University of HUDDERSFIELD

University of Huddersfield Repository

Armitage, Paul

The choanoflagellate translational machinery

Original Citation

Armitage, Paul (2019) The choanoflagellate translational machinery. Masters thesis, University of Huddersfield.

This version is available at http://eprints.hud.ac.uk/id/eprint/34849/

The University Repository is a digital collection of the research output of the University, available on Open Access. Copyright and Moral Rights for the items on this site are retained by the individual author and/or other copyright owners. Users may access full items free of charge; copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational or not-for-profit purposes without prior permission or charge, provided:

- The authors, title and full bibliographic details is credited in any copy;
- A hyperlink and/or URL is included for the original metadata page; and
- The content is not changed in any way.

For more information, including our policy and submission procedure, please contact the Repository Team at: E.mailbox@hud.ac.uk.

http://eprints.hud.ac.uk/

University of HUDDERSFIELD
Inspiring tomorrow's professionals

The choanoflagellate translational machinery

Paul Armitage

Supervisor 1: Dr Martin Carr

Supervisor 2: Dr Georgios Psakis

MSc by Research

I confirm that, unless indicated otherwise, I undertook all the work presented in this report. Any work performed by other parties is fully acknowledged.

Acknowledgements

I would like to take this opportunity to thank the people that went out of their way to help me. Firstly, Dr Martin Carr, I cannot thank you enough for what you have done for me over the years. Your knowledge and expertise in this area has not only helped guide me and helped me obtain the grades I have throughout my academic career. But also pushed me to, or at least try to, understand things as in depth as you do. Next, I would like to thank Dr Georgios Psakis. Again, you are the same as Dr Carr, your knowledge of the subjects you supervised me in, is second to none. You pushed me and supported me in everything I carried out in the lab and beyond. The ability to share knowledge and advise, not just me, but all your students, inspires many people to be better at what they do. I hope you both realise what special people you are and continue to inspire the next groups of students as you did for my cohort and me.

I would like to thank my wife also, whom without her support, this project would not have happened. For being there when my first four months of data got stolen, I would have had a breakdown if you were not there for me. Also, for raising our children on your own, as I try to provide a better future. Without you I would not be where I am today. Finally, I would like to thank my little brother, Ben, without you to help me and inspire me, I would not have made it through my academic journey. I hope to one day repay you all for the opportunities and support you have provided for me.

Abstract

Introduction

Choanoflagellates are the closest living, single celled eukaryotic relative to the Metazoa. As such, they provide an ideal model organism to investigate evolution of multicellular life. Over the past 20 years, the advancements in DNA sequencing technologies has provided the means to investigate evolution at the molecular level easier and in greater depth. One such evolution event, is the distribution of the elongation factors eEF1a and eEFL. The apparent random distribution of the two usually mutually exclusive proteins has had many studies undertaken. These studies have suggested what evolutionary forces have meant species have evolved the use of one factor over the other. It has also been identified that certain species, including choanoflagellate species, do in fact harbour and transcribe both genes. This study provides a new insight into the translational machinery makeup, dependant on which elongation factor is used by a species. The choanoflagellates are almost unique amongst other species analysed, in that, six species are known to harbour both elongation factors. This allows direct comparisons, within the same taxonomic group, to identify specific uses of certain components of the known translational machinery, dependant on which elongation factor is used by a specific species.

Methodology

Using a bioinformatical approach gene loss and gain was investigated along with gene use between the species. As well as investigating the eEF1a and eEFL genes, the eEF2, eEF3, eEF1bα, eEF1bγ and eEF1bδ genes were analysed. A small selection of ribosomal genes was also analysed to identify further areas of research in the area. Finally, a selection of ABCF genes were identified and analysed for distribution within the choanoflagellates. This study also investigated the use of EF1a in the dual encoding species *Salpingoeca punica*. Using a molecular biology approach the gene was successfully cloned into *Escherichia coli* cells for use in further protein works.

Results and conclusion

In this study, it was found that distinct patterns in certain genes were apparent, dependant on the elongation factor used. The identification of four ABCF genes within the choanoflagellates was found and the distribution of these proteins was solved phylogenetically. This included three horizontal gene transfer events in the ABCF2, ABCF4 and ABCF6 proteins. The small selection of ribosomal proteins indicated that the large ribosomal sub unit RPL5 showed an absence in the eEF1a species showing distribution that is elongation factor specific. This study was a broad investigation into mechanisms of translational machinery and evolution within the choanoflagellates. This allows identification for more in depth and specific studies in the future.

Table of contents

1.0 Introduction

1.1 The process of translation

Translation is the process of producing proteins essential for cellular function from the mRNA of cells. The biosynthesis of proteins from the genomic data held within the DNA of every organism is a process undertaken by a complex system of enzymes and factors (Sasikumar et al., 2012). This ancient process has been conserved throughout evolution, which is shown in its presence from prokaryotic to multicellular life. This sophisticated process involves the intricate formation of many molecular complexes to undertake the production of the proteins encoded within the DNA. Translation has four key stages, these being initiation, elongation, termination and ribosomal recycling (Dever et al., 2016). The initiation stage of translation is carried out by 12 known eukaryotic initiation factors (eIFs). This complex system includes the formation of the 12 eIFs into different complexes enabling the assembly of a functional 80S ribosome on to a mRNA molecule, which in turn allows translation to occur (Budkevich et al., 2008). The third stage of the process is termination, in eukaryotes this stage is carried out by two release factors (eRFs) eRF1 and eRF3. eRF1 recognises the three stop codons UAA, UAG and UGA (Ito et al., 2002). This then works in unison with eRF3 which in turn enables both separation of peptide and ribosome, as well as the ribosome and mRNA molecule (Denis et al., 2017). The final stage in translation is ribosomal recycling. Recycling of the ribosome after translation of the mRNA, allows the cyclic process to continue. The process is currently poorly understood at present, but does have one key factor, the ATPase ABCE1 (Neurenberg & Tampe, 2012). These three stages of translation are complex enough to warrant their own studies, as such, this study will be focussed mainly on the second stage, elongation.

1.2 The function and distribution of eEF1a and eEFL

The elongation phase refers to the cyclical process of adding a singular amino acid and the subsequent lengthening of the nascent peptide chain (Sasikumar et al., 2012). In eukaryotes, this process is performed by either eukaryotic elongation factor 1a (eEF1a), formerly known as eEF1α, or it's paralogue, eukaryotic elongation factor like (eEFL) enzymes, which form the eukaryotic elongation 1 complex. This complex is formed along with eukaryotic elongation factor 1b complex. Other elongation factors involved in elongation are eukaryotic elongation factor 2 (eEF2) and in fungi a third factor is required eukaryotic elongation factor 3 (eEF3). The EF1a protein was initially thought to be the only elongation factor involved in this process, but an absence of eEF1a in certain species led to further enquiries into this subject. How did species with no eEF1a, translate proteins? This was answered in a study by Keeling & Inagaki (2004), they showed how species including dinoflagellates, green algae, fungi and indeed the choanoflagellates, did not possess the eEF1a gene. Instead they had a GTPase, that although similar to eEF1a, showed distinct differences. This gene was coined eukaryotic Elongation Factor like (eEFL) (Keeling & Inagaki, 2004). As EFL contains domain structures which equate to EF1a indicates a presumable equivalence of functionality. This coupled with the usually mutual exclusivity of EF1a or EFL within species, gives a wide and random distribution between the two genes (Kamikawa et al., 2013). This random distribution, even between different species of the same taxon, sometimes deviates from known phylogenies and has given to explanations for the distribution. This pattern has been explained by horizontal gene transfer (Kamikawa et al., 2008), or lineage sorting (Gile et al., 2009). The distribution of EFL and EF1a has now been shown to be far more complex than was first understood.

EF1a has been studied in great depth (Sadritdinova et al., 2013; Serdyuk & Galzitskaya, 2007; Zhang et al., 2014) and its mechanics during translation are well understood, showing four individual domains

within the protein (Perez & Kinzy 2014). In its GTPbound form the eEF1a is able to bind aminoacyl-tRNA's (aa-tRNA) and then deliver them to the A-site of the ribosome. If a correct codon-anticodon pair is formed, this results in GTP hydrolysis, ribosomal conformational change, lengthening of the peptide chain and the subsequent release of eEF1a in its GDP state (Sasikumar et al., 2012). For this cyclic process to continue, GDP needs to be replaced with GTP in eEF1a for the process to continue. A study on the kinetics of EF1a in *Saccharomyces cerevisiae* by Gromadski et al. (2007) found that this process for GDP-GTP exchange was a slow process for EF1a on its own. The study went on to show that, on its own, EF1a carries out the exchange at a rate of 0.1 s^{-1}. This would not provide efficient translation of a protein and would hold up cellular processes. EF1a species have been shown to utilise an independent guanine exchange factor (GEF), found in the elongation factor 1b complex, which can increase the process by up to 320-fold and as such increases the rate of exchange to near 6s^{-1}. This value is compatible with cellular protein synthesis rates (Gromadski et al., 2007). The eEF1a/EFL and the eEF1b complex form the larger elongation factor 1 complex and seem to provide an essential function for cellular processes. The mechanics of EFL are less studied and as such poorly understood. eEFL is considered to carry out the role of eEF1a in species which do not harbour eEF1a, as mentioned earlier. The main difference between the two enzymes is that eEFL does not require the eEF1b complex for guanine exchange, as no eEFL species has been found to harbour the gene for one of the components of the eEF1b complex, eukaryotic elongation factor 1b alpha (eEF1bα) (Atkinson et al., 2014). As studies on eEFL are lacking, it is not known if this enzyme requires an unknown guanine exchange factor, or self-recharges in its GDP to GTP states.

1.3 The components and function of the eEF1b complex

The eEF1b complex has three components, the afore mentioned eEF1bα and the eukaryotic elongation factor 1b gamma (eEF1bγ) and in certain organisms eukaryotic elongation factor 1b delta (eEF1bδ). The role of the eEF1b complex has been found to have one GEF in all EF1a species in eEF1bα, but in multicellular organisms a second GEF has been identified in eEF1δ (Sasikumar et al., 2012). The third factor involved in this complex is eEF1bγ. This protein appears to provide a scaffolding role but is not essential for translation, a study by Kinzy et al. (2004) showed how deletion of the gene from *Saccharomyces cerevisiae* had little effect on protein synthesis. With little to no studies on the EFL gene and the subsequent proteins mechanics during translation available, little data is known about the role of the EF1b complex in EFL species.

1.4 The function of eukaryotic elongation factor 2

The GTPase protein eEF2 has been shown to be an essential component of the translational machinery, and has been used as a phylogenetic marker, as it is highly conserved (Kim & Graham, 2008). It is found in all kingdoms of life, with the bacterial EF-G and archaic EF2 being homologous with eEF2 for both the functional and structural elements of the protein (Kaul et al., 2011). For example, between mammalian species, the gene shows > 99% identity (Rapp et al., 1989). The role of eEF2 is to catalyse the translocation of the ribosome to allow peptide elongation (Susorov et al., 2018). The coordinated movement of the mRNA molecule and two tRNA molecules during peptide elongation is provided by eEF2 through a conformational change driven by GTP hydrolysis (Bartish et al., 2008).

1.5 The translational ATPase proteins

The ABC superfamily has identified that it is not only GTPase' that are involved in translation. Many of the ABC supergroup of proteins are crucial to the translational process and these proteins are ATPase's (Murina et al., 2017). The nomenclature of these proteins has been developed by an alphabetical approach, from the human equivalent proteins, and are named ABCA through to ABCH (Dean et al.,

2001). The ABCF proteins lack the fusion to membrane spanning domains of the other ABC families and as such the usual transporter function of this superfamily appears to be lost in the ABCF subgroup (Sharkey et al., 2016). Eukaryotic elongation factor 3 (eEF3) is an ATPase which belongs to the ABCF group of proteins (Kerr, 2004). In fungi, eEF3 has been shown to be a vital component of translation and without it, translation cannot happen (Mateyak et al., 2018). It appears that through the conclusion of two independent studies, eEF3 is a ribosomal dependent ATPase. In the study by Skogerson & Wakatama (1976), an early insight into eEF3 showed how in-vitro assays for translation elongation, required eEF3 when using ribosomes purified from yeast. An alternative study found, eEF1a and EF2 isolated from *Saccharomyces cerevisiae* could catalyse the translational process, successfully, with ribosomes from rat liver, a species which does not harbour the EF3 gene (Skogerson & Engelhardt, 1977). This led to further investigations into the eEF3 gene, to help further understand the process. Although not yet fully understood, studies have shown variable uses for the enzyme during translation within fungi, including the facilitation in the release of de-acetylated tRNA from the ribosomal E site (Triano-Alonso et al., 1995). Also, the EF3 enzyme has a significant role in the stimulation of EF1a to aminoacyl-tRNA (aa-tRNA) binding and its subsequent transfer to the ribosomal A-site in yeast (Uritani & Miyazaki, 1998).

As early studies into this area suggested that eEF3 was a fungal specific protein (Merrick, 1992), until the finding that eEF3 orthologs were found in several non-fungal species. One such study is the work by Ebstrup et al. (2005). Whilst carrying out proteomic work on the species *Phytophthora infestans*, this study found direct evidence of the eEF3 protein in 2d gels. This showed direct evidence that the EF3 gene is not fungal specific. It led to the question, can the eEF3 gene be found in any other species? As yet, the understanding of the role of eEF3, within non-fungal species, is not known. But Murina et al. (2017) found that the domain structure of eEF3 has been conserved in non-fungal species and as such suggests that the binding to the ribosome and eEF1a functions, may also have been conserved. This hypothesis was through eEF3 genes showing the same HEAT domains and C-terminus tract that contains polylysine and arginine rich areas as in the fungal eEF3 genes. Due to this, the Murina et al. (2017) study coined the EF3 gene in non-fungal species as eukaryotic elongation factor 3 like (eEF3L).

With transmembrane transportation properties not seen in the ABCF family, the roles of these proteins is still under consideration within the biological and medicinal fields. As yet, the role of the ABCF family within cellular processes, is poorly understood. With so many members of the ABCF family, the study by Murina et al. (2017) found 45 distinct sub-families within the ABCF group, it will take several studies to understand the roles of all singular members. Other ABCF proteins have been identified to play a role during translation. The role of ABCF1 has been shown to promote initiation of translation in eukaryotic cells (Paytubi et al., 2009) as well as regulation of the immune response and the development of embryos in mice (Wilcox et al., 2017). Another role of the ABCF family is in biogenesis of the ribosome. ABCF2 has been shown to be involved in this area in a study by Dong et al. (2005). In which they showed the absence of ABCF2 in *S. cerevisiae* disrupted the biogenesis of the ribosomes within the organisms.

1.6 The ribosomal proteins

The ribosome is an integral part of the translational machinery, used as the main site of protein biogenesis. In eukaryotes, for a ribosome to form, a 43S pre-initiation complex, formed by the eIFs, the 40S ribosomal subunit and methionyl-tRNAi are recruited to the mRNA (Gobet & Naef, 2017). Once the initiation codon (AUG) is recognised by the pre-initiation complex, the 60S ribosome is recruited. This then allows the 60S ribosome subunit to be recruited and the subsequent formation of the 80S ribosomal complex is formed, thus leading to translation of the mRNA (Hinnebusch & Lorsch, 2012).

The number of ribosomal proteins involved in the forming of the 80s ribosome is largely species specific (Blanchard et al., 2004). With many species having different make ups of the entire protein complex, for example, the L11 ribosomal arm is formed when the L10 ribosomal protein forms a connection with four to six copies of the ribosomal protein L7/L12, but this is dependent on the species involved (Moore, 2009). As the formation of the ribosomal complex is factored independently between species and with so many independent factors and proteins affecting the ribosomal complex, each species and/or taxonomic group, must be investigated individually to establish a firmer understanding of the ribosomal formational process.

1.7 Choanoflagellates as a model organism

As most studies on translation are angled toward specific factors within the process, this study will approach the subject from a new angle. In this study the translational machinery will be analysed dependant on which elongation factor a species uses. This will illustrate if any differences have evolved in the translational machinery, dependant on the use of EF1a or EFL within the species. For this study choanoflagellates will be used as the model organism, due to two factors. Firstly, the positioning of choanoflagellates in the tree of life and secondly the species EF1a and EFL distribution within the taxonomic group.

Choanoflagellates are single celled organisms which are positioned within the Holozoa lineage. Which in turn belongs to the super group of the Opisthokhonta (Adl et al., 2012). Choanoflagellates are ubiquitous to aquatic environments (Carr et al., 2017) and are bacterivorous, consuming their prey through phagocytic mechanisms (Wylezich et al., 2012). The choanoflagellates have been shown to hold a special position regarding the evolution of the metazoan kingdom. Figure 1.1 shows the phylogeny produced by the Brown et al. (2018) study, which clearly shows the choanoflagellate species *Monosiga brevicollis* as a sister group to the metazoan kingdom with the *Homo sapiens* and *Amphimedon queenslandica* branch. The positioning of the choanoflagellates within the super group of the opistokhonts, has led to the increased interest for evolutionary studies within the taxonomic group (Richter & King, 2013; Nitsche et al., 2011).

Secondly, choanoflagellates appear to be the ideal model organism for questions regarding the evolution and use of translational machinery, dependant on the elongation factor (EF1a or EFL) used in the making of the translational complex 1. This is due to the 21 species with genomic or transcriptomic data available to this study, four have the EF1a gene, eleven have the EFL gene, but most surprisingly six have both genes, these will be referred to as dual species from here on in. The dual species could shed further light on the evolution, loss and/or gain of EFL and EF1a genes, with direct comparisons of factors within one taxonomic group. This may help to understand how the Metazoa have evolved the use of only EF1a, losing the function of EFL during their evolution.

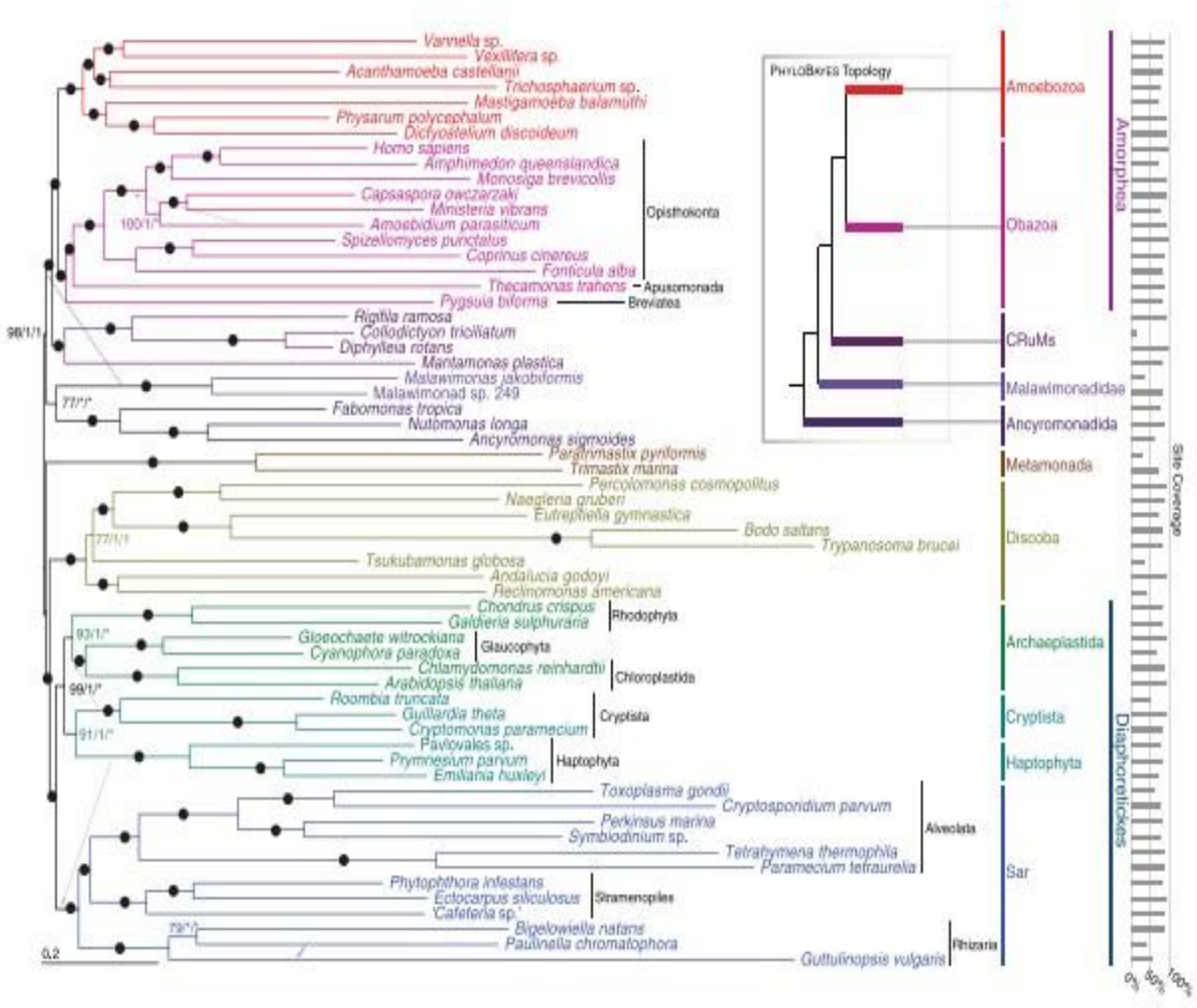


Figure 1.1: The evolutionary relationship between the choanoflagellates and the Metazoa. A phylogeny produced by Brown et al. (2018) by using 61 eukaryotic species, with the resulting alignment inferred from 351 proteins. Numbers on the branches show bootstrap values (1st) and posterior probabilities (2nd). Asterix represent unrecovered clade, full circles represent maximum support for both values. Phylogeny shows the choanoflagellate species *M. brevicollis* as a sister species to the Metazoa.

A study by Atkinson et al. (2014) produced a hypothesis for this process, which coincides with this study, due to the choanoflagellate elongation factor lineages. Their hypothetical model for evolutionary mechanism driving mutual exclusivity of eEF1a and eEFL states, suggests that the functional cycle of the two proteins may drive the evolution of the elongation factors. The guanine exchange factor EF1bα has not yet been identified in an EFL species. As previously stated, EF1a requires EF1bα to function at a high enough rate of exchange. It therefore appears that EFL do not need EF1bα to function. Their study showed association of EFL presence with the loss of both EF1a and EF1bα. By using this hypothesis and applying it to this study, it may be shown how this process is being applied to species that have either EFL, EF1a and in species that have both. By comparing the evidence between all three sub-groups and their comparative genes, further information may be shown to support this hypothesis and provide new insights into the differences in the translational machinery make-up from the different elongation factor species.

A bioinformatical approach will be utilised to investigate gene loss and gain throughout the choanoflagellate species, which currently have genomic or transcriptomic data available. By investigating such gene loss or gain, and then comparing the results between the elongation specific

species, we hope to identify any patterns of distribution that appear to be elongation factor species specific. The main area of focus will be in the translational machinery during the elongation phase. Secondly, a molecular biology approach to the dual encoding species, *S. punica's* eEF1a protein will be analysed. Of the two elongation factors to select from the dual encoding species the eEF1a gene was chosen over the eEFL gene due to the proposal by the Carr et al. (2017) study. In which they show that the eEF1a gene has undergone loss of function within the choanoflagellates and now appears that the eEF1a protein can no longer perform its role in translation. Their study suggested the role is now carried out by the eEFL protein within dual encoding species. By cloning the gene into *Escherichia coli* cells, we aim to express the eEF1a from a dual encoding species. This will then allow subsequent analysis of structural and functional properties of the eEF1a protein. This in turn will then allow further analysis to identify what processes the protein is still capable of performing within the dual encoding cells. This may help show why the dual encoding choanoflagellate species possess a protein that has been shown to have loss of function, for its known role within the cells and may highlight any moonlighting properties the protein is capable of performing.

2.0 Methodology

2.1 Bioinformatics

The initial transcriptomes of 19 species of choanoflagellates were provided by Dr Martin Carr but are publicly available from the data published by Richter et al. (2018). For the two species of choanoflagellates *S. rosetta* and *M. brevicollis* their genomic data is freely available from the NCBI data bank. For identification of the genes of interest, a sequence was identified through the website JGI, using the *M. brevicollis* page (https://genome.jgi.doe.gov) to identify any proteins of interest in this study. If corresponding genes were found for *M. brevicollis*, relative to the search, the sequences were used for blast searches for *S. rosetta* using NCBI and for all other species a local blast search, using databases produced using the NCBI+ v2.2.3 (Camacho et al., 2009). If genes were not available through this resource, a search using the NCBI databank for the name of a gene was used. Then a eukaryotic gene was selected from the NCBI database. This was then used to identify a corresponding choanoflagellate gene in a local BLAST search and this gene was then used for the BLAST of the remaining choanoflagellates. Once a correct choanoflagellate sequence was identified through these means, this sequence was used to search the remaining choanoflagellate species, transcriptomic data. Once individual genes were searched for through BLAST and the contig with the lowest E value were chosen to continue in analysis. These were then identified in the transcriptomic data and then reciprocal blasts using the NCBI database were carried out, to ensure the correct genes were identified.

The whole sequences from all available species were aligned using MAFFT 7.402 (Katoh & Toh, 2010). The MAFFT tool was used on the CIPRES science gateway v3 (Miller et al., 2010) with all default parameters set for analysis. From the output provided an initial tree was produced using FastTree 2.1.10 (with default parameters set). This allowed a fast maximum likelihood tree (in 3-4 minutes) to see a phylogenetic tree, this then allowed any changes to the number of sequences etc.. Once confident of an acceptable data set, a Bayesian tree was produced. This was created through the MrBayes 3.2.6 (Ronquist et al., 2012) tool again on the CIPRES science gateway. The parameters were changed by the following; the run time was changed to 96 hours. The rate matrix for amino acids was set to mixed. The number of generations (Ngen=) was set to 5,000,000. The stop early if the convergence diagnostic falls below the stop value option was changed to no. Both the Sumt Burnin

and Sump Burnin values were set at 1,250. The final change was to set the type of consensus tree to all compatible groups. The data produced from Mr Bayes, was then opened using FigTree (Rambout, 2009), to produce a phylogenetic tree. The output data was also used to identify the correct model for the ML analysis.

The next stage was to produce a maximum likelihood model using RAxML 8.2.10 (Stamatakis, 2014) on the CIPRES site. Each individual phylogeny produced by the MrBayes program had an individual model for the maximum likelihood phylogeny. This was changed in the RAxML parameters under the protein substitution model selection. The other changes to the parameters were, set the run time to 96 hours and the boot strap iterations were set at 1,000. Again, the output file was used to create a second phylogenetic tree using FigTree, to compare tree topology.

2.2 Molecular approach to the EF1a gene Bioinformatical analysis for identification of EF1a gene to continue for analysis

Due to the available species within the Carr lab, the species *S. punica* was chosen for analysis, due to the fact it is a dual encoding choanoflagellate species. The EF1a gene sequence identified by Carr et al. (2017) was provided by Dr Carr and used for bioinformatical analysis. The sequence was identified, and extraction of the total RNA from *S. punica* cells was attempted, for the subsequent production of total RNA and cDNA libraries. Unfortunately, due to poor cell growth, even after following several different protocols, the needed volume of cell density required for total RNA extraction could not be reached. In consequence, the EF1a genes from *S. punica* was synthetically produced (Geneart) for subsequent cloning experiments using the pET20b(+) plasmid vector.

General microbiology

For maintenance and propagation of plasmid vectors, DH5α and DH5α-T1 cells were used as the endA1 mutation prevents future degradation of the inserted plasmid, by inactivating intracellular endonucleases within the strain (Bryant, 1987). The cell types used and their corresponding genotypes can be seen in Table 2.1.

Table 2.1: The DH5α strains used in the study and their respective genotype.

Strain	Genotype
DH5α	F– endA1 glnV44 thi1 recA1 relA1 gyrA96 deoR nupG purB20 ϕ 80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17(rK –mK +), λ–
DH5α-T1	F- ϕ80lacZΔM15 Δ(lacZYA-argF)U169 recA1 endA1 hsdR17(rk - , mk +) phoA supE44 thi-1 gyrA96 relA1 tonA (confers resistance to phage T1)

Culture media

Media and maintenance stocks of Luria-Bertani (LB), were made using the following recipe; 10g Tryptone, 5g yeast extract, 10g NaCl and then the pH was adjusted to 7.5 pH by the addition of 1M NaOH. This was then made up to 1 litre with MilliQ H_2O. The media was then sterilised by autoclaving at 15 psi for 20 minutes at 121°C. Once the LB had reached a cooler temperature, the antibiotic was

aseptically added to the mixture at concentrations of 100µg/ml. For the agar plates, the same recipe was followed, with the addition of 1.5% w/v of agar prior to autoclaving. The antibiotic was added to the liquidised LB-agar, once the liquid medium was sufficiently cold to handle.

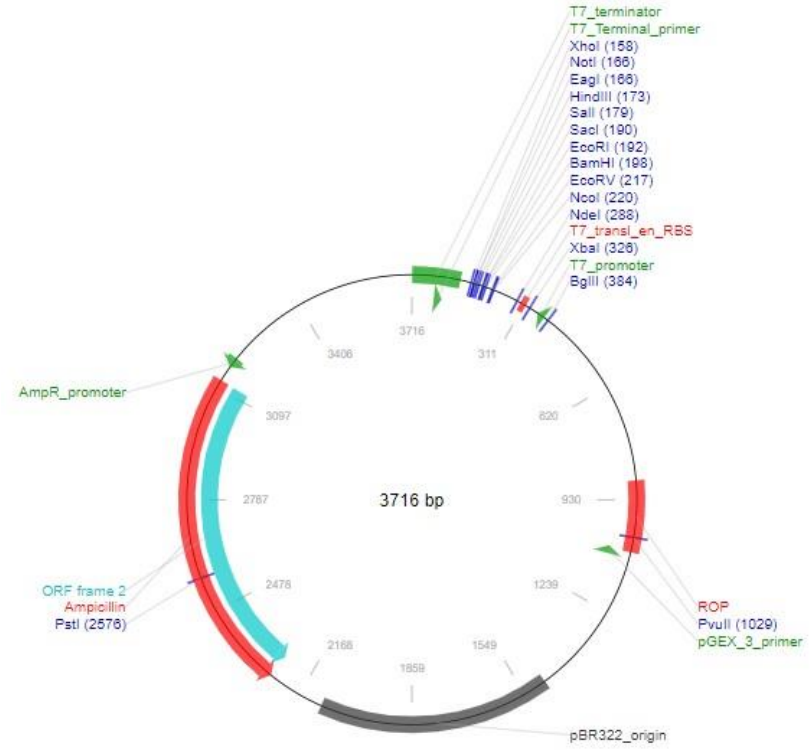


Figure 2.1: Schematic map of the pET20b(+) plasmid vector. Showing multiple cloning sites and ampicillin resistance. The map was obtained from addgene (www.addgene.org).

Restriction endonuclease digests

Plasmid preparations for the EF1a harbouring plasmid and the pET20b(+) plasmid were carried out on 5mL overnight cultures (LB-ampicilin) using the Qiagen®, QiAprep spin mini-prep kit®. Restriction endonuclease digests on both plasmids were carried out according to the NEB double digest finder using NEBuffer 2.1, in a total of 50µL, for a minimum of 4 hours at 37ºC.

Agarose gel Electrophoresis

The products of the digests were analysed on a 1% (w/v) agarose gel in 0.5X TBE (40mM Tris-Cl (pH 8.3), 45mM Boric acid, 1mM EDTA), supplemented with 1/1000 volume of Invitrogen® Sybr safe gel dye. Electrophoresis was conducted at 200mV for an hour, before examining the gel under UV. Bands corresponding to the EF1a insert and the digested pET20b(+) vector were excised from the gel and purified using the Qiagen® Gel extraction kit, according to the manufacturer's instructions.

Cloning

The concentrations of the *HinIII/NdeI*-digested insert and of the *HindIII/NdeI*-digested pET20b(+) plasmid vector were determined using 1% (w/v) agarose gel electrophoresis(section no.), by comparing fluorescent intensity of the Quick-load Purple 1kb DNA ladder (NEB). Ligation reactions,

using T4 DNA ligase (NEB), with different insert:vector molar ratios (Formuala; { [ng of vector] x [kb size of insert] } / {kb size of vector} x (insert:vector ratio) = ng of insert required) were attempted, at 16^{o}C overnight and at 25^{o}C for an hour, using the recommended protocol by NEB. Chemically competent cells, were transformed with 4-5µl of the ligation mixture and successful transformants were selected in LB-ampicillin plates (section number).

Bacterial transformation

The DH5α-T1® cells were subjected to a heat shock transformation protocol. 50µl of cells were maintained for 30 minutes on ice with an aliquot of plasmid vector and insert. The cells were then heat shocked at 42^{o}C for 30 seconds then returned on ice for a further 2 minutes. 800µl of LB-media was then added to the cells and then subsequently grown for 1 hour at 37^{o}C in an shaker incubator. The cells were then plated on to LB-ampicillin plates and grown at 37^{o}C overnight. Successful colonies were then picked for further analysis.

3.0 Results

3.1 eEF1a and EFL within the choanoflagellates

Carr et al. (2017) identified the 21 species of choanoflagellates used in this study and their subsequent varied distribution of EF1a and EFL. With 6 species showing dual encoding of both genes, 4 expressing EF1a and 11 species expressing the EFL genes. A simplified phylogenetic tree of the 21 species used can be seen in Figure 3.1, also introducing the choanoflagellate groups of the Craspedida and Acanthoecida.

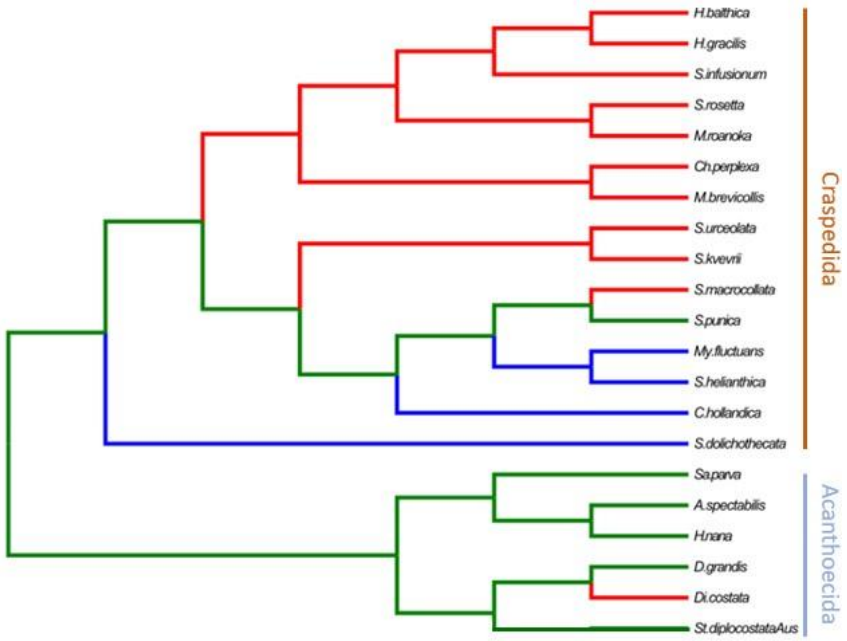


Figure 3.1: A simplified phylogeny of the species included in this study, produced from the larger tree produced by Carr et al. (2017). In which they used a six gene phylogeny to position the choanoflagellates, red branches show the EFL species, blue branches show the EF1a species and the green branches show the dual species. The figure also introduces the species belonging to both the Craspedida and Acanthoecida groups.

3.2 The eEF1b complex and components distribution within the choanoflagellates

All species were subjected to analysis through a BLAST search, to identify any other proteins involved in the formation of the translational EF1 complex. The proteins involved are EF1bα, EF1bγ, EF1bδ, EF2 and EF3, as stated in the Introduction, all the prior proteins are involved in certain aspects of the formation of the EF1a and EFL specific, EF1 complexes, prior to translation. This was to identify any specific relationships between all the known factors involved, such as gene loss/gain, from all the species from all three, elongation factor sub-groups. Table 3.1 shows the resulting data and the relationships involved.

The eEF1b complex, which is formed by the subunits eEF1bα, eEF1bγ and eEF1bδ (Sasikumar et al., 2012), shows a distinct distribution between the eEF1a, eEFL and dual encoding species. All eEF1a species harbour the eEF1bα and eEF1bγ genes, needed to recharge the GDP-GTP state of the eEF1a enzyme. Surprisingly, the three eEF1a encoding species, *C. hollandica, M. fluctuans* and *S. dolichothecata* harbour the eEF1bδ gene and shows this factor is elongation factor specific. The eEF1a distribution mirrors Metazoa, as no metazoan species has been shown to encode eEFL (Carr et al., 2017).

Elongation Factor subgroup	Species	eEF1bα	eEF1bγ	eEF1bδ	eEF2	eEF3L
EF1a	*Salpingoeca helianthica*	Yes	Yes	No	Yes	Yes
	Codosiga hollandica	Yes	Yes	Yes	Yes	No
	Mylnosiga fluctuans	Yes	Yes	Yes	Yes	Yes
	Salpingoeca dolichothecata	yes	Yes	Yes	Yes	No
EFL	*Monosiga brevicollis*	No	Yes	No	Yes	No
	Salpingoeca rosetta	No	Yes	No	Yes	No
	Salpingoeca urceolata	No	Yes	No	Yes	No
	Hartaetosiga gracilis	No	No	No	Yes	No
	Microstomoeca roanoka	No	No	No	Yes	No
	Salpingoeca macrocollata	No	Yes	No	Yes	Yes
	Salpingoeca infusionum	No	No	No	Yes	No
	Didymoeca costata	No	No	No	Yes	Yes
	Salpingoeca kvevrii	No	No	No	Yes	Yes
	Hartaetosiga balthica	No	No	No	Yes	No
	Choanoeca perplexa	No	Yes	No	Yes	Yes
Dual	*Salpingoeca punica*	No	Yes	No	Yes	Yes
	Diaphanoeca grandis	No	No	No	Yes	Yes
	Stephanoeca diplocostata	No	No	No	Yes	Yes
	Acanthoeca spectabilis	No	No	No	Yes	Yes
	Helgoeca nana	No	No	No	Yes	Yes
	Savillea parva	No	No	No	Yes	Yes

Table 3.1: The distribution of elongation factors involved in the formation of the eEF1 complex in relation to the elongation factor species within the choanoflagellates.

As proposed by the Atkinson et al. (2014) study, no EFL encoding species have been found to harbour an eEF1bα gene. This is supported by this study, with no eEF1bα gene transcription in either the EFL or dual encoding species. This is also the first suggestion that the dual encoding species utilise EFL during translation, as they do not possess the means to efficiently recharge the GDP-GTP states of their eEF1a protein. The distribution of eEF1bγ within the eEF1a species is as expected, as it provides scaffolding to eEF1bα during the formation of the eEF1 complex. The distribution of eEF1bγ within the EFL and dual encoding species, shows a random distribution as a product of selection and/or genetic drift. Within the EFL encoding species, 5 species, *Monosiga brevicollis, Salpingoeca rosetta, Salpingoeca urceolata, Salpingoeca macrocollata and Cohoanoeca perplexa* were shown to harbour the eEF1bγ gene. From the dual encoding species only, *Salpingoeca punica* was shown to possess the gene. Regarding eEF1bδ, the three eEF1a species *Codosiga hollandica, Mylnosiga fluctuans* and *Salpingoeca dolichothecata* possess the gene and thus shows distribution specific to the eEF1a choanoflagellate species.

3.3 eEF2 distribution within the choanoflagellates

As mentioned in the Introduction, eEF2 has been shown to be fundamental to cellular processes in all known species. As such it would be expected to be found in any species analysed. This has been shown

also in this study, with all species harbouring the eEF2 gene. This further supports that eEF2 is an integral part of translation within cellular function and is conserved throughout the choanoflagellates.

3.4 The identification and distribution of eEF3L within the choanoflagellates

Whilst performing the BLAST search for the eEF3 gene within the choanoflagellates, many important factors were identified. During the eEF3 BLAST search of each independent choanoflagellate transcriptome, 10 species gave a 0.0 value for eEF3. But multiple high scoring -e values were given during the search, which usually shows a closely related amino acid sequence. The other -e values were shown even in the species that already gave a 0.0 value. This raised questions as to what these genes were and thus the sequences were aligned and then incorporated in to a publicly available data set, on the ABCF family of proteins (Murina et al., 2017). This alignment consisted of 580 nonchoanoflagellate species, showing a diverse selection of species, with known ABCF genes. 190 choanoflagellate genes were then incorporated into the alignment, including the 0.0 valued genes. This totalled 770 sequences within the alignment, and the alignment was made for a 6860 amino acid long alignment. The phylogeny produced was used to identify the different ABCF genes within the choanoflagellates from this closely related group of ATPase-hydrolysing enzymes. The resulting tree can be seen in Appendix 3.1, which shows the relationship between eEF3 and the related ABCF proteins. From this large alignment, several smaller trees were produced to identify where phylogenetically, the identified choanoflagellate genes were placed for each gene within the tree and help to identify the transcriptomic data for protein identification.

The first gene analysed from the ATPase classes was the EF3 gene, which as previously stated was originally thought to be a fungal only gene. The related sequences from the choanoflagellates were placed into an alignment of other known EF3 genes from fungal, viral, bacterial and other eukaryotic organisms. To maximise the chances of a correct tree, two outgroups were selected from closely related ABCF gene sequences. For this tree the genes ABCF4 and ABCF5 were chosen due to their close relationship to EF3. For the ABCF4 gene sequences from three non-related species were chosen, the marine diatom species *Thalassiosira pseudonana* (Heterokonta), the species of alga *Aureococcus anophagefferens* (Heterokonta) and finally, the species of marine coccolithophore *Emiliania huxleyi* (Haptophyta). For the ABCF5 gene the choice for differing species is limited, due to the gene being mainly in fungal species. Two fungal species were chosen, these being *Gaeumannomyces graminis* (Ascomycota) and *Auricularia delicata* (Basidiomycote) and one algal species in *Chlorella variabilis* (Chlorophyta). The alignment contained 89 species, including ten choanoflagellates and was a 1948 amino acid long sequence. The resulting phylogenetic tree can be seen in figure 3.3.

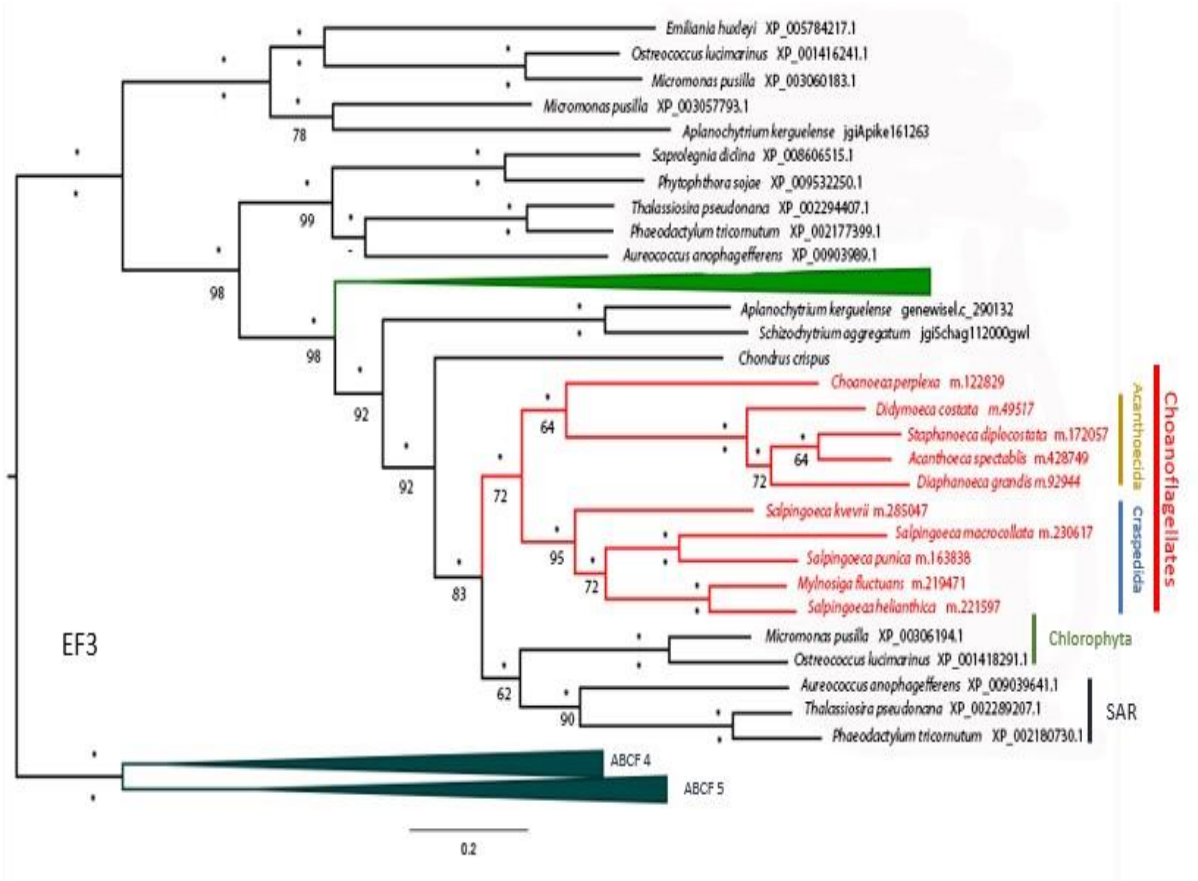


Figure 3.3: The resulting tree from the EF3 analysis. The outgroups ABCF4 and ABCF5 have been used. The tree shows both the maximum likelihood scores (top) and bootstrap values (bottom). Weakly supported branches which had mlBP <50% and biPP <0.70 values were omitted or marked -. The collapsed sections of the tree are; Blue= outgroups and Green= fungal EF3 genes.

The EF3 phylogenetic tree shown in Fig. 3.3 shows that, the choanoflagellate group to be monophyletic with strong support (biPP 1.00; mlBP 83%) and ten species of choanoflagellates present do possess the eEF3L gene. The exact positioning of the choanoflagellate sequences within their clade shows moderate support throughout. With the choanoflagellate EF3 clade being monophyletic and a sister group to the alga species from the super groups of the heterokonts and the Viridiplantae, appears to show evidence of horizontal gene transfer. It suggests that an ancestor of the choanoflagellates acquired the gene and then subsequent gene loss within the choanoflagellates has then occurred. Within the Acanthoecida group of choanoflagellates, all species analysed possess the EF3 gene, so appears that no gene loss has occurred within this group. With other species belonging to this group, yet do not have their transcriptomic data available at present, this may not be the case throughout this group. As proposed by Murina et al. (2017), the eEF3L gene appears to be an ancient protein lineage, yet this study shows it was gained in the choanoflagellates through horizontal gene transfer, with subsequent loss on multiple occasions, throughout the phylogenetic tree.

This phylogeny shown in Figure 3.3 shows similarity to the phylogeny produced by Carr et al. (2017) and positions each choanoflagellate species in similar positions, within their revised taxonomy of the group. Here we find *Didymoeca costata, Stephanoeca diplocostata, Acanthoeca spectabilis* and *Diaphanoeca grandis* all positioned within the Acanthoecida group of choanoflagellates. This phylogeny also shows the species *Salpingoeca kvevrii, Salpingoeca macrocollata, Salpingoeca punica, Mylnosiga fluctuans* and *Salpingoeca helianthica* all positioned within the Craspedida choanoflagellate group. The one exception is in *C. perplexa,* this species is part of the Craspedida group, but appears as

a sister group to the Acanthoecida with moderate to strong support (biPP 1.0 and mlBP 64%). The strong support values show that, eEF3L genes have been identified within the choanoflagellates and the resulting data has been placed in table 3.2. The identification numbers from the Richter (2018) transcriptomes have also been shown for verification.

Species	ID no. for transcriptomic data
C. perplexa	m.122829
D. costata	m.49517
S. diplocostata	m.172057
A. spectablis	m.428749
A. grandis	m.92944
S. kvevrii	m.285047
S. macrocollata	m.230617
S. punica	m.163838
M. fluctuans	m.219471
S. helianthica	m.221597

Table 3.2: The identified eEF3L genes from the choanoflagellate species analysed. The m. numbers represent the ID numbers from the Richter et al (2018) data set.

3.5 The distribution of the ABCF1 gene within the choanoflagellates

As the large ABCF multifamily analysis showed the ABCF genes to be present, the following phylogenetic trees were produced for all clear ABCF genes present throughout the choanoflagellates. Out groups were selected from closely related ABCF genes as above and the following trees were produced. First was the ABCF1 genes, which is shown in Figure 3.4. This phylogeny was created with 74 species, including 18 species of choanoflagellates and the alignment was 1916 amino acids long. The outgroups for this class of ABCF were chosen as ABCF2 and ABCF6. The species used for the ABCF2 were the plant species *Zea mays* (Archaeplastida), the single celled eukaryote *Capsaspora owczarzaki* (Opisthokhonta) and the fungal species *Agaricus bisporus* (Opisthokhonta). For the ABCF6 outgroup, the species chosen were the protist species *Dictyosetelium purpureum* (Amoebozoa), the red algae species *Cyanidioschyzon merole* (Archaeplastida) and the alveolate species *Perkinsus marinus* (Alveolata).

The initial BLAST analysis found almost all of the choanoflagellate species harbour the ABCF1 gene. The exceptions to this are *S. rosetta* and *M. brevicollis,* which appears unusual, as these two species have full genomic data available. If clade 1 from Fig 3.4 is seen, the species branches do however show areas of low support. As such, the exact phylogenetic positioning cannot be fully resolved, within the taxonomic group, through this evidence. Looking at clade 1 and comparing it to the known phylogeny by Carr et al. (2017), the most recent phylogeny for these species, this group of choanoflagellates is positioned as the sister group to the Metazoan clade with strong support (biPP 1.0 and mlBP 97%). Although the group branches are not showing strong support within this phylogeny, the positioning of the choanoflagellates in this group closely matches that of the most relevant phylogeny. With two exceptions, in the species *Salpingoeca urceolata* and *Salpingoeca dolichothecata*, which do not have the data within their transcriptomes. This shows that ABCF1 was present in the LCA of the metazoan/choanoflagellate lineages, and that gene loss has been shown to have occurred twice, on

two different branches within this taxonomic group. It is also apparent that *S. diplocostata* possesses two similar copies of the ABCF1 gene, both positioned within the choanoflagellate clade.

On clade 2, the species *Helgoeca nana* and *A. spectabilis* possess two distinctly different copies of the ABCF1 gene. One copy appears to be closely related to the SAR supergroup homolog of ABCF1. These two species of choanoflagellate are also extremely closely related within the Acanthoecida clade (Carr et al., 2017). The relationship shown in Fig. 3.4 mirrors the known phylogeny and as such would be indicative of gene gain through horizontal gene transfer.

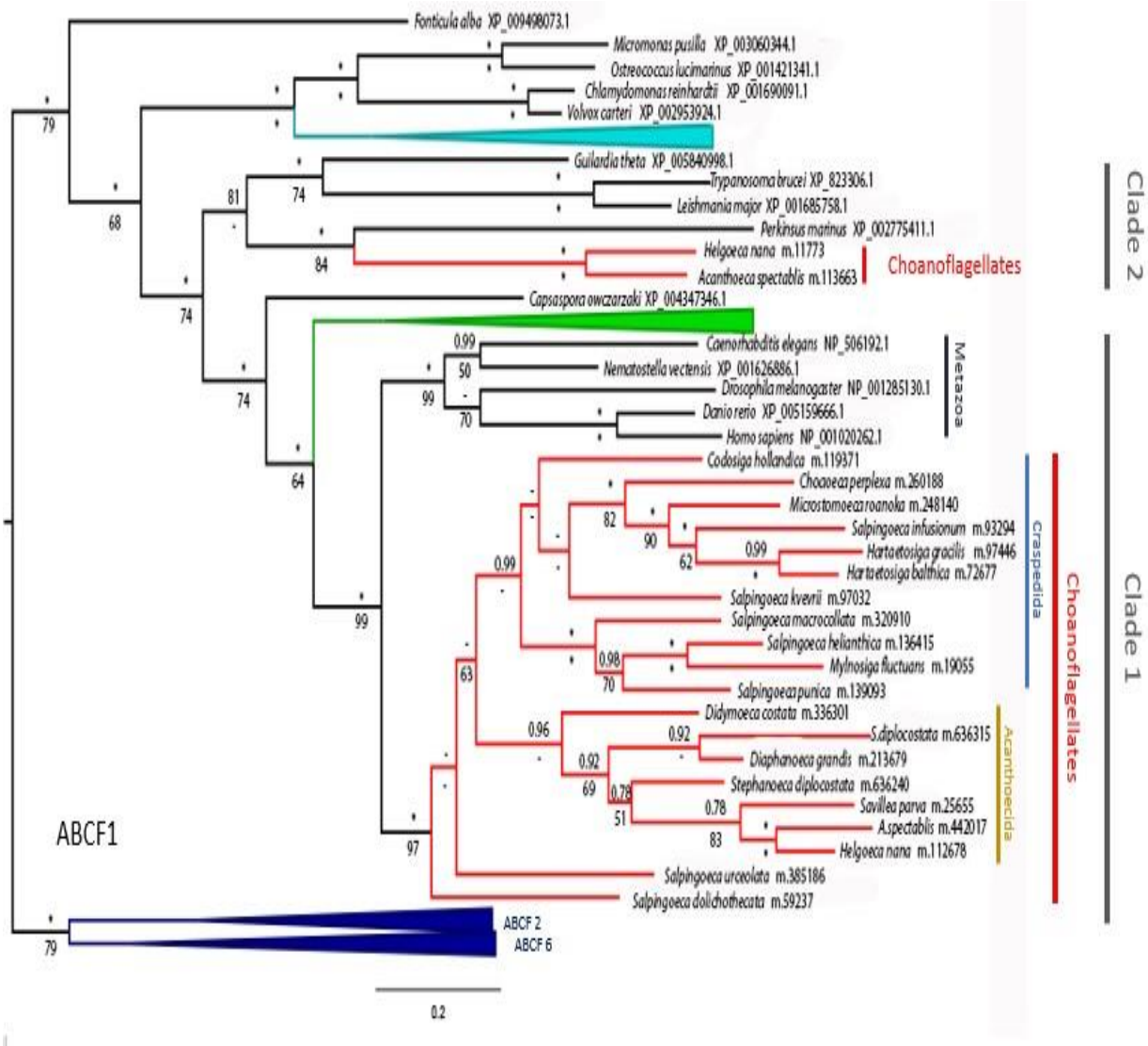


Figure 3.4: ABCF1 tree, showing probability and maximum likelihood values. All maximum support values have been marked with an * (biPP (1.0) and mlBP (100)) and those that are not well supported have been omitted or marked – (biPP <70 and mlBP ,50).

With strong support throughout the tree, and the choanoflagellate group (clade 1) being monophyletic and a monophyletic group within clade 2, the ABCF1 genes can be identified. In Table 3.3 the species and their identification number from the Richter (2018) data set, have been shown.

Species	Gene ID no. for ABCF1 gene.
Salpingoeca helianthica	m.206417
Codosiga hollandica	m.119371
Mylnosiga fluctuans	m.19055
Salpingoeca dolichothecata	m.59237
Salpingoeca urceolata	m.385186
Hartaetosiga gracilis	m.97446
Microstomoeca roanoka	m.248140
Salpingoeca macrocollata	m.320910
Salpingoeca infusionum	m.93294
Didymoeca costata	m.336301
Salpingoeca kvevrii	m.97032
Hartaetosiga balthica	m.72677
Choanoeca perplexa	m.260188
Salpingoeca punica	m.139093
Diaphanoeca grandis	m.213679
Stephanoeca diplocostata	(8) m.636315 (7) m.636240
Acanthoeca spectabilis	(4) m.113663 (8) m.442017
Helgoeca nana	(5) m.112678 (16) m.11773
Savillea parva	m.25655

Table 3.3: the corresponding gene numbers to the found ABCF1 genes, to identify the ABCF1 genes in the Richter AA transcriptome sequences. Brackets () show species with two genes.

3.6 the distribution of the ABCF2 gene within the choanoflagellates

For the ABCF2 analysis, all genes were placed into an alignment consisting of 129 species and an alignment of 2310 amino acids. The outgroups for this data set were the ABCF1 gene, using the following species; *Caenorhabditis elegans* (nematode), *Thalassiosira pseudonana* (Heterokonta) and *Aureococcus anophagefferens* (Heterokonta). The other outgroup is the ABCF6 gene, using the following species; *Dictyostelium purpureum* (Amoebozoan), *Perkinsus marinus* (alveolate) and *Cyanidiaschyzan merolae* (Rhodophyta). These groups were chosen for their close relationship to ABCF2 and from varied species. Fig 3.5 shows the resulting tree.

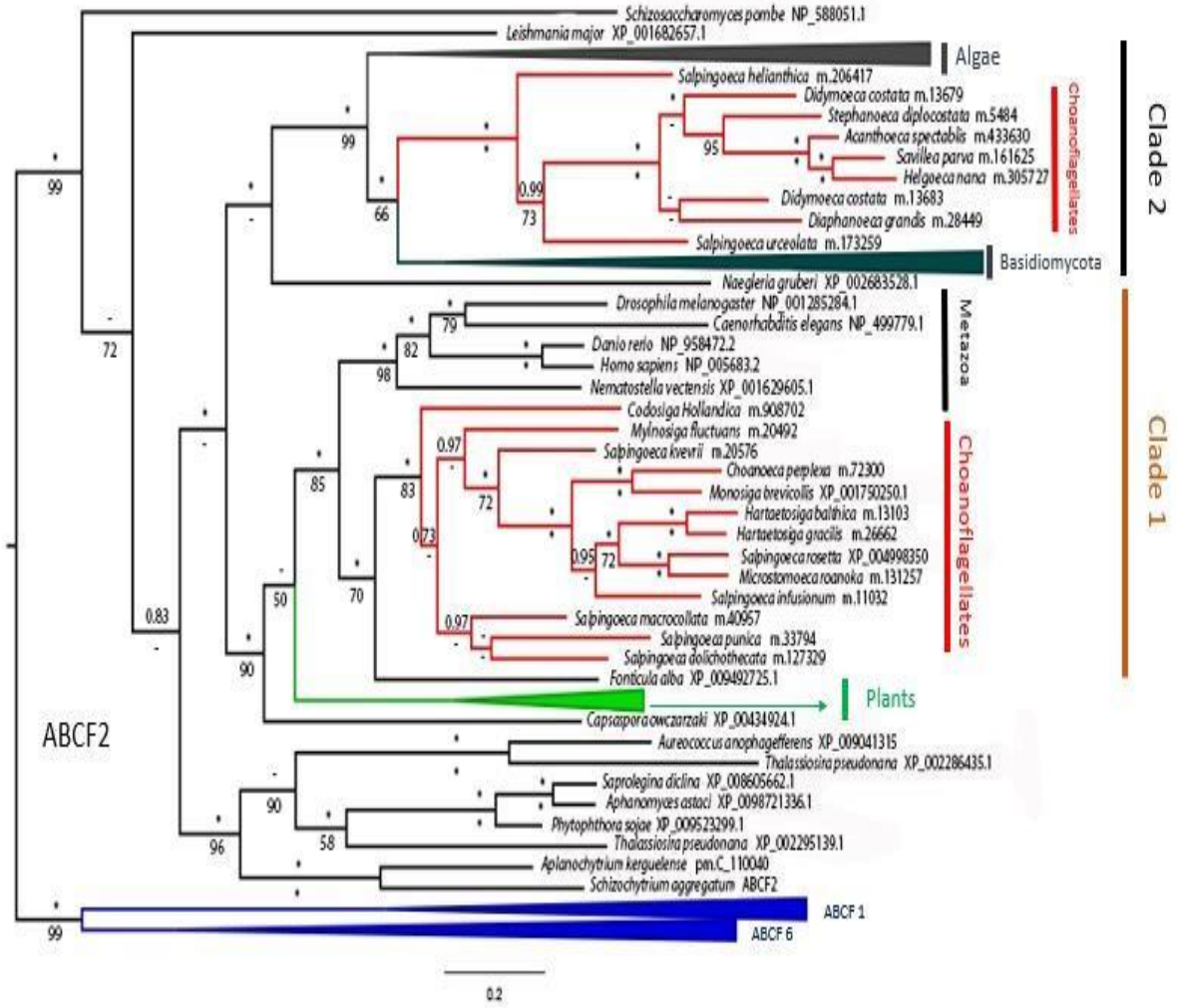


Figure 3.5: ABCF2 tree, showing probability and maximum likelihood values. All maximum support values have been marked with an * (biPP (1.0) and mlBP (100)) and those that are not well supported have been omitted or marked – (biPP <70 and mlBP ,50).

As can be seen from Fig.3.5 all species of choanoflagellates were shown to express the ABCF2 gene, except *D. grandis.* Also, two distinct groups can be identified within the choanoflagellates. Again, as with the ABCF1 gene, one group is strongly supported as sister groups to the metazoan lineage (biPP 1.0 and mlPP 83%) and the other moderate to strongly supported (biPP 1.0 and mlBP 66%) as a sister group to the Basidiomycota ABCF2 lineage. It is shown that clade one contains only species allocated to the Craspedida group of choanoflagellates. This part of the phylogeny also shows close inter species relationships, as in the phylogenetic relationships in the Craspedida group shown by Carr et al. (2017). Although branches of this clade are not well supported, the species positioning within the clades, in comparison to the recent phylogenetic positioning of the choanoflagellates, adds credence to support the positioning of the species is correct.

In clade 2, the group of choanoflagellates belong to the Acanthoecida, with well supported data for the positioning of the individual species within the Acanthoecida group. Also, this clade shows moderate support for the data, in two exceptions from the Craspedida group; *S. urceolata* and *S. helianthica* being present in this part of the tree. This clade is also showing moderate to strong support in its relationship with the Basidiomycota lineage (biPP 1.0 and mlBP66%), for the ABCF2 homologs.

With evidence suggesting the gene has undergone two vertical inheritance events from an ancestor of the choanoflagellates, with substantial gene loss of the Basidiomycota type ABCF2, within the Craspedida group. The identified ABCF2 genes have been placed in table 3.4. These show the corresponding gene ID numbers from the Richter et al. (2018) transcriptomes.

Species	Gene ID no. for ABCF2 gene.
Salpingoeca helianthica	m.226195
Codosiga hollandica	m.908702
Mylnosiga fluctuans	m.20492
Salpingoeca dolichothecata	m.127329
Monosiga brevicollis	XP_001750250
Salpingoeca urceolata	m.173259
Hartaetosiga gracilis	m.26662
Microstomoeca roanoka	m.131257
Salpingoeca macrocollata	m.40957
Salpingoeca infusionum	m.11032
Didymoeca costata	m.13679
Salpingoeca kvevrii	m.20576
Hartaetosiga balthica	m.13103
Choanoeca perplexa	m.72300
Salpingoeca punica	m.33794
Salpingoeca rosetta	XP_004998350.1
Stephanoeca diplocostata	m.5484
Acanthoeca spectabilis	m.433630
Helgoeca nana	m.412183
Savillea parva	m.161625

Table 3.4: The corresponding identification numbers for the ABCF2 genes identified in the analysis. The m. numbers are for the Richter (2018) transcriptomes. XP_ numbers are from NCBI.

3.7 The distribution of the ABCF4 gene within the choanoflagellates

The choanoflagellates as a group do transcribe the ABCF4 gene, but in this data set only four species harbour the gene and the analysis is shown in Fig. 3.6. Again, closely related outgroups were selected as ABCF5, with the following species selected; *Auricularia delicata* (Basidiomycota), *Pestalotiopis fici* (Ascomycota) and *Gibberella zeae* (Ascomycota). For the EF3 outgroup, the species chosen were; *Chondrus crispus* (Rhodophyta), *Emiliania huxleyi* (Haptophyta) and *Setosphaeria turcica* (Ascomycota). The resulting alignment consisted of 33 species, including 4 choanoflagellates, in a 1805 long amino acid sequence. The resulting phylogeny can be seen in Fig. 3.6.

As can be seen in Fig. 3.6, this phylogeny has strong support running throughout the tree, with only weak support at the second and third branches of the tree. This suggests that the relationship between the ABCF4 groupings, inter specially, as being poorly resolved. Clade 1 (Figure 3.6) shows strong support values within the branches, to show that four species of choanoflagellates do possess the ABCF4 gene. All species within this clade belong to the Acanthoecida choanoflagellate group. Their identical positioning within the known choanoflagellate phylogeny, coupled with the strong support values within the species group in clade 1, suggests that the positioning is correct. Their positioning as

a sister group to the Haptophyta and Heterokonta lineages on clade 1, shows moderate support (biPP 0.96 and mlBP 51%), as to the positioning of the groups, but strong support in the branches following. This indicates that the correct positioning is shown, but has not been fully resolved, within this part of the phylogeny.

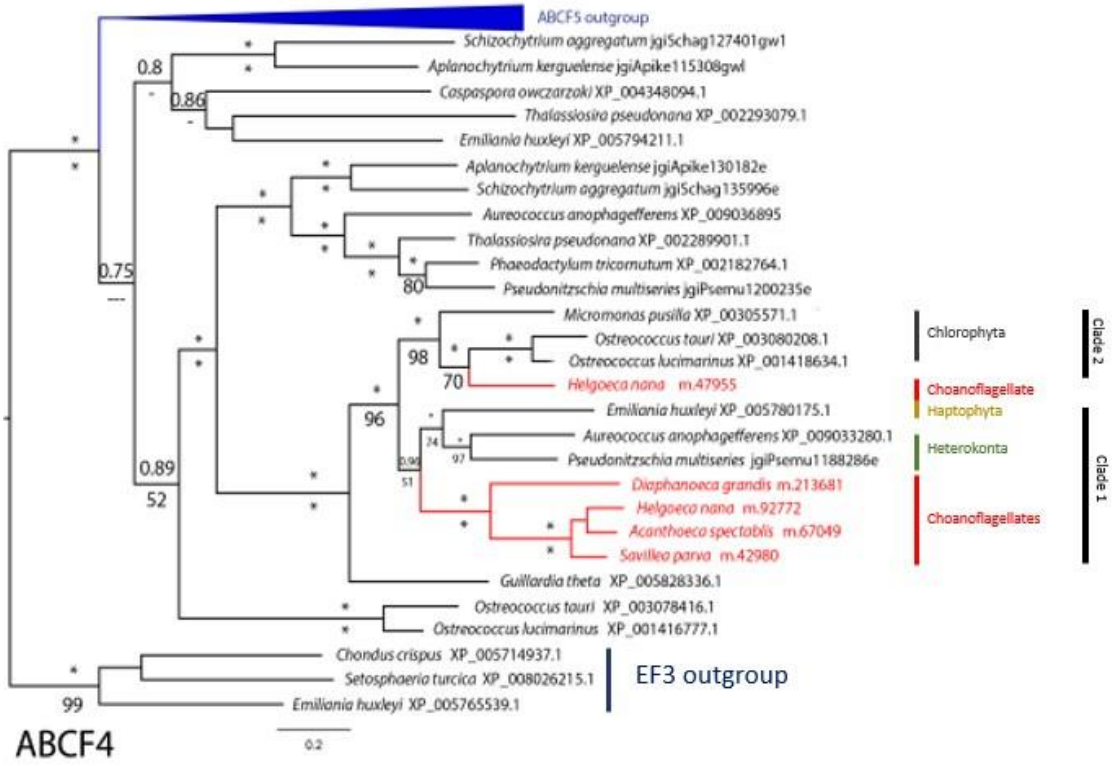


Figure 3.6: ABCF tree, showing probability and maximum likelihood values. All maximum support values have been marked with an * (biPP (1.0) and mlBP (100)) and those that are not well supported have been omitted or marked – (biPP <70 and mlBP ,50).

H. nana shows strong support values towards the species harbouring two copies of the ABCF4 gene. One from the Heterokonta and Haptophyta origin (Clade 1) and a second from Chlorophyta origin (Clade 2). As can be seen on clade 2, in Fig. 3.6, the support values would seem to indicate this evidence clearly (biPP 1.0 and mlBP 70%) and coupled with two genes being found during the BLAST search, can be shown to have independent two copies with confidence. This tree shows support to the ABCF4 gene being gained through horizontal gene transfer on two separate occasions within the choanoflagellates. The identified genes and corresponding ID numbers from the Richter (2018) transcriptomes can be found in table 3.5.

Species	ID number for ABCF4 gene
Helgoeca nana	m.92772 & m.47955
Diaphanoeca grandis	m.213681
Savillea parva	m.42980
Acanthoeca spectablis	m.213681

Table 3.5: The identified ABCF4 genes in the four species of choanoflagellates which harbour the gene. The corresponding ID numbers are to identify the sequences in the Richter (2013) transcriptomic data set.

3.8 The distribution of the ABCF6 gene within the choanoflagellates

The final gene to be analysed was the ABCF6 gene, the two outgroups were selected from closely related genes. These were ABCF1 and ABCF2, the species chosen were for the ABCF1 genes; *C. elegans* (Nematoda), *T. pseudonana* (Heterokonta) and *A. anophagefferens* (Heterokonta) and for the ABCF2 genes the species were; *Saccharomyces cerevisiae* (Ascomycota), *Z.mays* (Angiosperms) and

C.owczarzaki (Opisthokonta). The alignment was made using 21 species, including 5 choanoflagellates, and the sequence was 973 amino acids long. The resulting phylogeny can be seen in figure 3.7.

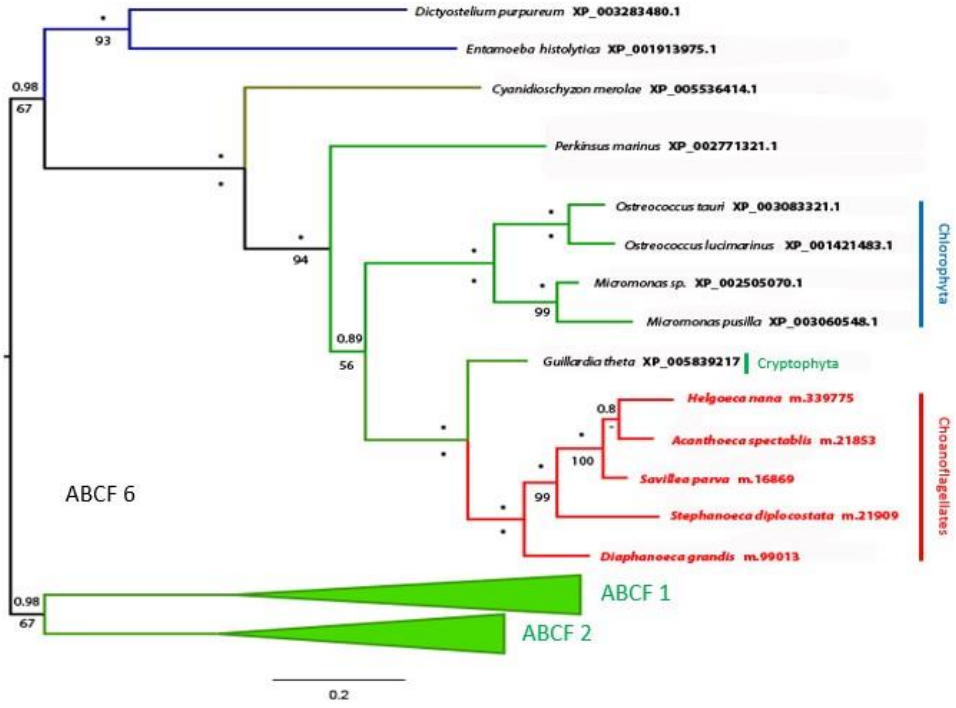


Figure 3.7: ABCF6 tree, showing probability and maximum likelihood values. All maximum support values have been marked with an * (biPP (1.0) and mlBP (100)) and those that are not well supported have been omitted or marked – (biPP <70 and mlBP ,50).

The phylogeny shown in Figure 3.7 has strong support throughout the tree, and as the outgroups are on independent branches and show strong support values, it can be confidently stated that the five species of choanoflagellates shown do harbour the ABCF6 gene. As can be seen, the choanoflagellates have been placed in a moderately supported sister group to the Cryptophyta clade (biPP 0.89 and mlBP 56%). Within the choanoflagellate group, very strong support values are shown within the branches and their phylogenetic positioning within this clade mirrors the known phylogeny of the choanoflagellates. The evidence supports horizontal gene transfer within the choanoflagellates.

As with the ABCF4 phylogeny, it is the Acanthoecida species only, that harbour the ABCF6 gene and shows that their LCA shared a lineage with the algal species present in this phylogeny. The resulting identification of ABCF6 genes and relative ID numbers have been placed in table 3.6. Table 3.7 shows the distribution of the ABCF genes throughout the 21 species of choanoflagellates.

Species	Identification number for ABCF6 genes
Helgoeca nana	m.339775
Acanthoeca spectabilis	m.21853
Savillea parva	m.16869
Stephanoeca diplocostata	m.21909
Diaphanoeca grandis	m.99013

Table 3.6: The identified ABCF6 gene identification numbers for the Richter (2013) transcriptomes and related choanoflagellates.

Species	ABCF1	ABCF2	ABCF4	ABCF6
Salpingoeca helianthica	Yes	Yes	No	No
Codosiga hollandica	Yes	Yes	No	No
Mylnosiga fluctuans	Yes	Yes	No	No
Salpingoeca dolichothecata	Yes	Yes	No	No
Monosiga brevicollis	Yes	Yes	No	No
Salpingoeca rosetta	No	Yes	Yes	No
Salpingoeca urceolata	Yes	Yes	No	No
Hartaetosiga gracilis	Yes	Yes	No	No
Microstomoeca roanoka	Yes	Yes	No	No
Salpingoeca macrocollata	Yes	Yes	No	No
Salpingoeca infusionum	Yes	Yes	No	No
Didymoeca costata	Yes	Yes	No	No
Salpingoeca kvevrii	Yes	Yes	No	No
Hartaetosiga balthica	Yes	Yes	No	No
Choanoeca perplexa	Yes	Yes	No	No
Salpingoeca punica	Yes	Yes	No	No
Diaphanoeca grandis	Yes	No	Yes	Yes
Stephanoeca diplocostata	Yes	Yes	No	Yes
Acanthoeca spectabilis	Yes	Yes	Yes	Yes
Helgoeca nana	Yes	Yes	Yes	Yes
Savillea parva	Yes	Yes	Yes	Yes

Table 3.7: Showing the distribution of the ABCF proteins within the choanoflagellates.

3.9 The distribution of a selection of Ribosomal Proteins within the choanoflagellates

A selection of the proteins from both the small and large ribosomal sub-units were analysed to indicate gene presence within the choanoflagellates. Firstly, the small (40S) ribosomal sub unit was analysed to see which genes were present in the transcriptomes of the 19 species, except for *S. rosetta* and *M. brevicollis* whose entire genomes have been sequenced. The resulting data can be seen in Table 3.8. for the small (40S) ribosomal sub unit and Table 3.9 for the large (60S) ribosomal sub unit. Table 3.9 shows a completely different story, As is shown, there appears to be multiple gene loss of large ribosomal genes within the choanoflagellates.

EF type	Choanoflagellate species	Ribosomal gene								
		RPS4	**RPS5**	**RPS6**	**RPS7**	**RPS8**	**RPS12**	**RPS15**	**RPS19**	**RPS25**
EF1a	*Salpingoeca helianthica*	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Codosiga hollandica	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Mylnosiga fluctuans	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Salpingoeca dolichothecata	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
EFL	*Monosiga brevicollis*	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Salpingoeca rosetta	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Salpingoeca urceolata	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Hartaetosiga gracilis	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Microstomoeca roanoka	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Salpingoeca macrocollata	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Salpingoeca infusionum	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Didymoeca costata	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Salpingoeca kvevrii	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Hartaetosiga balthica	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Choanoeca perplexa	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
DUAL	*Salpingoeca punica*	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Diaphanoeca grandis	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Stephanoeca diplocostata	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
	Acanthoeca spectabilis	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Helgoeca nana	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Savillea parva	yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 3.8: The corresponding analysis for the 40S or small ribosomal sub units within the choanoflagellates. The resulting table shows gene presence within the species.

EF type	Choanoflagellate species	Ribosomal gene												
		RPL3	**RPL4**	**RPL5**	**RPL6**	**RPL7**	**RPL7a**	**RPL8**	**RPL10**	**RPL18**	**RPL21**	**RPL27a**	**RPL31**	**RPL37**	**RPL44**
EF1a	*Salpingoeca helianthica*	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	yes	Yes	Yes	Yes	No
	Codosiga hollandica	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No
	Mylnosiga fluctuans	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Salpingoeca dolichothecata	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No
EFL	*Monosiga brevicollis*	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Salpingoeca rosetta	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Salpingoeca urceolata	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	No	Yes
	Hartaetosiga gracilis	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No
	Microstomoeca roanoka	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	No
	Salpingoeca macrocollata	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Salpingoeca infusionum	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	No	No
	Didymoeca costata	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Salpingoeca kvevrii	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Hartaetosiga balthica	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No
	Choanoeca perplexa	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
DUAL	*Salpingoeca punica*	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Diaphanoeca grandis	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No
	Stephanoeca diplocostata	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes	No	Yes
	Acanthoeca spectabilis	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	No	Yes	No	No
	Helgoeca nana	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Savillea parva	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes

Table 3.9: The corresponding large ribosomal gene analysis, showing gene loss across all 21 species of choanoflagellates. Shows gene presence within the choanoflagellates.

•

Regarding RPL27a, the transcriptional data shows gene loss to have occurred on four separate occasions and is not elongation factor species specific. The same could be said about RPL37, this gene has been lost a total of five times across the choanoflagellate group. Again, this is not EF species specific, but one branch containing the species *H. balthica, H. gracilis* and *S. infusionum* on subgroup 1 appears to have lost the gene in an ancestor of the three species and the branch containing the species *C. perplexa* and *M. brevicollis,* which both possess the gene. RPL44 also shows the same relationship in the sub-group 1 as RPL37. The rest of the tree shows that the RPL44 gene has been lost a minimum of seven times, including the loss in the subgroup 1. Again, this does not look elongation factor dependant, as it has been lost across all elongation factor species type.

3.10 The Molecular biology approach to S. punica EF1a

With six species to choose from, a bioinformatical process was followed to identify an ideal protein candidate to subject to the gene-to-protein structure pipeline. Three programs were used to analyse each of the six dual encoding species EF1a sequences. First using the SCOOBY globular domain software, then through the ScanProsite and finally through the RONN program. Of the six proteins available, the *S. punica* eEF1a protein was chosen to continue forward. Due to its globular domains being identified, the correct function of the protein predicted, and areas of high order predicted within the protein. The full bioinformatical analysis results can be seen in figure 3.9.

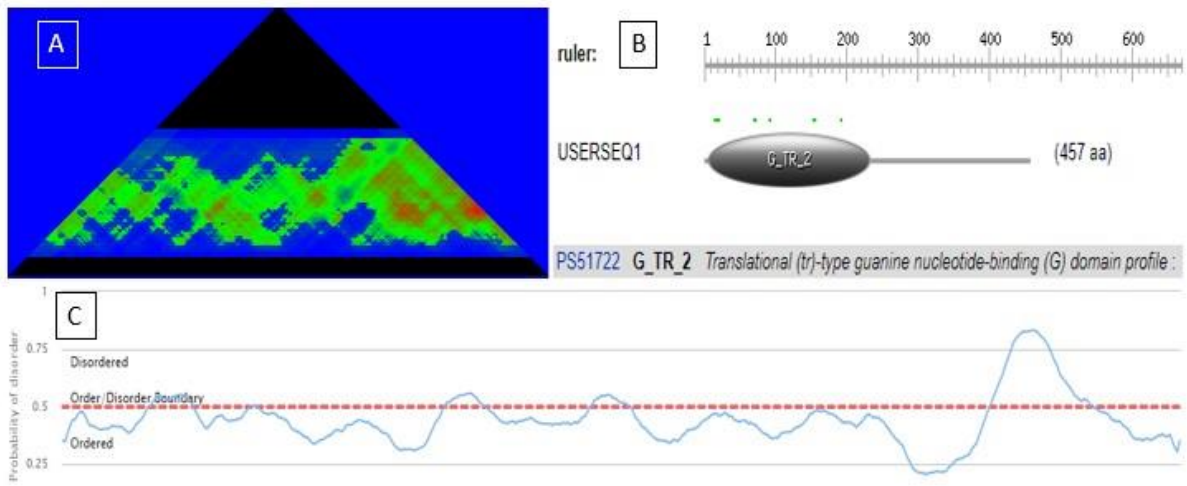


Figure 3.9: Bioinformatic analysis on eEF1a from *S. punica*, (A) shows the SCOOBY analysis, putatively revealing the presence of four globular domains, shown in red. The larger area of red on the right-hand side also coincided with the analysis performed by ScanProsite in panel (B), which identified a larger domain. This amino acid stretch also was identified as a translational guanine nucleotide-binding domain, which EF1a is expected to comprise. (C) shows the predicted ordered/disordered regions within the protein. Large areas of this protein appear ordered and the regions of disorder appear to be the regions between the conserved domains, as predicted by the SCOOBY analysis.

As the initial analysis of the *S. punica* eEF1a protein showed four globular domains, the sequence was subjected to model building using the SWISS-MODEL builder (Waterhouse et al., 2018), to investigate further. As can be seen in figure 3.10, the resulting model also showed four domains and correlated with the known structure of eEF1a. As eEF1a has been subjected to many studies, the protein domains are well understood (Zagari et al., 2001) and the protein is known to have four domains to carry out

•

its cellular function. This allowed the study to progress using the correct protein sequence to investigate the role of the eEF1a protein within the dual encoding species of the choanoflagellates.

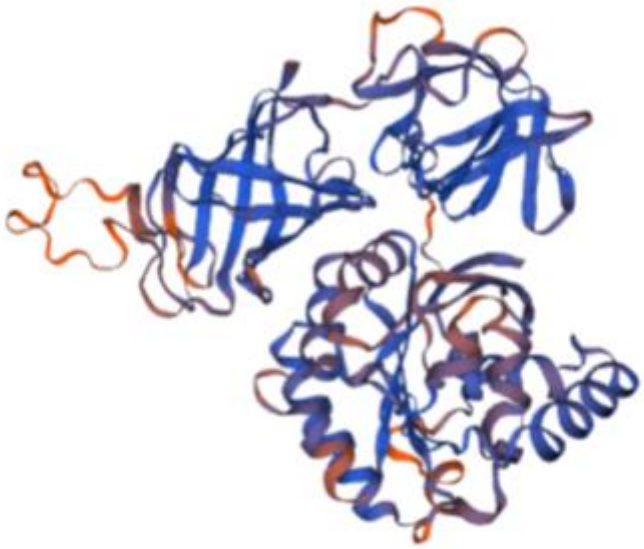
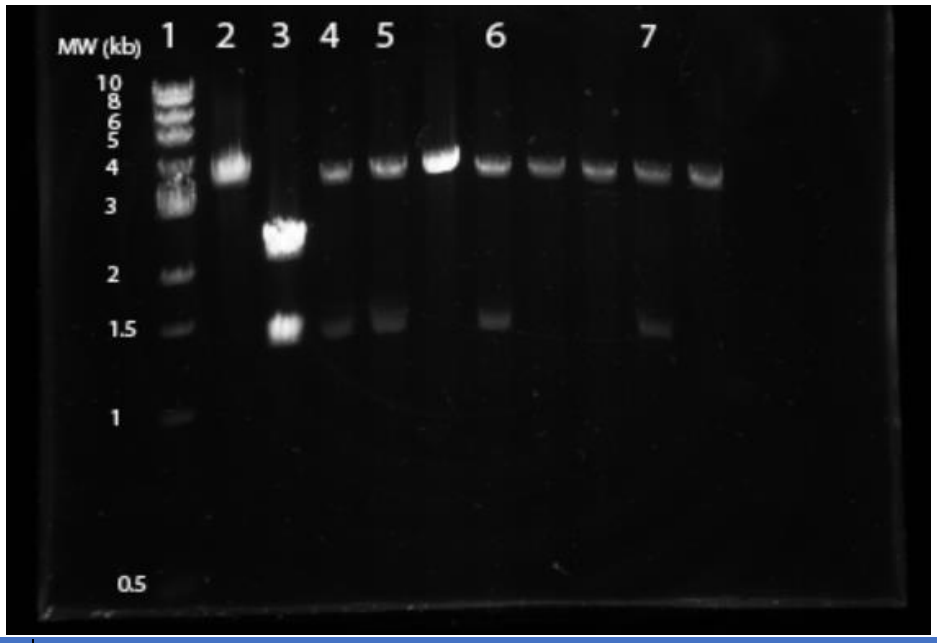


Figure 3.10: The resulting model produced using the SWISS-MODEL software for the *S. punica* eEF1a protein sequence. The domains for the produced model using the software, matches that of the known eEF1a structure.

For improved transformation efficiency One Shot® chemically competent cells (NEB) were used. Using 100ng of vector and for the 1:2, and 1:3 insert:vector molar ratios, ligation reactions at room temperature were successful. To confirm the eEF1a gene had been cloned successfully into the vector, single colonies were picked and used to inoculate 5mL of LB containing ampicillin for the generation of mini preps. Plasmid preps were subjected to restriction endonuclease digests (methodology) and outcomes were analysed by agarose-gel electrophoresis (Figure 3.11).

Lane 2 was for the digested, linear, pet20b(+) vector, the expected DNA fragment size was 3428bp, the estimate for the size of the linear DNA fragments is shown to be 3980bp . In lane 3 was the EF1a gene and the storage vector, the sizes expected for these were; vector 2393 bp and EF1a gene 1394 bp. The gel and calibration graph estimated the storage vector to be 2570bp and EF1a was estimated to be 1660bp. Picked colonies were in lanes 4, 5, 6 and 7. As can be seen in Table 3.7 the estimated sizes of both vector and eEF1a were not accurate. The pet20b (+) vector was consistent throughout, with all values given at 3980bp. This would be a large difference, but if the ladder in lane one is used, by eye, the DNA fragments fall within the expected area. With the pet20b(+) vector having a linear length of 3428bp, it would be expected to be shown between bands 5 and 6 in lane 1. As can be seen, this appears to be true. Also, in the case of the EF1a, the linear DNA fragment has an expected size of 1394bp. From these results the range of EF1a fragments is between 1660 and 1860 bp. Again, by eye you would expect to see these bands, just below band 8 in lane 1. Once again, the bands do appear to be below the expected band.



DNA Fragment	Lane number and estimated MW (kbp)					
	2	3	4	5	6	7
A	3.98	X	3.98	3.98	3.98	3.98
B	X	2.57	X	X	X	X
C	X	1.66	1.78	1.78	1.86	1.66

Figure 3.11: Restriction analysis of putatively successful ligation products by 1% (w/v) agarose gel electrophoresis in 0.5x TBE. Showing successful cloning in 4 colonies. The table shown is using the calibration graph, using the form of y= -2.3689 x + 1.882 was used to estimate the MW of the DNA fragments produced by the transformed *E.coli* cells.

With the ladder being produced to have a known weight, these discrepancies can be analysed further. To calculate the estimated weight of the known ladder, using band 7 as a reference, as it is close to the EF1a expected length. The known length of this band is 1500bp. Using the calibration graph produced in this study, the value given for this band is 1780bp. With this showing difference within DNA fragments of known length, may explain the discrepancies with the other fragments estimated MW and still supports that the cloning was successful.

•

DISCUSSION 4.0

4.1 The distribution of eEF1a, eEFL and dual species and corresponding eEF1b complex proteins

As eEF1a and eEFL studies usually are interested in reconstructing the apparent punctate distribution of the gene throughout the tree of life, this study decided to take a new view point. As a novel study into the area, this study looked at the relationship between species which harbour eEF1a, eEFL and the dual encoding species. With a first insight into which proteins interact with the eEF1 complex dependant on which elongation factor the species harbours. This study was also unique as it looked at the closest living unicellular relatives to the metazoan kingdom, by using choanoflagellates as a model organism.

As the components of the elongation factor 1 complex in eukaryotes usually consists of eEF1a/eEFL, eEF1bα, eEF1bγ, eEF1bδ, and additional proteins eEF2 and in some cases eEF3 (Mateyak et al., 2017; Sasikumar et al., 2012), this was the starting point for the investigation. As stated in the Introduction, eEF1a species require eEF1bα for efficient translation of proteins (Gromadski et al., 2007). As stated by Atkinson et al. (2014), no eEFL species has been found to possess a transcribed copy of the eEF1bα gene. As can be seen by table 3.1 this still holds true through the choanoflagellate species analysed in this study. As expected, only eEF1a choanoflagellate species do possess the gene. As for the dual species, no species analysed possessed a copy of the gene, showing similar patterns of distribution as the eEFL species. This would indicate that the dual encoding species could not use eEF1a for efficient protein translation, as the speed in which GDP is replaced by GTP is far too slow for cellular processes. This must indicate therefore that the dual encoding species most likely utilise EFL as the main elongation factor.

As previously mentioned, eEFL is theorised to be able to spontaneously recharge its GTP without the GEF, eEF1bα or use an unknown GEF (Atkinson et al., 2014). As stated by Carr et al. (2017), through their phylogenetic studies, the evidence supports that the last common ancestor (LCA) of the choanoflagellate species possessed both eEF1a and eEFL genes. But subsequent gene loss has occurred, on multiple occasions on multiple branches of the choanoflagellate phylogeny. The hypothesise by Atkinson et al. (2014) can be seen in the choanoflagellates, where if a species has both eEFL and eEF1a, a loss of function by mutation in eEF1a or eEF1bα would cause the process to be hindered. If an individual with both factors can continue to function, using eEFL it would be inherited throughout the species evolution, with no ill effect upon the individual. This trait would continue to its progeny and continue forward in this manner throughout the evolution of the species. Mirrored to that, if a species has a random mutation which effects eEFL's function, then eEF1a and eEF1bα would carry on the process, again without hindrance to an individual's translation capabilities. This would then carry the same scenario for eEF1a and eEF1bα being maintained for elongational properties of the cell and again continued throughout the evolution of the species.

Presuming null mutations are equally likely in all three genes, for an eEFL species to arise a functionally terminal mutation must occur in either eEF1a or eEF1bα. Doing so would make both unusable for their function during translation. This leads to a 2 in 3 chance of a random mutation happening from all three mentioned elongation factor genes. For an eEF1a species to arise one mutation must occur in eEFL but not the other two proteins, so a 1 in 3 chance of this phenomenon happening. As a mutation is twice as likely to happen to produce an eEFL species, one would expect to find more eEFL species within a taxonomic group. This is true within the choanoflagellates as there are ten species of eEFL encoding species and only four eEF1a species, this shows support to this theory, through probability

alone. It has also been shown by Carr et al. (2017) how in dual encoding species, eEFL has been conserved for functional processes. Where in stark contrast, the eEF1a gene in the dual species have shown loss of translational function, due to accumulated mutations. As none of the dual species harbour eEF1bα, the gene loss of eEF1bα may be the cause of the loss of function of eEF1a and the subsequent use of eEFL during translation. This may indicate the reason for the sporadic distribution of the elongation factors throughout the eukaryotic phylogeny is due to the loss of eEF1bα, leaving eEF1a too slow for cellular function and the subsequent use of eEFL for the process.

The choanoflagellate eEF1a protein being shown to have loss of function in the dual encoding species (Carr et al., 2017). This has also been shown in a different species, in the study by Kamikawa et al. (2008). In which they identified the marine diatom species *Thalassiosira pseudonana* to be a dual encoding species also. Their study also showed that the eEF1a protein in *T. pseudonana* is also losing functional constraint. The other known dual encoding species, *Basidiobolous ranarams* also provides another explanation for elongation factor evolution. The study by Henk & Fisher (2012) showed how *B. ranarams's* gene copy number for eEFL is three times that of its eEF1a gene. Henk & fisher (2012) propose that the copy number of a gene may have an effect on the evolutionary choice between the two genes. As production of more copies of a gene can favour the preservation of gene duplicates, through gene conservation. As a loss of function of eEF1a has been seen in both the dual encoding choanoflagellates as well as *B. ranarams,* and a model of gene loss provided by the Henk & Fisher (2012) study, provides further areas of research within the choanoflagellates. By identifying the gene copy number within the choanoflagellates, it could be seen if they too show similar traits. If so, this would suggest the hypothesis plays a role in the evolution of the dual encoding species and an explanation towards the EFL mutual exclusivity in other species. With the three known dual encoding species showing similar traits and yet from different taxonomic groups, leads to the idea of convergent evolution of dual encoding species.

This leads to the question, why do the dual encoding species harbour an actively transcribed eEF1a protein? As said in the Introduction, the molecular analysis in this study aimed to find an answer to this, as it has been proposed that many proteins possess "moonlighting" properties. As shown by the review on the eEF1 complex by Sasikmuar et al. (2012), the eEF1a protein may not just play a role in translation. Studies have shown links between the role of the cytoskeletal organisation and translation. One such study has shown how, in permeabilised cells, slight changes to the actin cytoskeleton had greatly negative effects on the level of translation within the cells. This was shown to happen even though the cells showed similar levels of translational factors to normal cells (Stapulionis et al., 1997). This coupled with other studies on eEF1a itself may provide answers to why the dual species still transcribe active copies of the gene.

As yet, no studies have been shown to advance the knowledge in this area, probably due to a lack of genomic data for dual species, not just in the choanoflagellates, but in all other known dual encoding species. This, as said, was the initial basis behind the molecular portion of this study. However, due to poor choanoflagellate cell growth at the beginning of this study, coupled with failed attempts at transformation with certain *E.coli* strains, hindered the initial investigation. Due to time constraints, this side of the project was left incomplete. But as shown, the final attempt proved successful in cloning the gene into a bacterial host. Now the eEF1a gene has been cloned, the plasmid containing the gene can now continue through the gene-protein pipeline. This will allow further studies to characterise the protein and investigate the structural and functional properties of the eEF1a protein

within the dual encoding species. Only then could we understand in what function these species still utilise eEF1a.

4.2 The EF1b complex

The eEF1bα gene is an ancestral gene and has prokaryotic origins (Sasikumar et al., 2012). The distribution of this gene shows a distinct pattern within the choanoflagellates, which is also apparent in other eukaryotic species. This gene is only present within the genomic data of eukaryotic eEF1a species, as stated by Atkinson et al. (2014). A loss of function within the EFL or dual encoding choanoflagellates, appears to have led to deletion of the gene from the analysed species. As such, the gene has been lost multiple times, Figure 4.1 shows gene loss of all genes for the eEF1b complex within the analysed species. As all eEF1a species belong to the Craspedida choanoflagellate group, the gene loss for the Craspedids has happened a minimum of 3 times. For the Acanthoecida group, with no eEF1a species known to the group, it appears to have been lost in an ancestor of the group. Until all species within the Acanthoecids have been screened for their use of elongation factors, this will not be fully resolved.

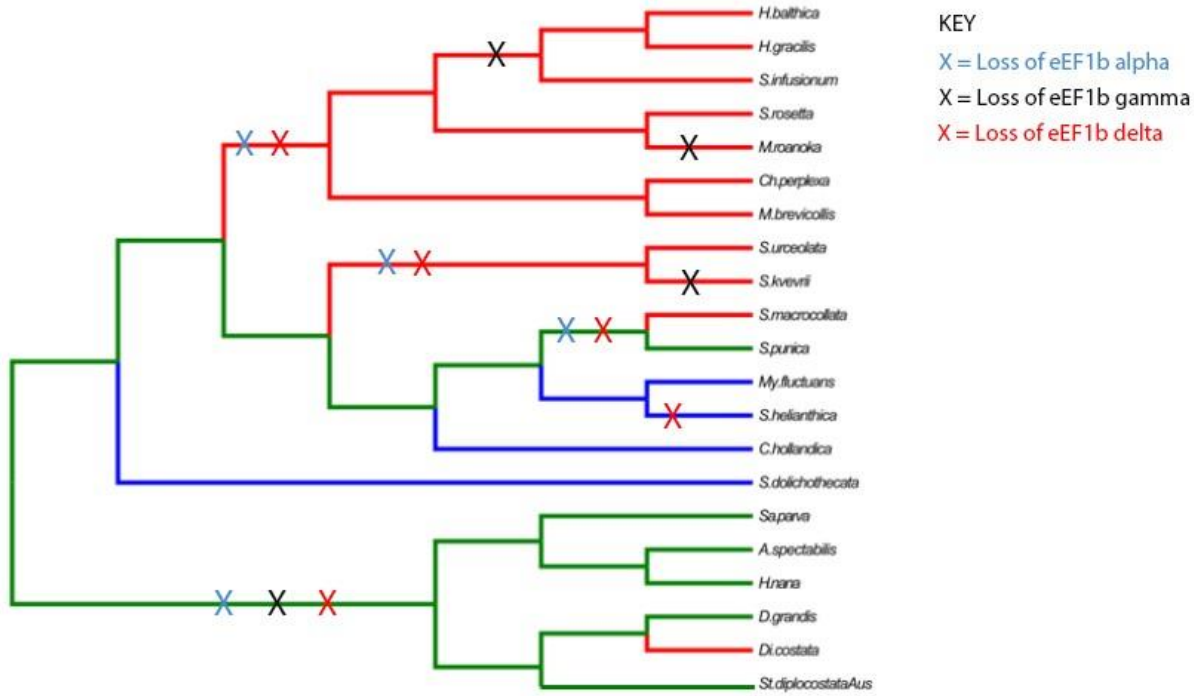


Figure 4.1: The gene loss for the genes involved in the formation of the eEF1b complex within the analysed choanoflagellate species. Crosses represent gene loss in the subsequent branches of the tree.

The explanation for the eEF1bγ gene being present within the Craspedida group in the EFL and dual encoding species, shows that if gene loss does not occur, then this protein must have other properties to enhance cellular processes. The eEF1bγ gene appears to encode a versatile protein and as such, possesses many roles or moonlighting properties. It has been shown that eEF1bγ's role in translation is to provide support through scaffolding means for eEF1 complex, but in *S. cerevisiae,* the deletion of the gene had little effect on protein synthesis (Kinzy et al., 2004). This may provide an explanation for how a species which does not encode the gene, shows little hindrance to the transcription levels. Another finding by the Kinzy et al. (2004) study, also showed that, cells may gain an extra advantage from the loss of eEF1bγ. They found that a deletion of the eEF1bγ gene in *S. cerevisiae* increased resistance to oxidative stress. It appears apparent that, many of the proteins that are involved with

transcription, possess the capabilities to perform many functions. This suggests that in the absence of these genes within a genome, other proteins must carry out the essential processes in the absence of the known protein for the purpose.

The findings of eEF1bδ in the choanoflagellates was a surprising find. The available data on this protein is limited. A search of the NCBI database provided few examples of this gene and/or protein data. With no examples of this gene within single celled eukaryotes provided by the database, suggesting this is the first example within a single celled eukaryote. The distribution within the choanoflagellates also suggests that the protein is elongation factor specific, as only eEF1a species transcribe the gene. This is also shown by the Sasikumar et al. (2012) study, which showed its presence in metazoan lineages which are all eEF1a species.

It is also apparent that the Acanthoecida species do not possess any of the genes for the formation of the eEF1b complex. This may be explained by the elongation factors utilised by the group. As all species are EFL or dual encoding species, and the findings in this study and that of Carr et al. (2017), in which the dual species utilise EFL for translation, suggests that species that utilise EFL do not require the eEF1b complex to function. This may suggest further work into other species which use EFL for translation to explore the EFL species from any taxonomic group, to investigate EFL translation in greater detail. This area of work is severely lacking in data, hence EFL studies must be undertaken to understand this area in greater detail.

4.3 The distribution of eEF2 within the choanoflagellates

As was shown in Fig.3.8, all species show copies of the eEF2 gene. This would be expected as of its important role in translation. This protein appears to be found in every organism due to the importance of its role within translation (Kaul & Rafeequi, 2011). Without this protein, translation would not be able to continue, and the function would provide devastating effects upon an individual with a mutation within this gene.

4.4 The distribution and function of eEF3 within the choanoflagellates

As was said in the Introduction, eEF3 was considered a fungal only elongation factor. The finding of the gene in other organisms, originally found by Ebstrup et al. (2005), in the organism *P. infestans*. Since this finding, eEF3 has been found in other organisms and subsequently coined eEF3L by the Murina et al. (2017) study. As this gene is found in both the Craspedida and Acanthoecida species, it appears an ancestor of both groups inherited this gene. As the tree in Figure 3.3 shows, the choanoflagellates appear to have obtained this gene through a horizontal gene transfer event. The evidence suggests the donor of the gene was of Chlorophyta origin. One other explanation for the horizontal gene transfer of this gene is through viral mediation. A study by Yamada et al. (1993) showed how the EF-3 gene in the protist virus species *Chlorella,* showed 63.6% identity and 92.4% identity to the *S. cerevisiae* EF3 protein, which may suggest the eukaryotic lineage was from a viral donor. Although in this study, the phylogenetic analysis showed the choanoflagellates were not grouped with the viral clades, the possibility of this mode of transfer cannot be ruled out. Even though the exact mechanism of transfer is not entirely known, the event must have happened before the divergence of the two choanoflagellate groups, and as such has been lost on five occasions within the Craspedids as has been shown in Figure 4.2.

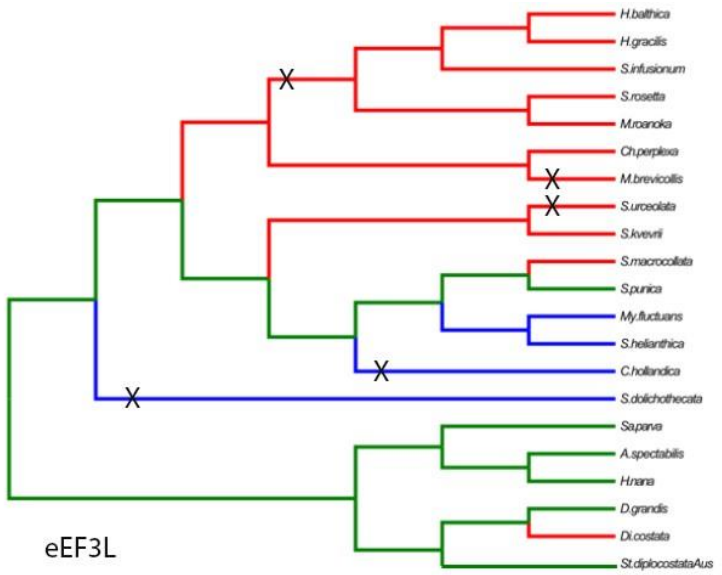


Figure 4.2: The analysis for eEF3L gene loss within the choanoflagellate species analysed. Crosses represent gene loss in the subsequent branches of the tree.

The fact that these species do harbour the eEF3L gene would indicate that the protein is utilised by the choanoflagellates. The purpose may be used in translation as in the fungal kingdom. This has been suggested by the Murina et al. (2017) study, using the data published by Spahn et al. (2006) in that the eEF3L in the eEF1a choanoflagellates, still possess the required domains for use in translation. This consists of the choanoflagellate eEF3L gene containing the HEAT domain and a putative peptide binding site, required for binding of the ribosome (Appendix 4.1). Also, the C-terminus in the eEF1a species *S. helianthica* and *M. fluctuans* possess the required EF1a binding C-terminus extension. Showing an area of polylysine/arginine rich area within the C-terminus. This also coincides with the findings of Spahn et al. (2006) and would allow the protein to bind both ribosome and EF1a. For the dual encoding species, *S. macrocollata* shows no HEAT domains, so would not bind the ribosome, but does have an area of polylysine and arginine richness. In *D. costata, S. kvevrii* and *C. perplexa* do possess the HEAT domains required for ribosomal binding (Appendix 4.1) and show areas of polylysine areas, but with little to no arginine within these areas. This suggests that in *S. macrocollata* this protein is no longer used in translational processes involving the ribosome. In the other species from the EFL species, it appears that the gene may still utilised in translation, the absence of arginine within the expected C-terminus area, may suggest that binding to EFL does not require arginine rich areas, and may utilise another amino acid. As work on EFL is limited, this may provide ideas for future research in the area. This can also be found in the dual encoding species, where the species *S. diplocostata* and *A. spectabilis* both contain the HEAT domain and the C-terminus extension, with polylysine areas, but little arginine within this area (Appendix 4.1). This also shows further support to the fact that dual encoding species utilise EFL as their elongation factor.

4.5 The ATPase ABCF genes distribution within the choanoflagellates

This group of ATPase enzymes were an unexpected finding, initially not part of this study. The identification of several high scoring -e values, during the eEF3 screening of the transcriptomes of the choanoflagellates, gave an interesting proposition. What were these genes that obviously are related to the EF3 sequence? Using a publicly available data set (Murina et al., 2017) the genes of interest were placed in a large phylogeny, containing a wide array of species, from all lineages of the eukaryotic

kingdom. The resulting phylogenetic tree was then used to identify which sequences were ABCF genes. The study by Murina et al. (2017) shows that there are 15 subfamilies within the ABCF group of ATPase enzymes. Of all the ABCF genes analysed four were used in this study to identify specific groups of ABCF genes, within the choanoflagellates.

4.5.1 The ABCF1 distribution and evolution within the choanoflagellates

ABCF1 was identified and the phylogeny can be seen in Fig. 3.4. With strong support through maximum likelihood and Bayesian probability scores throughout the tree, the tree could be classed as resolved. The choanoflagellate branches themselves, show areas of poor support. So, although the ABCF1 proteins have been identified correctly within the choanoflagellates, the inner relationship between the choanoflagellate ABCF1 phylogeny cannot be fully resolved. The positioning of the choanoflagellate genes in comparison to the known phylogeny of the taxonomic group matches, with closely related species such as *H. balthica* and *H. gracilis,* showing the same relationships as the most recent phylogeny (Carr et al., 2017), this would indicate that this phylogeny is correct. If the transcriptomic data is available, the addition of more choanoflagellate genes, not present in this phylogeny, may help to resolve the tree fully. This may suggest an area of future research, if/when the data is collected.

As far as gene loss with this protein, it has certainly been lost on two separate occasions. This is missing from the genomic data of both *M. Brevicollis* and *S. rosetta.* As the data for these two species is obtained from fully sequenced genomes, the absence of data for this gene may be due to the gene being found in a region of the genome that is difficult to fully sequence. As such, it may suggest future work for screening the available *S. rosetta* transcriptomic data, to see if the species transcribes for the protein and clarify this area. The other 19 species analysed in this study, only have transcriptomic data available at present. It also appears that two species have shown gene gain for ABCF1. The species *H. nana* and *A. spectabilis* have two distinctly different ABCF1 genes. One as with the others, shows a relationship to the metazoan proteins, but one copy has closer links to the *Perkinsus marinus* (alveolate) gene. As this copy is found in only these two species, coupled with their direct relationship phylogenetically (they are sister species within the choanoflagellate phylogeny), it appears that an ancestor between these two species gained the gene through a horizontal gene transfer event from a species belonging to the SAR supergroup. The other species in this study, within this clade within the most recent choanoflagellate phylogeny (Carr et al., 2017), *S. parva,* does not possess the SAR copy of ABCF1 gene. This suggests that the LCA between the *H. nana* and *A. spectabilis* and *S. parva* branch underwent a horizontal gene transfer event. It may also have been in an ancestor of both and subsequently lost in *S. parva.* The evidence to support that the exact ancestor to inherit the gene in a horizontal gene gain event is less parsimonious. More evidence to support this claim could be shown by the genomic data of the species *Polyoeca dichotoma*. This species appears on the same branch as *H. nana* and *A. spectablis*, this species may show further support to the hypothesis if it contains the horizontally gained ABCF1 gene.

4.5.2 The distribution and evolution of ABCF2 within the choanoflagellates

The ABCF2 gene was found in all species of choanoflagellate analysed, except in *D. grandis*, but this may have been due to the data for the species being transcriptomic, the gene may not have been transcribed at the time of data collection. With expression data now available on NCBI future work on this subject could be proposed. To understand if gene loss has occurred in *D. grandis* it could be shown that gene expression is variable, or if relatives of the species also show low expression of this gene. If

•

this is shown to be true, then a molecular approach using degenerate primers to identify the gene would have to be implemented to know if the species does possess the gene.

With all other species analysed an interesting pattern of distribution is shown. It is shown in Fig. 3.5 that two distinct groups can be seen. Clade 1 shows this protein to be related closely to the metazoan protein on clade 2 as a sister group to the Basidiomycota protein lineage within the Opisthokont super group. As choanoflagellates are positioned as a sister group to both the metazoan and fungal lineages within the tree of life and as both the Craspedida and Acanthoecida choanoflagellate groups are represented within clade 2 of this phylogeny, it appears that the ancestors of the choanoflagellates possessed two copies of the ABCF2 protein. This suggests that choanoflagellates had both copies of the gene and then the loss of one, left the other in place to carry out its role within the individual species. This shows that the ABCF2 gene has undergone two separate vertical inheritance events and substantial gene loss within the choanoflagellates, shown in Figure 4.3.

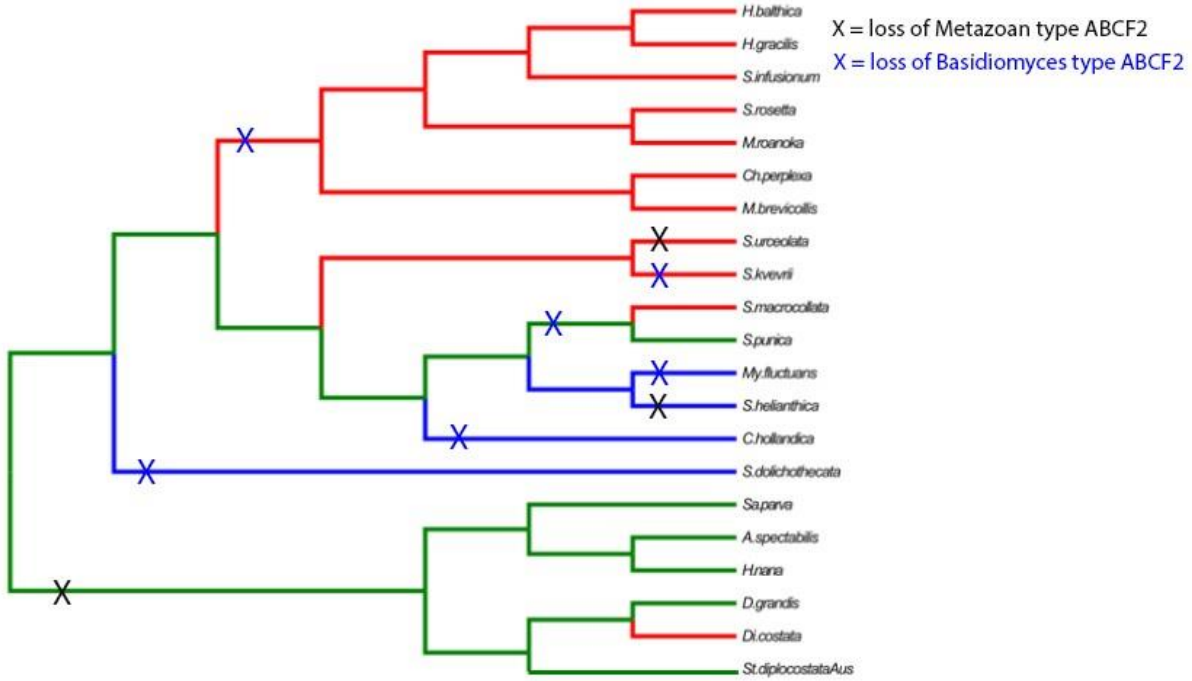


Figure 4.3: The gene loss of the metazoan type and Basidiomycota type ABCF2 genes within the choanoflagellates. Crosses represent gene loss in the subsequent branches of the tree.

4.5.3 The distribution and evidence of horizontal gene transfer of ABCF4 within the Acanthoecida

Fig.3.6 shows the phylogeny for the ABCF4 gene within the choanoflagellates. The data supports this gene being gained through a horizontal gene transfer event, within the Acanthoecida group of the choanoflagellates. It appears that an ancestor of this group inherited the gene from a species of Cryptophyta. With the Craspedida group not being represented within this phylogeny, suggests that it was an ancestor of this group which inherited the gene. Within this group it appears that the gene has been lost on two occasions. The Acanthoecida group is split into two separate branches, the group containing *S. parva, A. spectabilis* and *H. nana* all possess the gene. In the other grouping, only *D. grandis* possesses the gene. This shows that an ancestor of the two Acanthoecida subgroups possessed the gene, but subsequent loss in the two species *D. costata* and *S. diplocostata* has occurred. Strong evidence of horizontal gene transfer is also seen in the species *H. nana.* As this species appears to have two distinct copies of the gene, it would suggest that a more recent horizontal gene transfer event

•

happened within this species. The gene appears to have undergone horizontal transfer into *H. nana* from a species of Chlorophyta, related to *Ostreococcus tauri.*

4.5.4 The distribution and evidence of horizontal gene transfer of ABCF6 within the Acanthoecida

Shown in Fig. 3.7 is the phylogeny produced for the ABCF6 gene. As is shown, again in the Acanthoecida group of choanoflagellates, horizontal gene transfer has occurred. The LCA between the Acanthoecida choanoflagellates appears to have gained the gene through horizontal gene transfer. The evidence supports that the species the gene was inherited from was an ancestor to *Guillardia theta,* so appears to, again, be from Cryptophyta origin.

4.5.5 The conclusion to the ABCF gene lineages within the choanoflagellates

In this study it appears that of the four ABCF genes analysed, that the two genes ABCF1 and ABCF2 have been a constant lineage throughout eukaryotic evolution. This is shown by the gene being present consistently through the phylogeny of the eukaryotes. The ABCF1 and ABCF2 phylogenies also lend support to this idea, the phylogenies produced resemble the known phylogeny of species. This ancient gene lineage appears to play a vital role in translation in the eukaryotes, but as of yet, the exact role of these proteins is unknown. These genes are indicative of vertical inheritance, passed through the eukaryote's evolution from parent to progeny. In the cases of the ABCF2 (fungal copy), ABCF4 and ABCF6 genes it is apparent that horizontal gene transfer has occurred. It is proposed by this study that the gene gain shown in this study, was incorporated into the choanoflagellates horizontally. The first description of horizontal gene transfer in choanoflagellates was in a study by Bapteste et al. (2003). The enzyme phosphofructokinase was found to have been gained horizontally in the species *Monosiga ovata*. Later work went on to show how the choanoflagellate species *M. brevicollis* had 405 genes (4.4% of nuclear genome) of algal and prokaryotic origin (Yue et al., 2013).

In eukaryotes the fact that horizontal gene transfer, although controversial on the mechanisms, has occurred, is accepted. The main explanation is horizontal gene transfer from mitochondrial and plastid origin (Thomason et al., 2003; Rot et al., 2006). The study by Huang (2013), explains that the mitochondria and plastid acquisition of genes, could not adequately support the amount of foreign DNA within eukaryotic lineages alone. One other explanation for such gene gain was proposed by Doolittle (1998) and provides an explanation to the gene gain within the choanoflagellates. In this study, it is proposed that in species belonging to the amitochondriate protists, such as the *Giardia theta* species, gain horizontal gene transfer from the food source of the organisms. It was proposed that bacterivorous species have greater exposure to foreign DNA fragments during the phagocytosis of bacterial cells. The genes then become incorporated into the hosts genome and continue to their progeny. As choanoflagellates are bacterivorous and obtain cellular energy by consuming bacteria (King, 2005), this may explain how the species obtained foreign genes, within their genome, and may be in ancestors of the species gaining DNA in such a way, that has continued throughout evolution and into multicellular organisms.

The ABCF4 and ABCF6 genes within the choanoflagellates in this study appear to be from both Haptophyta and Cryptophyta origins. Both Haptophyta and Cryptophyta species are both contain species which help make up the phytoplankton found within marine, brackish and occasionally fresh water habitats globally (Reed et al., 2013; Curtis et al., 2012). As choanoflagellates also are found within the same environments as the phytoplankton and are ubiquitous to aquatic habitats (Carr et al., 2008), the encounters between these species must be extremely common.

•

4.6 The distribution of ribosomal genes within the choanoflagellate species

As mentioned earlier, the ribosomal part of this study was to investigate a small section of ribosomal proteins, to identify if patterns of distribution are apparent within the different elongation factor species. Of the small selection of genes analysed, no small ribosomal subunits were missing from the transcriptomic or genomic data of the choanoflagellates. There does appear to be gene loss within the large ribosomal subunits. Although not elongation factor specific, it does highlight the ribosomal genes are not uniform, even within the same taxon. This provides an area of further research to investigate the entire ribosomal genes of all the choanoflagellates with available data, to understand the ribosome make-up of all the species.

4.7 Conclusion

As a novel study into this area of evolutionary biology, the findings within this study give a broad over sight of the choanoflagellate translational machinery. The findings on the EF1a, EFL and dual encoding species, provides the groundwork for future study. In this study, areas of future research have been identified to further understand the translational machinery and its evolution within the group. Evidence throughout this study has shown that the dual encoding species show such strong resemblance to the EFL pattern of distribution of all genes analysed, that these species must utilise EFL as their elongation factor. This opens a line of enquiry as to what the EF1a gene is transcribed for within the dual encoding species. Through a molecular approach, continued from this study's work, the answer to this may be found, by expression studies and then continuation along the gene-protein pipeline. Another such area is in the ABCF proteins. As this study identified four classes of ABCF proteins and identified new genes that encode the proteins within this group, further identification of more ABCF proteins could be made. It was found in this study (not shown), that the choanoflagellates may harbour further ABCF genes. Time constraints hindered further analysis, but further ABCF genes were identified and left unresolved. These genes were the ABCF7 and Yhes ABCF genes. To continue this part of the study would be advantageous in gaining the knowledge of which proteins have been acquired and by what means. This may help understand the evolution of this family of proteins further, not only in the choanoflagellates, but in the eukaryotic tree in general. With gene loss being shown across the analysed genes, ways to confirm gene loss and gain is to be able to sequence the entire genomes of all species within this study. Alternatively, designing degenerate primers, specific to the gene of interest, could be designed. This would allow amplification through PCR of any genes not present in the transcriptome but are within the genomic data of the choanoflagellate species. Only then can the findings within this study be confirmed.

With gene gain within the choanoflagellates being apparent, this can only lend to the ideas produced by studies such as the Doolittle (1998) study. It appears that the predation habits of the choanoflagellates, has allowed the group to gain genes in which the individuals gain a benefit through horizontal gene transfer. With this finding, that the choanoflagellates have gained genes from other eukaryotic species, it may show further areas of interest, to identify further genes gained in such a manner. This may help identify and further advance knowledge in the area, to show that gene gain is not just from prokaryotic or viral to eukaryotic lineages. This mode of gene transfer may help to understand how multicellular genomes have been shown to contain foreign DNA but may have been gained in a variety of ways in a unicellular ancestor to the multicellular species (Huang, 2013).

It was apparent in this study also, that the ribosomal make-up of a species is variable. Within the choanoflagellates analysed, the large ribosomal proteins showed most variability. It is apparent that

•

to understand the role of the ribosomal genes, an independent study, focussed solely on the ribosomal genes of each species of choanoflagellates would be required to understand this principle and to fully understand the roles within each elongation specific species.

4.8 Molecular approach

The molecular biology approach to the EF1a protein within the dual encoding species proved a difficult challenge. The first problem encountered was with the species chosen for total RNA extraction, *D. grandis*. The cells obtained from Dr Carr proved impossible to culture into significant numbers for RNA extraction. Many attempts were made to correct the protocol and all attempts failed to produce the needed quantity of cell density. This may be an area for further research into the husbandry of the choanoflagellates, to identify optimal growth conditions for further analysis of the species.

Once the decision was made to have the *S. punica* gene synthesised, further experiments failed to produce any cloned colonies. The idea that the protein may be toxic to prokaryotic cells pushed the experiment forward. One cell line was used in which as shown, was successful in the cloning procedure. This, unfortunately for this study, was produced in the final experiment made during the process. The successful cloning has now left the possibility for further study. The cloned cells can now be continued forward into expression studies. If successfully expressed within an expression strain of *E. coli,* then the gene to protein pipeline may be utilised to analyse the function of eEF1a within the dual encoding species of choanoflagellates. This may then show what, if any, processes the eEF1a gene carries out within the dual encoding species. With the absence of eEF1b within the dual encoding species, which this study has shown, it is highly unlikely that eEF1a carries out elongation in the dual species. This provides an exciting prospect to carry the work forward, in showing what function this protein is utilised for. This may answer the question proposed at the beginning of this study, what functions the eEF1a enzyme provides to cells, other than elongation. This may shed further light on the moonlighting properties of not just EF1a, but opens up possibilities of study for other proteins and their moonlighting roles within the choanoflagellates.

References

New references

- Adl, S. M., Simpson, A. G. B., Lane, C. E., Lukeš, J., Bass, D., Bowser, S. S., . . . Spiegel, F. W. (2012). The revised classification of eukaryotes. Journal of Eukaryotic Microbiology, 59(5), 429-514. doi:10.1111/j.1550-7408.2012.00644.x
- Atkinson, G. C., Kuzmenko, A., Chicherin, I., Soosaar, A., Tenson, T., Carr, M., . . . Umeå universitet. (2014). An evolutionary ratchet leading to loss of elongation factors in eukaryotes. BMC Evolutionary Biology, 14(1), 35-35. doi:10.1186/1471-2148-14-35
- Bapteste, E., Moreira, D., & Philippe, H. (2003). Rampant horizontal gene transfer and phospho-donor change in the evolution of the phosphofructokinase. Gene, 318, 185-191. doi:10.1016/S0378-1119(03)00797-2
- Blanchard, S. C., Kim, H. D., Gonzalez, R. L., Puglisi, J. D., & Chu, S. (2004). tRNA dynamics on the ribosome during translation. Proceedings of the National Academy of Sciences of the United States of America, 101(35), 12893-12898. doi:10.1073/pnas.0403884101
- Brown, M. W., Heiss, A. A., Kamikawa, R., Inagaki, Y., Yabuki, A., Tice, A. K., . . . Roger, A. J. (2018). Phylogenomics places orphan protistan lineages in a novel eukaryotic super-group. Genome Biology and Evolution, 10(2), 427-433. doi:10.1093/gbe/evy014
- Budkevich, T. V., El'skaya, A. V., & Nierhaus, K. H. (2008). Features of 80S mammalian ribosome and its subunits. Nucleic Acids Research, 36(14), 4736-4744. doi:10.1093/nar/gkn424
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. BMC Bioinformatics, 10(1), 421-421. doi:10.1186/1471-2105-10-421
- Carr, M., B. S. C. Leadbeater, Hassan, R., Nelson, M., Baldauf, S. L., Teknisk-naturvetenskapliga vetenskapsområdet, . . . Institutionen för organismbiologi. (2008). Molecular phylogeny of choanoflagellates, the sister group to metazoa. Proceedings of the National Academy of Sciences of the United States of America, 105(43), 16641-16646. doi:10.1073/pnas.0801667105
- Carr, M., Richter, D. J., Fozouni, P., Smith, T. J., Jeuck, A., Leadbeater, B. S. C., & Nitsche, F. (2017). A six-gene phylogeny provides new insights into choanoflagellate evolution. Molecular Phylogenetics and Evolution, 107, 166-178. doi:10.1016/j.ympev.2016.10.011
- Curtis, B. A., Tanifuji, G., Burki, F., Gruber, A., Irimia, M., Maruyama, S., . . . Uppsala universitet. (2012). Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. Nature, 492(7427), 59-65. doi:10.1038/nature11681
- Dean, M., Rzhetsky, A., & Allikmets, R. (2001). The human ATP-binding cassette (ABC) transporter superfamily. Genome Research, 11(7), 1156-1166. doi:10.1101/gr.GR-1649R
- Denis, C. L., Richardson, R., Park, S., Zhang, C., Xi, W., Laue, T. M., & Wang, X. (2018). Defining the protein complexome of translation termination factor eRF1: Identification of four novel eRF1-containing complexes that range from 20S to 57S in size. Proteins: Structure, Function, and Bioinformatics, 86(2), 177-191. doi:10.1002/prot.25422
- Dever, T. E., Kinzy, T. G., & Pavitt, G. D. (2016). Mechanism and regulation of protein synthesis in saccharomyces cerevisiae. Genetics, 203(1), 65. doi:10.1534/genetics.115.186221
- Dong, J., Lai, R., Jennings, J. L., Link, A. J., & Hinnebusch, A. G. (2005). The novel ATP-binding cassette protein ARB1 is a shuttling factor that stimulates 40S and 60S ribosome biogenesis. Molecular and Cellular Biology, 25(22), 9859-9873. doi:10.1128/MCB.25.22.9859-9873.2005
- Doolittle, W. (1998). You are what you eat: A gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. Trends in Genetics, 14(8), 307-311. doi:10.1016/S0168-9525(98)01494-2
- Ebstrup, T., Saalbach, G., & Egsgaard, H. (2005). A proteomics study of in vitro cyst germination and appressoria formation in phytophthora infestans. Proteomics, 5(11), 2839.
- Flanagan, J. F., Namy, O., Brierley, I., & Gilbert, R. J. C. (2010). Direct observation of distinct A/P hybrid-state tRNAs in translocating ribosomes. Structure, 18(2), 257-264. doi:10.1016/j.str.2009.12.007
- Gile, G. H., Faktorová, D., Castlejohn, C. A., Burger, G., Lang, B. F., Farmer, M. A., . . . Keeling, P. J. (2009). Distribution and phylogeny of EFL and EF-1α in euglenozoa suggest ancestral co-occurrence followed by differential loss. PLoS One, 4(4), e5162. doi:10.1371/journal.pone.0005162
- Gobet, C., & Naef, F. (2017). Ribosome profiling and dynamic regulation of translation in mammals. Current Opinion in Genetics & Development, 43, 120-127. doi:10.1016/j.gde.2017.03.005
- Gromadski, K. B., Schümmer, T., Strømgaard, A., Knudsen, C. R., Kinzy, T. G., & Rodnina, M. V. (2007). Kinetics of the interactions between yeast elongation factors 1A and 1Balpha, guanine nucleotides, and aminoacyl-tRNA. The Journal of Biological Chemistry, 282(49), 35629.

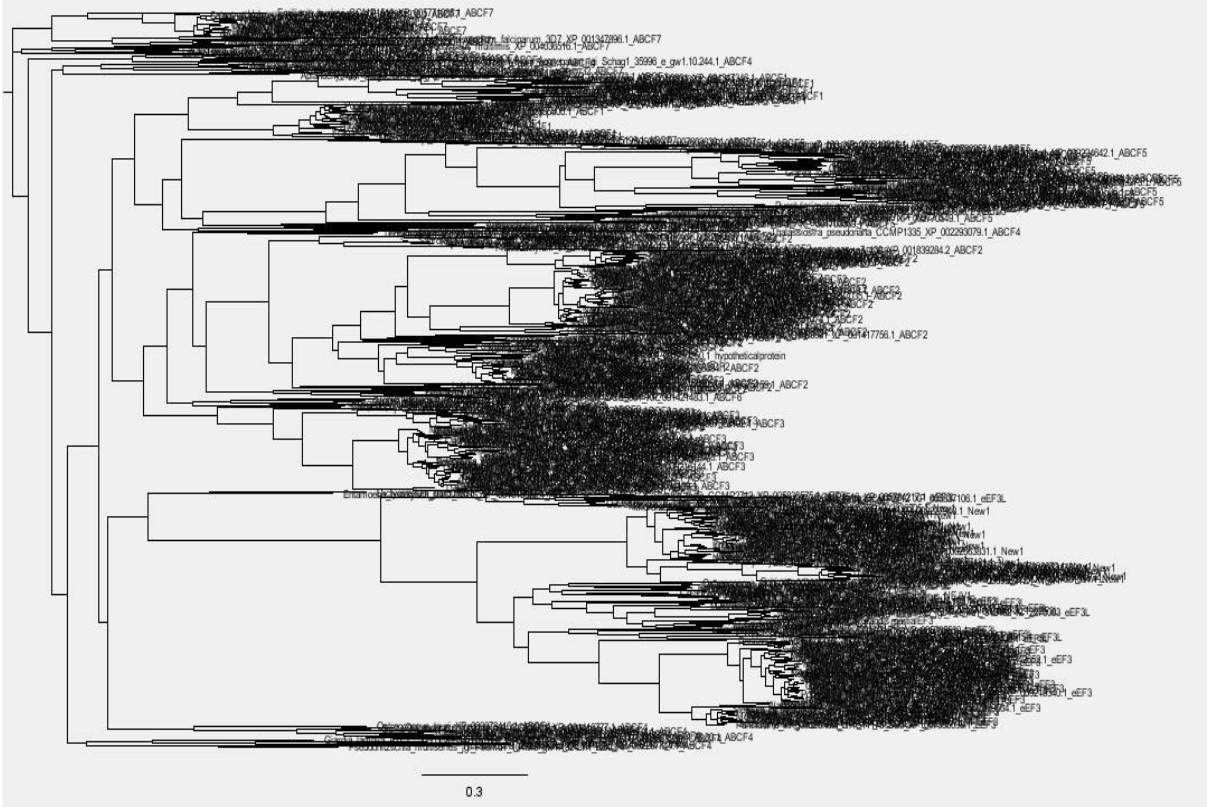
- Henk, D. A., & Fisher, M. C. (2012). The gut fungus basidiobolus ranarum has a large genome and different copy numbers of putatively functionally redundant elongation factor genes. PloS One, 7(2), e31268. doi:10.1371/journal.pone.0031268
- Hinnebusch, A. G., & Lorsch, J. R. (2012). The mechanism of eukaryotic translation initiation: New insights and challenges. Cold Spring Harbor Perspectives in Biology, 4(10), a011544-a011544. doi:10.1101/cshperspect.a011544
- Huang, C. Y., Chen, P., Huang, M., Tsou, C., Jane, W., & Anthony H. C. Huang. (2013). Tandem oleosin genes in a cluster acquired in brassicaceae created tapetosomes and conferred additive benefit of pollen vigor. Proceedings of the National Academy of Sciences, 110(35), 14480. doi:10.1073/pnas.1305299110
- Ito, K., Frolova, L., Seit-Nebi, A., Karamyshev, A., Kisselev, L., & Nakamura, Y. (2002). Omnipotent decoding potential resides in eukaryotic translation termination factor eRF1 of variant-code organisms and is modulated by the interactions of amino acid sequences within domain 1. Proceedings of the National Academy of Sciences of the United States of America, 99(13), 8494-8499. doi:10.1073/pnas.142690099
- Kamikawa, R., Brown, M. W., Nishimura, Y., Sako, Y., Heiss, A. A., Yubuki, N., . . . Inagaki, Y. (2013). Parallel re-modeling of EF-1α function: Divergent EF-1α genes co-occur with EFL genes in diverse distantly related eukaryotes. BMC Evolutionary Biology, 13(1), 131. doi:10.1186/1471-2148-13-131
- Kamikawa, R., Inagaki, Y., & Sako, Y. (2008). Direct phylogenetic evidence for lateral transfer of elongation factor-like gene. Proceedings of the National Academy of Sciences of the United States of America, 105(19), 6965-6969. doi:10.1073/pnas.0711084105
- Katoh, K., & Toh, H. (2010). Parallelization of the MAFFT multiple sequence alignment program. Bioinformatics (Oxford, England), 26(15), 1899-1900. doi:10.1093/bioinformatics/btq224
- Kaul, G., Pattan, G., & Rafeequi, T. (2011). Eukaryotic elongation factor-2 (eEF2): Its regulation and peptide chain elongation. Cell Biochemistry and Function, 29(3), 227-234. doi:10.1002/cbf.1740
- Keeling, P. J., Inagaki, Y., & Palmer, J. D. (2004). A class of eukaryotic GTPase with a punctate distribution suggesting multiple functional replacements of translation elongation factor 1α. Proceedings of the National Academy of Sciences of the United States of America, 101(43), 15380-15385. doi:10.1073/pnas.0404505101
- Kerr, I. D. (2004). Sequence analysis of twin ATP binding cassette proteins involved in translational control, antibiotic resistance, and ribonuclease L inhibition. Biochemical and Biophysical Research Communications, 315(1), 166-173. doi:10.1016/j.bbrc.2004.01.044
- Kim, E., & Graham, L. (2008). EEF2 analysis challenges the monophyly of archaeplastida and chromalveolata. PLoS One, 3(7) doi:10.1371/journal.pone.0002621
- King, N. (2005). choanoflagellates. Current Biology, 15(4), R113-R114. doi:10.1016/j.cub.2005.02.004
- Kinzy, T. G., Chowdhury, W. Q., Chatterjee, I., Ortiz, P. A., & Olarewaju, O. (2004). The translation elongation factor eEF1B plays a role in the oxidative stress response pathway. RNA Biology, 1(2), 89-94. doi:10.4161/rna.1.2.1033
- Mateyak, M. K., Pupek, J. K., Garino, A. E., Knapp, M. C., Colmer, S. F., Kinzy, T. G., & Dunaway, S. (2018). Demonstration of translation elongation factor 3 activity from a non-fungal species, phytophthora infestans. PloS One, 13(1), e0190524. doi:10.1371/journal.pone.0190524
- Merrick, W. C. (1992). Mechanism and regulation of eukaryotic protein synthesis. Microbiological Reviews, 56(2), 291-315
- Miller, M. A., Pfeiffer, W., & Schwartz, T. (2010). Creating the CIPRES science gateway for inference of large phylogenetic trees. Paper presented at the 1-8. doi:10.1109/GCE.2010.5676129
- Moore, P. (2009). The ribosome returned. Journal of Biology, 8(1), 8-8. doi:10.1186/jbiol103
- Murina, V., Atkinson, G., Hauryliuk, V., Reith, M., & Kasari, M. (2017). ABCF ATPases involved in protein synthesis, ribosome assembly and antibiotic resistance: Structural and functional diversification across the tree of life doi:10.1101/220046
- NITSCHE, F., CARR, M., ARNDT, H., & LEADBEATER, B. S. C. (2011). Higher level taxonomy and molecular phylogenetics of the choanoflagellatea. Journal of Eukaryotic Microbiology, 58(5), 452-462. doi:10.1111/j.1550-7408.2011.00572.x
- Nürenberg, E., & Tampé, R. (2013). Tying up loose ends: Ribosome recycling in eukaryotes and archaea. Trends in Biochemical Sciences, 38(2), 64-74. doi:10.1016/j.tibs.2012.11.003
- Paytubi, S., Wang, X., Lam, Y. W., Izquierdo, L., Hunter, M. J., Jan, E., . . . Proud, C. G. (2009). ABC50 promotes translation initiation in mammalian cells. Journal of Biological Chemistry, 284(36), 24061-24073. doi:10.1074/jbc.M109.031625

-
- Perez, W. B., & Kinzy, T. G. (2014). Translation elongation factor 1A mutants with altered actin bundling activity show reduced aminoacyl-tRNA binding and alter initiation via eIF2α phosphorylation. The Journal of Biological Chemistry, 289(30), 20928-20938. doi:10.1074/jbc.M114.570077
- Rambaut, A - http://tree. bio. ed. ac. uk/software/figtree/, 2009
- Rapp, G., Klaudiny, J., Hagendorff, G., Luck, M. R., & Scheit, K. H. (1989). Complete sequence of the coding region of human elongation factor 2 (EF-2) by enzymatic amplification of cDNA from human ovarian granulosa cells. Biological Chemistry Hoppe-Seyler, 370(10), 1071.
- Read, B. A., Kegel, J., Klute, M. J., Kuo, A., Lefebvre, S. C., Maumus, F., . . . Emiliania huxleyi Annotation Consortium. (2013). Pan genome of the phytoplankton emiliania underpins its global distribution. Nature, 499(7457), 209-213. doi:10.1038/nature12221
- Richter, D. J., Fozouni, P., Eisen, M. B., & King, N. (2018). Gene family innovation, conservation and loss on the animal stem lineage. Elife, 7 doi:10.7554/eLife.34226
- Richter, D. J., & King, N. (2013). The genomic and cellular foundations of animal origins.Annual Review of Genetics, 47(1), 509-537. doi:10.1146/annurev-genet-111212-133456
- Ronquist, F., Teslenko, M., Paul van der Mark, Ayres, D. L., Darling, A., Höhna, S., . . . Naturvetenskapliga fakulteten. (2012). MrBayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. Systematic Biology, 61(3), 539-542. doi:10.1093/sysbio/sys029
- Rot, C., Goldfarb, I., Ilan, M., & Huchon, D. (2006). Putative cross-kingdom horizontal gene transfer in sponge (porifera) mitochondria. BMC Evolutionary Biology, 6(1), 71-71. doi:10.1186/1471-2148-6-71
- Sadritdinova, A. F., Snezhkina, A. V., Dmitriev, A. A., Krasnov, G. S., Astakhova, L. N., Kudryavtsev, A. A., . . . Kudryavtseva, A. V. (2013). A new reference gene, Ef1A, for quantitative real-time PCR assay of the starfish asterias rubens pyloric ceca. Doklady Biological Sciences, 452(1), 310-312. doi:10.1134/S0012496613050050
- Sasikumar, A. N., Perez, W. B., & Kinzy, T. G. (2012). The many roles of the eukaryotic elongation factor 1 complex. Wiley Interdisciplinary Reviews: RNA, 3(4), 543-555. doi:10.1002/wrna.1118
- Serdyuk, I. N., & Galzitskaya, O. V. (2007). Disordered regions in elongation factors EF1A in the three superkingdoms of life. Molecular Biology, 41(6), 949-961. doi:10.1134/S002689330706012X
- Sharkey, L. K. R., Edwards, T. A., & O'Neill, A. J. (2016). ABC-F proteins mediate antibiotic resistance through ribosomal protection. Mbio, 7(2), e01975. doi:10.1128/mBio.01975-15
- Skogerson, L., & Engelhardt, D. (1977). Dissimilarity in protein chain elongation factor requirements between yeast and rat liver ribosomes. Journal of Biological Chemistry, 252(4), 1471.
- Skogerson, L., & Wakatama, E. (1976). A ribosome-dependent GTPase from yeast distinct from elongation factor 2. Proceedings of the National Academy of Sciences of the United States of America, 73(1), 73-76.
- Spahn, C. M. T., Kinzy, T. G., Becker, T., Pedersen, J. S., Balar, B., Beckmann, R., . . . Anand, M. (2006). Structure of eEF3 and the mechanism of transfer RNA release from the E-site. Nature, 443(7112), 663-668. doi:10.1038/nature05126
- Spilka, R., Ernst, C., Mehta, A. K., & Haybaeck, J. (2013). Eukaryotic translation initiation factors in cancer development and progression. Cancer Letters, 340(1), 9-21. doi:10.1016/j.canlet.2013.06.019
- Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics, 30(9), 1312-1313. doi:10.1093/bioinformatics/btu033
- Stapulionis, R., Kolli, S., & Deutscher, M. P. (1997). Efficient mammalian protein synthesis requires an intact F-actin system. Journal of Biological Chemistry, 272(40), 24980-24986. doi:10.1074/jbc.272.40.24980
- Susorov, D., Zakharov, N., Shuvalova, E., Ivanov, A., Egorova, T., Shuvalov, A., . . . Alkalaeva, E. (2018). Eukaryotic translation elongation factor 2 (eEF2) catalyzes reverse translocation of the eukaryotic ribosome. The Journal of Biological Chemistry, 293(14), 5220-5229. doi:10.1074/jbc.RA117.000761
- Thomason, B., Bergthorsson, U., Palmer, J. D., & Adams, K. L. (2003). Widespread horizontal transfer of mitochondrial genes in flowering plants. Nature, 424(6945), 197-201. doi:10.1038/nature01743
- Triana-Alonso, F. J., Chakraburtty, K., & Nierhaus, K. H. (1995). The elongation factor 3 unique in higher fungi and essential for protein biosynthesis is an E site factor. Journal of Biological Chemistry, 270(35), 20473-20478. doi:10.1074/jbc.270.35.20473
- Uritani, M., & Miyazaki, M. (1988). Role of yeast peptide elongation factor 3 (EF-3) at the AA-tRNA binding step. Journal of Biochemistry, 104(1), 118.
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., . . . Schwede, T. (2018). SWISS-MODEL: Homology modelling of protein structures and complexes. Nucleic Acids Research, 46(W1), W296-W303. doi:10.1093/nar/gky427

-
- Wilcox, S. M., Arora, H., Munro, L., Xin, J., Fenninger, F., Johnson, L. A., . . . Jefferies, W. A. (2017). The role of the innate immune response regulatory gene ABCF1 in mammalian embryogenesis and development. PloS One, 12(5), e0175918. doi:10.1371/journal.pone.0175918
- Wylezich, C., Karpov, S. A., Mylnikov, A. P., Anderson, R., & Jürgens, K. (2012). Ecologically relevant choanoflagellates collected from hypoxic water masses of the baltic sea have untypical mitochondrial cristae. BMC Microbiology, 12(1), 271. doi:10.1186/1471-2180-12-271
- Yamada, T., Fukuda, T., Tamura, K., Furukawa, S., & Songsri, P. (1993). Expression of the gene encoding a translational elongation factor 3 homolog of chlorella virus CVK2. Virology, 197(2), 742-750. doi:10.1006/viro.1993.1650
- Yue, J., Sun, G., Hu, X., & Huang, J. (2013). The scale and evolutionary significance of horizontal gene transfer in the choanoflagellate monosiga brevicollis. BMC Genomics, 14(1), 729. doi:10.1186/1471-2164-14-729
- Zagari, A., Masullo, M., Sica, F., Vitagliano, L., & Bocchini, V. (2001). The crystal structure of sulfolobus solfataricus elongation factor 1α in complex with GDP reveals novel features in nucleotide binding and exchange. The EMBO Journal, 20(19), 5305-5311. doi:10.1093/emboj/20.19.5305
- Zhang, X., Shi, D., Shi, H., Chen, J., Li, C., & Feng, L. (2014). EF1A interacting with nucleocapsid protein of transmissible gastroenteritis coronavirus and plays a role in virus replication. Veterinary Microbiology, 172(3-4), 443-448. doi:10.1016/j.vetmic.2014.05.034

Appendix

3.1: The large ABCF gene tree. Produced to analyse the different ABCF genes within the choanoflagellates.



•

4.1: The NCBI analysis of the eEF3L genes within the choanoflagellates. Showing the predicted putative HEAT domains. The panels represent (A) *S. helianthica* (B) *M. fluctuans* (C) *S. macrocollata* (D) *S. kvevrii* (E) *C. perplexa* (F) *S. diplocostata* (G) *A. spectabilis*

Any data produced in this thesis may be obtained upon request from the author.