukaryotes. *Molecular biology and evolution*, *23*(1), pp.23-29. https://doi.org/10.1093/molbev/msj001

Song, E., Park, S., & Kim, S., 2019. Primers for complete chloroplast genome sequencing in *Magnolia*. *Applications in Plant Sciences*, 7(9), p. e11286. PMCID: PMC6764489, PMID: 31572627, doi: 10.1002/aps3.11286

Sorof-Uddin, M., & Cheng, Q., 2015. Recent application of biotechniques for the improvement of Mango Research. *Applied Plant Genomics and Biotechnology*, (pp.195–212). Woodhead Publishing. https://doi.org/10.1016/B978-0-08-100068-7.00012-4

Spang, A., Mahendrarajah, T. A., Offre, P., & Stairs, C. W., 2022. Evolving Perspective on the Origin and Diversification of Cellular Life and the Virosphere, *Genome Biology and Evolution*, 14(6), p. evac034. PMID: 35218347, PMCID: PMC9169541, DOI: 10.1093/gbe/evac034

Stechmann, A., & Cavalier-Smith, T., 2002. Rooting the eukaryote tree by using a derived gene fusion. *Science*, 297(5578), pp. 89–91. PMID: 12098695, DOI: 10.1126/science.1071196

Subirana, L., Péquin, B., Michely, S., et al., 2013. Morphology, genome plasticity, and phylogeny in the genus *Ostreococcus* reveal a cryptic species, *O. mediterraneus* sp. nov. (Mamiellales, Mamiellophyceae). *Protist*, 164(5), pp. 643–659. PMID: 23892412, DOI: 10.1016/j.protis.2013.06.002

Suganuma, Y., & Yamamoto, H., 1992. Conjugation in Tetrahymena: ultrastructure of the meiotic prophase of the micronucleus. *European journal of protistology*, *28*(4), pp.434-441. https://doi.org/10.1016/S0932-4739(11)80008-8

Tanifuji, G., Cenci, U., Moog, D., et al., 2017. Genome sequencing reveals metabolic and cellular interdependence in an amoeba-kinetoplastid symbiosis. *Scientific Reports*, 7(1), p. 11688. PMID: 28916813, PMCID: PMC5601477, DOI: 10.1038/s41598-017-11866-x

Tashyreva, D, Simpson A. G. B., Prokopchuk G., et al., 2022. Diplonemids – a review on "new" flagellates on the Oceanic Block. *Protist*, 173(2), p.125868. PMID: 35339983, https://doi.org/10.1016/j.protis.2022.125868

Tikhonenkov, D. V., Gawryluk, R., Mylnikov, A., & Keeling, P., 2021. First finding of free-living representatives of Prokinetoplastina and their nuclear and mitochondrial genomes. *Scientific Reports*, 11(1), pp. 1-21. PMID: 33536456, PMCID: PMC7859406, DOI: 10.1038/s41598-021-82369-z

Ueki, N., Ide, T., Mochiji, S., et al., 2016. Eyespot-dependent determination of the phototactic sign in Chlamydomonas reinhardtii. *Proceedings of the National Academy of Sciences*, *113*(19), pp.5299-5304. https://doi.org/10.1073/pnas.1525538113

Vávra, J., & Lukeš, J., 2013. Microsporidia and 'the art of living together'. *Advances in parasitology*, *82*, pp.253-319. https://doi.org/10.1016/B978-0-12-407706-5.00004-6

Velle, K. B., & Fritz-Laylin, L. K., 2020. Conserved actin machinery drives microtubule-independent motility and phagocytosis in *Naegleria*. Journal of Cell Biology, 219(11), p. e202007158. PMID: 32960946, PMCID: PMC7594500, DOI: 10.1083/jcb.202007158

Visvesvara, G. S., Moura, H., & Schuster, F.L., 2007. Pathogenic and opportunistic free-living amoebae: Acanthamoeba spp., Balamuthia mandrillaris, Naegleria fowleri, and Sappinia diploidea. *FEMS Immunology & Medical Microbiology*, *50*(1), pp.1-26. PMID: 17428307, DOI: 10.1111/j.1574-695X.2007.00232.x

Viswanadha, R., Sale, W. S., & Porter, M. E., 2017. Ciliary motility: Regulation of Axonemal Dynein Motors. *Cold Spring Harbor Perspectives in Biology*, 9(8), p. a018325. PMID: 28765157, PMCID: PMC5538414, DOI: 10.1101/cshperspect.a018325

Von der Heyden, S., & Cavalier-Smith, T., 2005. Culturing and environmental DNA sequencing uncover hidden kinetoplastid biodiversity and a major marine clade within ancestrally freshwater Neobodo designis. *International journal of systematic and evolutionary microbiology*, *55*(6), pp.2605-2621. DOI 10.1099/ijs.0.63606-0

Vonlaufen, N., Kanzok, S. M., Wek, R. C., & Sullivan Jr, W. J., 2008. Stress response pathways in protozoan parasites. *Cellular Microbiology*, 10(12), pp.2387-2399. PMID: 18647172, DOI: 10.1111/j.1462-5822.2008.01210.x

Votýpka, J., Kostygov, A. Y., Kraeva, N., et al., 2014. Kentomonas gen. n., a new genus of endosymbiont-containing trypanosomatids of *Strigomonadinae* subfam. n. *Protist*, 165(6), pp.825–838. PMID: 25460233, DOI: 10.1016/j.protis.2014.09.002

Wheeler, R. J., Gluenz, E., & Gull, K., 2013. The Limits on Trypanosomatid Morphological Diversity. *PLoS ONE*, 8(11), p.e79581. PMCID: PMC3834336, PMID: 24260255, doi: 10.1371/journal.pone.0079581

Whittaker, R. H., & Margulis, L., 1978. Protist classification and the Kingdoms of Organisms, *Biosystems*, 10(1-2), pp.3-18. DOI: 10.1016/0303-2647(78)90023-0

Worden, A. Z., 2006. Picoeukaryote diversity in coastal waters of the Pacific Ocean. *Aquatic Microbial Ecology*, *43*(2), pp.165-175. Doi: 10.3354/ame043165

Yabuki, A., Tanifuji, G., Kusaka, C., et al., 2016. Hyper-eccentric structural genes in the mitochondrial genome of the algal parasite *Hemistasia phaeocysticola*. *Genome Biology and Evolution 8(9)*, pp.2870-2878. PMID: 27566761, PMCID: PMC5630924, DOI: 10.1093/gbe/evw207

Yamaguchi, A., Yubuki, N., & Leander, B. S., 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(1), pp. 12-16. PMID: 22401606, PMCID: PMC3374381, https://doi.org/10.1186/1471-2148-12-29

Yazaki, E., Yabuki, A., Imaizumi, A., et al., 2022. The closest lineage of Archaeplastida is revealed by phylogenomics analyses that include Microheliella maris. *Open Biology*, *12*(4), p.210376. https://doi.org/10.1098/rsob.210376

Yu, B., Niu, J., Feng, J., et al., 2018. Regulation of ferredoxin-NADP+ oxidoreductase to cyclic electron transport in high salinity stressed Pyropia yezoensis. *Frontiers in plant science*, *9*, p.1092. https://doi.org/10.3389/fpls.2018.01092

Zakharova, A., Saura, A., Butenko, A., et al., 2021. A New Model Trypanosomatid, *Novymonas esmeraldas*: Genomic Perception of Its "*Candidatus Pandoraea novymonadis*" Endosymbiont. *mBio* 12(4), p. e0160621. PMID: 34399629, PMCID: PMC8406214, DOI: 10.1128/mBio.01606-21.

Zhang, Q., Simpson, A., & Song, W., 2012. Insights into the phylogeny of systematically controversial haptorian ciliates (Ciliophora, Litostomatea) based on multigene analyses. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1738), pp.2625-2635. https://doi.org/10.1098/rspb.2011.2688

Zhou, X., Ren, L., Li, Y., et al., 2010. The next-generation sequencing technology: A technology review and future perspective. *Sci. China Life Sci.* 53(1), pp.44–57. PMID: 20596955, https://doi.org/10.1007/s11427-010-0023-6

Zingone, A., Natale, F., Biffali, E., et al., 2006. Diversity in morphology, infectivity, molecular characteristics and induced host resistance between two viruses infecting *Micromonas pusilla*. *Aquatic Microbial Ecology*, *45*(1), pp.1-14. ISSN: 1616-1564, DOI:10.3354/ame045001

Appendix

Appendix 1

Sample Site	Sample Name	Sample Number
HE-DP	Deposition Pond	11
		12
HW-D	Derwent	2
		10
HE-L	Lake	3
		4
HE-R	Reedbed	7
		8
HW-BL	Biology Lake	7
		10

Table 13: Original samples used in identification experiments from the University of York, of which only four samples having yielded results.

Appendix 2

Figure 34: Geographical locations of samples sites from the University of York, indicating where samples were originally isolated from.

Copywrite Statement

i. The author of this thesis (including any appendices and/ or schedules to this thesis) owns any copyright in it (the "Copyright") and s/he has given The University of Huddersfield the right to use such Copyright for any administrative, promotional, educational and/or teaching.

ii. Copies of this thesis, either in full or in extracts, may be made only in accordance with the regulations of the University Library. Details of these regulations may be obtained from the Librarian. This page must form part of any such copies made.

iii. The ownership of any patents, designs, trademarks and any and all other intellectual property rights except for the Copyright (the "Intellectual Property Rights") and any reproductions of copyright works, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property Rights and Reproductions cannot and must not be made available for use without permission of the owner(s) of the relevant Intellectual Property Rights and/or Reproductions