Evolution of Nucleomorph and Plastid Genomes in the Chlorarachniophytes

A Dissertation Submitted to
the Graduate School of Life and Environmental Sciences,
University of Tsukuba
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy in Science
(Doctoral Program in Biological Sciences)

Shigekatsu SUZUKI

Table of Contents

Abstract

Plastids were acquired by eukaryotic cells via several endosymbiotic events between a heterotrophic eukaryote and a photosynthetic organism. Plants and some algae (glaucophytes, green, and red algae) acquired plastids via a single primary endosymbiosis event between a eukaryote and a cyanobacterium. The other algal groups have complex plastids that originated through the uptake of green and red algal endosymbionts via multiple secondary endosymbioses. These primary and secondary endosymbioses resulted in the remarkable diversity of photosynthetic eukaryotes throughout the tree of life. During endosymbiosis, the endosymbiont genes are lost, or transferred to the host genome, and the algal endosymbiont nuclei disappear in most cases. However, chlorarachniophytes and cryptophytes still possess a relict nucleus, known as the nucleomorph, of the green and red algal endosymbiont, respectively. Thus, they have two endosymbiotically-derived genomes, nucleomorph, and plastid genome, which can retain intermediate traits of the secondary endosymbiosis in the complex plastids. The nucleomorph and plastid genomes are an interesting and suitable model to study the reductive evolution of endosymbiotically-derived genomes. To date, studies on the nucleomorph and plastid genomes of chlorarachniophytes are highly limited compared to those of cryptophytes. The purpose of this study was to elucidate the detailed process of evolution of the endosymbiotically-derived genomes during secondary endosymbioses using chlorarachniophytes as model organisms.

Chapters 2 and 3 of this thesis include the analysis and comparison of three plastid genomes and two nucleomorph genomes of chlorarachniophytes, which elucidate the detailed processes of the reductive genome evolution. The plastid genomes were highly conservative *i.e.,* shared gene contents, gene orders, and genome sizes. However, lineage-specific intron losses might have caused the loss of a protein-coding gene. Furthermore, I also performed a phylogenetic analysis using plastid-encoded protein

genes to reveal a more specific phylogenetic position of chlorarachniophytes plastids compared to that in previous studies. It reveals that the chlorarachniophytes have acquired their plastids from an ulvophycean closely related to Bryopsidales. Four nucleomorph genomes shared 171 function-predicted genes, which are 86% of the total 198 function-predicted nucleomorph genes, including the same set of plastid-targeted protein genes. Furthermore, I could not detect any evidence for any endosymbiotic gene transfer (EGT) after the radiation of major lineages of chlorarachniophytes. These findings suggest that genome reduction with gene loss and EGT have mainly occurred before the radiation. The genomes were predicted to have had dynamic genome rearrangements after the radiation, which differs from the plastid genomes. The difference is probably because the nucleomorph genomes possess many open reading frames (ORFs) with unknown function (ORFans). The nucleomorph genomes of *Lotharella* species have unique duplicated genes, suggesting that their nucleomorph genomes have undergone secondary expansion due to gene duplication.

Nucleomorph genomes of chlorarachniophytes and cryptophytes possess genes with rapid evolutionary rates. I examined the relative gene expression levels of the nucleomorph-encoded protein genes of two chlorarachniophytes and three cryptophytes using RNA-seq. The genes of two heat-shock proteins, Hsp70 and Hsp90, were highly expressed under normal conditions. It has been shown that molecular chaperone overexpression allows the accumulation of genetic mutations in bacteria. The results of my study suggest that the overexpression of heat-shock proteins in nucleomorph genomes may play a role in buffering the mutational destabilization of proteins, which might enable the high evolutionary rates of nucleomorph-encoded proteins (Chapter 4).

Chapter 5 includes the gene expression profiles of nuclear- and nucleomorph-encoded protein genes of chlorarachniophytes throughout their cell cycles, which clarify the evolution of gene regulation

of the endosymbiont's proteins. In the nuclear-encoded genes, 75% plastid-targeted and 34% nucleomorph-targeted protein genes were differently expressed along the cell cycle, suggesting that the host cell controls the plastid and nucleomorph by regulating gene expression. In contrast, 99% of the nucleomorph-encoded protein genes did not exhibit gene expression patterns along the cell cycle. This is probably because most transcriptional regulatory regions are missing in the short intergenic regions of the nucleomorph genomes.

In this thesis, I describe the evolutionary processes of endosymbiont genome compaction during secondary endosymbioses using chlorarachniophytes. Comparative analysis of nucleomorph genomes reveals that the reductive evolution of nucleomorph genomes in chlorarachniophytes has mostly reached an endpoint, and that the genome reduction of chlorarachniophyte nucleomorphs has progressed further than that of cryptophytes. Furthermore, I report gene expression profiles of nuclear- and nucleomorph-encoded protein genes in chlorarachniophytes. These data suggest that the regulation of endosymbiont genes by the endosymbiont was transferred to the host during secondary endosymbiosis. This study contributes toward understanding the process of acquiring of endosymbionts in chlorarachniophytes and other algae with secondary plastids.

Abbreviations

AA/aa	Amino acid
ANOVA	A one-way analysis of variance
BIC	Bayesian information criterion
bp	Base pair
BP	Bootstrup probablilities
BPP	Bayesian posterior probabilities
ctDEG	Cytoplasm-targeted DEG
DEG	Differently expressed gene
EGT	Endosymbiotic gene transfer
ER	Endoplasmic reticulum
ERAD	ER associated degradation
EST	Expressed sequence tag
FDR	False discovery rate
Gbp	Gigabase pair
HEG	Highly expressed gene
IR	Inverted repeat
kbp	Kilobase pair
L/D cycle	Light/Dark cycle
LEG	Lowly expressed gene
LGT	Lateral gene transfer
LSC region	Large single-copy region

LSU	Large subunit
MCMCMC	Metropolis-coupled Markov chain Monte Carlo
ML	Maximum likelihood
mRNA	Messenger RNA
mtDEG	Mitochondrion-targeted DEG
ncRNA	non-coding RNA
nt	Nucleotide
ORF	Open reading frame
OTU	Operational taxonomic unit
PCR	Polymerase chain reaction
PFGE	Pulsed-field gel electrophoresis
PPC	Periplastidal compartment
ptDEG	Plastid-targeted DEG
rDNA	Ribosomal DNA
RPKM	Reads per kilobase of exon per million mapped reads
rRNA	Ribosomal RNA
RT	Reverse transcriptase
RT-qPCR	Quantative reverse transcription PCR
RuBisCO	Riblose 1,5-bisphosphate carboxylase/oxygenase
SELMA	Symbiont-specific ERAD-like machinary
snRNA	Small nuclear RNA
SP	Signal peptide

SSC region	Short single-copy region
SSU	Small subunit
TIC	Translocons on the inner chloroplast membrane
TOC	Translocons on the outer chloroplast membrane
TPL	Transit peptide-like sequence
tRNA	Transfer RNA
UTC clade	Ulvophyceae-Trebouxiophyceae-Chlorophyceae clade

1. General introduction

1.1. Plastid acquisition via endosymbiosis, and genome reduction of the endosymbionts

Plastids were acquired via several endosymbiotic events between a heterotrophic eukaryote and a photosynthetic organism. Plants and some algae (glaucophytes, green, and red algae) acquired plastids via a single primary endosymbiosis between a eukaryote and a cyanobacterium (Rodriguez-Ezpeleta et al. 2005; Price et al. 2012). The other algal groups (chlorarachniophytes, cryptophytes, dinoflagellates, euglenophytes, haptophytes, and heterokonts) and a non-photosynthetic parasite group (apicomplexans) have complex plastids that originated by the uptake of green and red algal endosymbionts via multiple secondary endosymbiosis events (Ishida 2005; Gould et al. 2008; Archibald 2009; Keeling 2010). These primary and secondary endosymbioses resulted in the remarkable diversity of photosynthetic eukaryotes across the tree of life.

During the endosymbiotic events, genomes of the symbionts are reduced in size and in gene content (Palmer 1997). In plastid genomes, the genome sizes are generally less than ~500 kbp (Green 2011), and most gene-rich plastid genomes of red algae have ~250 protein-coding genes (Janouškovec et al. 2013); however, others have fewer protein genes (Martin and Herrmann 1998; Green 2011), which are more reductive than the genomes of cyanobacteria. During endosymbiosis, a vast amount of the endosymbionts' genes have been deleted, or transferred to the host genomes, via endosymbiotic gene transfer (EGT) (Martin and Herrmann 1998; Timmis et al. 2004). During secondary endosymbiosis, the nuclei of the endosymbionts are reduced, and finally deleted. Currently, most of the algae which have acquired plastids via secondary endosymbiosis lack endosymbiont nuclei; however, chlorarachniophytes and cryptophytes retain the endosymbiont nuclei, known as "nucleomorph" (Greenwood 1974; Hibberd

and Norris 1984) (Fig. 1.1). Thus, nucleomorphs can be considered as an "intermediate stage" of the secondary endosymbiosis process (Gilson et al. 2006).

1.2. Nucleomorph of chlorarachniophytes and cryptophytes

Chlorarachniophytes and cryptophytes retain nucleomorphs, but the origin of their plastids differ (Fig. 1.1). Chlorarachniophytes acquired complex plastids by the ingestion of a green algal endosymbiont. The endosymbiont is closely related to the ulvophyceae-trebouxiophyceae-chlorophyceae (UTC) clade (van de Peer et al. 1996; Ishida et al. 1997, 1999; Rogers et al. 2007; Tanifuji, Onodera, Brown et al. 2014), and the host is a cercozoan, such as the small predator *Minorisa minuta* of the Rhizaria supergroup (del Campo et al. 2012). In contrast, cryptophytes acquired plastids from red algae (Douglas et al. 1991; Douglas and Penny 1999). The nucleomorph is localized in the periplastidal compartment (PPC), the space between the inner and outer pair of plastid membranes, which is the remnant cytoplasm of the endosymbiont (Gillott and Gibbs 1980; Hibberd and Norris 1984).

The nucleomorphs of chlorarachniophytes and cryptophytes possess nucleomorph genomes (Hansman et al. 1985; Ludwig and Gibbs 1985, 1987), which have interesting common features for genome reductive evolution despite their different origins. Pulsed-field gel electrophoresis (PFGE) analyses reveals that all known nucleomorph genomes consist of three chromosomes, of which the predicted genome sizes range from 330 to 1033 kbp, and from 495 to 750 kbp in chlorarachniophytes, and cryptophytes, respectively (Eschbach et al. 1991; Rensing et al. 1994; Gilson and McFadden 1999; Lane and Archibald 2006; Silver et al. 2007; Phipps et al. 2008; Tanifuji et al. 2010; Ishida et al. 2011). The genome sizes have been highly reduced compared to the nuclear genomes of green algae or red algae during the secondary endosymbiotic processes.

The nucleomorph genomes are reductive; however, they retain essential transcriptional and translational activities. Several EST analyses show that genes of nucleomorph genomes are transcribed (Archibald et al. 2003; Williams et al. 2005; Slamovits and Keeling 2009), and an RNA-seq analysis shows that more than 99% of nucleomorph genomes, including intergenic regions, are transcribed, and the transcription levels are higher than those of nuclear genes (Tanifuji, Onodera, Moore et al. 2014). An *in situ* hybridization analysis indicates that SSU rRNA is located in the PPC of a chlorarachniophyte *Chlorarachnion reptans*, suggesting that transcripts of nucleomorph genomes are translated in the PPC (McFadden et al. 1994). Indeed, parts of the nucleomorph-encoded proteins of a chlorarachniophyte *Bigelowiella natans* are detected by proteomic analysis using isolated plastids (Hopkins et al. 2012).

1.3. EGT and protein transportation in the chlorarachniophytes and cryptophytes

Chlorarachniophytes and cryptophytes have complex plastids with nucleomorphs, suggesting that many proteins for nucleomorphs, PPC, and plastids have been encoded by the nuclear genomes via EGT during the secondary endosymbiotic processes. The nuclear genomes of a chlorarachniophyte *Bigelowiella natans* and a cryptophyte *Guillardia theta* have been sequenced recently, and are predicted to possess 694 plastid- and 1,002 PPC/nucleomorph-targeted proteins, and 755 plastid- and 2,401 PPC/nucleomorph-targeted proteins, respectively (Curtis et al. 2012) (Fig. 1.1). The nuclear genome sequences contain parts of the mitochondria genome sequences, and a fragmented *B. natans* cytochrome c oxidase (*cox1*), which is generally encoded in the mitochondria genome, is located in the first intron of an alpha subunit paralog of a guanine nucleotide-binding protein gene (Curtis and Archibald 2011). In contrast, the nuclear genomes do not possess any current nucleomorph and plastid genome sequences, suggesting that ongoing EGTs do not originate from the nucleomorphs or plastids to the nucleus (Curtis

et al. 2012). *B. natans* possesses only single plastid and nucleomorph per cell, and it is not able to digest the organelles for EGT. Chlorarachniophytes acquired their plastids from a green alga; however, the origins of the nuclear-encoded endosymbiont's proteins are divergent. Plastid-targeted proteins of *B. natans* are composed of some components of "red lineage" or bacterial proteins, probably because *B. natans* is mixotrophic, and able to engulf such organisms (Archibald et al. 2003). The same aspects are also observed in an amoeboid chlorarachniophyte *Amorphochlora amoebiformis* (Yang et al. 2014). Nuclear genomes of *B. natans* and *G. theta* also reveal a phylogenetically mosaic genome, which possess both "green lineage" and "red lineage" genes (Curtis et al. 2012).

Most of the endosymbiont's genes, which have been acquired via EGT, are transported into plastids, PPC, and nucleomorphs. The plastid-targeted protein genes were transported to their proper subcellular locations following signal peptides (SP) and transit peptide-like sequence (TPL) at the N-termini of the proteins in a chlorarachniophyte *B. natans* (Rogers et al. 2004). The SPs of plastid-targeted proteins are rich in hydrophobic amino acid residues, like SPs of ER (endoplasmic reticulum) -targeted proteins. The TPL sequences are difficult to predict because of their poor conservation (Rogers et al. 2004). A transient genetic transformation analysis using *A. amoebiformis* reveals that the SPs need to transport proteins into the ER, and TPL sequences need to transport proteins from the ER to the plastid stroma (Hirakawa et al. 2009). PPC-targeted proteins also possess similar N-terminal bipartite sequences (Gile and Keeling 2008), but the TPL of plastid- and PPC-targeted proteins have a different electrical charge (Hirakawa et al. 2009, 2010, 2011). Such N-terminal bipartite sequences are observed in many organisms with secondary plastids originating from red algae, including cryptophytes (Stork et al. 2013). The TPL sequences of cryptophytes often have Phe at the +1 position after the predicted cleavage sites of the SPs, and Phe is important for distinguishing the destination of the

proteins. If proteins have Phe in the TPL sequences, the proteins are transported into the plastid, and if not, they remain in the PPC (Gould, Sommer, Hadfi et al. 2006; Gould, Sommer, Kroth et al. 2006).

To date, the plastid transport system in plastid membranes of chlorarachniophytes is not known (Hirakawa et al. 2012); however, it has been shown nucleomorph genomes encode homologs of translocons of the plastid inner two membranes, *tic22* and *toc75* (Gilson et al. 2006). In cryptophytes, the plastid transport system is more known than chlorarachniophytes. Plastids of cryptophytes are surrounded by four membranes. Initially, plastid-targeted proteins pass through the outermost membrane using the Sec61-complex (Wastl and Maier 2000). The second outermost membrane has a symbiont-specific ERAD-like machinery (SELMA), which is an ERAD (ER associated degradation) -like system for protein transport into ERs; plastid-targeted proteins pass through this membrane using the SELMA (Sommer et al. 2007; Hempel et al. 2009). Several components of the SELMA are encoded by the nucleomorph genomes (Hempel et al. 2009). The inner two membranes use the TIC/TOC (translocons on the inner/outer chloroplast membrane) system (Stork et al. 2013), and a component of TIC, Tic22, is encoded by the nucleomorph genome (Douglas et al. 2001).

1.4. Endosymbiont's genome evolution of cryptophytes

1.4.1. Plastid genomes of cryptophytes

To date, plastid genome sequences are reported in four cryptophytes, *Guillardia theta*, *Rhodomonas salina*, *Cryptomonas paramecium*, and *Teleaulax amphioxeia* (Douglas and Penny 1999; Khan et al. 2007; Donaher et al. 2009; Kim et al. 2015). In these cryptophytes, only *C. paramecium* is a non-photosynthetic species, which lacks photosynthetic skills incidentally. Plastid genomes of the photosynthetic cryptophytes range from 121.5 kbp to 135.9 kbp, which are large for secondary plastid

genomes (Khan et al. 2007). The genomes possess 179 to 183 genes, including 143 to 147 protein-coding genes. Most of the protein-coding genes are shared among the photosynthetic cryptophytes; however, *R. salina* has three unique pseudogenes of light-independent chlorophyllide reductase, *chlB*, *chlN*, and *chlL*, which are not found in the other cryptophytes (Khan et al. 2007; Kim et al. 2015). The plastid genome of *R. salina* has two group II introns, none of which are present in the other cryptophytes (Khan et al. 2007). One of them resides within *psbN*; however, the others are degenerated, and located between *ycf37* and *ycf12*. Both the introns have an intronic reverse transcriptase (RT)-like ORF, but the ORF within the *ycf37-ycf12* intron is a pseudogene. The RT-like ORF is not identified in the plastid genomes of the other cryptophytes (Kim et al. 2015). The plastid genome of the non-photosynthetic cryptophyte *C. paramecium* lacks many genes related to photosynthesis (Donaher et al. 2009). The gene encoding the β subunit of phycoerythrin (*cpeB*), and photosynthetic regulator and electron transfer component *ftrB*, are missing in *C. paramecium*. The plastid genome of *C. paramecium* lacks all genes encoding the protein subunits of photosystem I (*psa*) and II (*psb*), and most *pet* protein family genes required for oxygenic photosynthesis. Plastid genome structure of all three photosynthetic cryptophytes is highly conserved in genome rearrangements (Kim et al. 2015). For *C. paramecium*, the plastid genome is slightly disrupted compared to the others, and it differs from the photosynthetic cryptophytes with three large inversions.

The plastid genomes of *R. salina* and *T. amphioxeia* have a gene, *dnaX*, which encodes the tau/gamma components of bacterial DNA polymerase (Khan et al. 2007; Kim et al. 2015). A phylogenetic analysis reveals that the *dnaX* genes of *R. salina* and *T. amphioxeia* have been acquired by lateral gene transfer (LGT). The *dnaX* genes are monophyletic with some bacteria, indicating that this lineage has acquired the gene from bacteria. In LGT of plastid genomes, genes encoding proteobacteria-derived RuBisCO large and small subunits (Delwiche and Palmer 1996), and non-cyanobacterial type *rpl36* (Rice

and Palmer 2006) are identified.

1.4.2. Nucleomorph genomes of cryptophytes

Douglas et al. (2001) first reported the complete nucleomorph genome of the cryptophyte *G. theta*. To date, nucleomorph genomes of four cryptophytes, *G. theta*, *H. andersenii*, *C. paramecium*, and *Chroomonas mesostigmatica* are completely sequenced (Douglas et al. 2001; Lane et al. 2007; Tanifuji et al. 2011; Moore et al. 2012). Despite being a non-photosynthetic species, *C. paramecium* possesses the slightly smaller nucleomorph genome (485.9 kbp) composed of three chromosomes (Tanifuji et al. 2011). Their genome sizes range from 485.9 kbp to 702.9 kbp, and *C. mesostigmatica* has the largest nucleomorph genome in them. Four nucleomorph genomes are composed of three chromosomes with rRNA clusters (5'–18SrRNA-5.8SrRNA-28SrRNA-5SrRNA–3') at each subtelomeric region. Interestingly, *H. andersenii* lacks the rRNA clusters at three chromosomal termini (Lane and Archibald 2006; Lane et al. 2007). In contrast, *C. mesostigmatica* possesses longer subtelomeric repeats, including rRNAs and several ORFs (Moore et al. 2012). The nucleomorph genomes of four cryptophytes are compact with high gene density (0.83–1.09 genes/kbp); however, the intergenic length ranges from 93 bp to 200 bp, which is the most significant factor in genome size variation (Moore et al. 2012). The nucleomorph genomes of *G. theta*, *C. paramecium*, and *C. mesostigmatica* have a small number of introns (2–24 introns). *H. andersenii* possesses no introns (Lane et al. 2007), suggesting that the intron loss occurred uniquely along the *Hemiselmis* lineage (Moore et al. 2012). This intron loss might have been derived from the loss of spliceosome-related protein genes, and five spliceosome-specific snRNAs, U1, U2, U4, U5, and U6 (Lane et al. 2007).

The nucleomorph genomes possess 466–505 genes encoding proteins, most of which have

housekeeping functions, *e.g.,* translation, transcription, and DNA metabolism. Particularly all photosynthetic cryptophytes share an identical set of 31 plastid-associated protein genes, although non-photosynthetic *C. mesostigmatica* lacks 13 genes of these. This suggests that this set of plastid-associated protein genes was highly conserved in the common ancestor of cryptophytes (Moore et al. 2012). Approximately 70% of the function-predicted genes are shared in four cryptophytes; however, some of their content is missing in the specific lineage. *C. paramecium* lacks portions of the plastid-associated protein genes (Tanifuji et al. 2011), *H. andersenii* lacks all of the spliceosome-related protein genes (Lane et al. 2007), and *C. mesostigmatica* lacks most of the proteasome subunit genes (Moore et al. 2012). The nucleomorph genomes have many hypothetical ORFs with unknown function, which are not homologous to known genes, so-called ORFans. The ORFans account for approximately 30% of the genes in the nucleomorph genome of *C. mesostigmatica* (Moore et al. 2012).

Gene order of the nucleomorph genomes is moderately conservative within the syntenic blocks of the four cryptophytes. The syntenic blocks consist of 6.7–19.1 genes on average between two species (Moore et al. 2012), and cover the entire nucleomorph genomes (Lane et al. 2007; Tanifuji et al. 2011; Moore et al. 2012). This is probably because the frequent recombination, which derives the recombination-mediated disruption of ORFs, is prevented in compact genomes, like the nucleomorph genomes composed of short spacers and single-copy genes (Archibald and Lane 2009). Lane et al. (2007) proposes syntenic ORFs, which are pairs of ORFs without homology, but with similar sizes on the same loci within syntenic blocks. This likely suggests that the pairs of syntenic ORFs originated in the same genes.

The nucleomorph genomes of the cryptophytes are highly reductive in gene content, with fewer non-coding regions. Additionally, they have lineage-specific gene loss, indicating that the

cryptophyte nucleomorph genome reduction has not yet reached an endpoint (Moore et al. 2012).

1.5. Chlorarachniophytes as a model organism to understand evolution during secondary endosymbiosis

Chlorarachniophytes are a small algal group presently composed of eight genera and 14 species (Hirakawa 2014). Chlorarachniophytes acquired their plastids from a green alga (described above), but the plastids are different lineage to that of euglenophytes (Rogers et al. 2007). Chlorarachniophytes have several cell forms during their cell cycle: coccoids, amoeboid cells, and flagellates (*e.g.,* Ota et al. 2005). In a model chlorarachniophyte, *B. natans*, flagellates are dominant in the cell cycle (Moestrup and Sengco 2001). The *B. natans* cells typically have a single plastid and nucleomorph per cell; therefore, cell division is accompanied by the division of the plastid and nucleomorph. By controlling the light/dark cycle, *B. natans* cells can be cultivated in synchrony, and the synchronized cells divide during the dark phase (Hirakawa et al. 2011). An amoeboid chlorarachniophyte, *Amorphochlora amoebiformis*, can be used for transient gene transformation using a particle bombardment method (Hirakawa et al. 2008).

1.6. Purpose of this study

To date, nucleomorph and plastid genomes have been completely sequenced for two chlorarachniophytes, *B. natans* and *Lotharella oceanica* (Gilson et al. 2006; Rogers et al. 2007; Tanifuji, Onodera, Brown et al. 2014). However, this genetic information is insufficient to understand the detailed organelle genome evolution in the chlorarachniophytes compared with cryptophytes. Recently, the nuclear genome of *B. natans* also has been sequenced, and reveals that the nuclear genome encodes a vast amount of endosymbiont-targeted proteins (Curtis et al. 2012). These genes aid in understanding the detailed process

of endosymbiont enslavement; however, there is no evidence that the endosymbiont-targeted protein genes of the nuclear or nucleomorph genome are transcriptionally controlled.

In this study, I reveal: (i) the detailed process employed in minimizing organelle genomes by analyzing the additional three plastid and two nucleomorph genomes in chlorarachniophytes, (ii) that the expression level of nucleomorph genes changed with the reductive evolution of the nucleomorph genomes, using an RNA-seq based gene expression analysis, and (iii) the regulation of the transcriptional patterns of the endosymbiont-targeted protein genes of the nuclear genome, using serial analyses of gene expression during the cell cycle. This study contributes toward understanding the process of acquiring of endosymbionts in chlorarachniophytes and other algae with secondary plastids.

2. Plastid genome sequences of *Gymnochlora stellata*, *Lotharella vacuolata*, and *Partenskyella glossopodia* reveal remarkable structural conservation among chlorarachniophyte species

2.1. Abstract

Chlorarachniophyte algae have complex plastids acquired by the uptake of a green algal endosymbiont, and this event is called secondary endosymbiosis. Interestingly, the plastids possess a relict endosymbiont nucleus, referred to as the nucleomorph, in the intermembrane space, and the nucleomorphs contain an extremely reduced and compacted genome in comparison with green algal nuclear genomes. Therefore, chlorarachniophyte plastids consist of two endosymbiotically-derived genomes, *i.e.*, the plastid and nucleomorph genomes. To date, complete nucleomorph genomes have been sequenced in four different species, whereas plastid genomes have been reported in only two species in chlorarachniophytes. To gain further insight into the evolution of endosymbiotic genomes in chlorarachniophytes, I newly sequenced the plastid genomes of three species, *Gymnochlora stellata*, *Lotharella vacuolata*, and *Partenskyella glossopodia*. My findings reveal that chlorarachniophyte plastid genomes are highly conserved in size, gene content, and gene order among species, but their nucleomorph genomes are divergent in such features. Accordingly, the current architecture of the plastid genomes of chlorarachniophytes evolved in a common ancestor, and changed very little during their subsequent diversification. Furthermore, my phylogenetic analyses using multiple plastid genes suggest that chlorarachniophyte plastids are derived from a green algal lineage that is closely related to Bryopsidales in the Ulvophyceae group.

2.2. Introduction

Chlorarachniophytes have acquired plastids from a green alga via the secondary endosymbiosis (see general discussion). Chlorarachniophyte plastids contain two different genomes, a nucleomorph and plastid genome, which have existed over the same evolutionary time after the secondary endosymbiosis. It should be interesting to investigate and compare evolutionary histories of these two genomes in chlorarachniophytes. Whereas four nucleomorph genomes have been sequenced so far, complete plastid genome sequences are available for only two chlorarachniophytes, *Bigelowiella natans* (Rogers et al. 2007) and *Lotharella oceanica* (Tanifuji, Onodera, Brown et al. 2014). To gain further insight into endosymbiotic genome evolution in chlorarachniophytes, I sequenced the plastid genomes of three species, *Gymnochlora stellata*, *Lotharella vacuolata*, and *Partenskyella glossopodia*. Comparative analyses of five chlorarachniophyte plastid genomes revealed that they were highly conserved in size, gene content, and gene order, although their nucleomorph genomes are divergent in these features. Architectural conservation of these plastid genomes may be related to their high gene density because frequent rearrangements are likely to disrupt the coding sequences. A remarkable finding was the presence of group I and II introns in the plastid genomes of four chlorarachniophytes, but not *B. natans*, suggesting that the loss of introns occurred in at least one lineage during the reductive evolution of chlorarachniophyte plastid genomes. Furthermore, I performed phylogenetic analyses using multiple plastid genome-encoded proteins, suggesting that chlorarachniophyte plastids are derived from a green algal lineage that was closely related to Bryopsidales in the group of Ulvophyceae.

2.3. Materials and methods

2.3.1. DNA extraction and plastid genome sequencing

G. stellata CCMP2053 and *P. glossopodia* RCC365 were cultivated at 20ºC under white illumination (80–100 µmol photon/m^2) on a 14:10 h light:dark cycle in 250- to 500-mL flasks containing ESM medium (Kasai et al. 2009) or IMK medium (Wako Pure Chemical Industries, Ltd., Osaka, Japan). Cultivation and DNA extraction of *G. stellata* were performed by Dr. Hirakawa (University of Tsukuba). The cells were collected by gentle centrifugation from two- to three-week-old cultures. The total DNA was extracted by a standard phenol-chloroform protocol and plastid DNA was isolated by Hoechst dye-cesium chloride density gradient ultracentrifugation at 50,000 rpm for 24 hours with a Vti 65.2 rotor (Beckman Coulter, Inc., Brea, CA, USA). Plastid DNA of *P. glossopodia* was also separated using pulsed-field gel electrophoresis according to the methods described in Ishida et al. (2011), and the DNA was purified from the gels using a GELase Agarose Gel-Digesting Preparation Kit (Epicentre, Illumina, Inc., Madison, WI, USA). Plastid DNA of *G. stellata* was Sanger sequenced and assembled by Dr. Sugita (Nagoya university) and Dr. Kofuji (Kanazawa university) at the National Institute of Genetics in Japan. To fill the gaps between resulting contigs, multiple polymerase chain reactions (PCR) were performed, followed by sequencing with an ABI 3130 Genetic Analyzer. Plastid DNA of *P. glossopodia* was sequenced by three runs using the 454 GS Junior System (454 Life Sciences, Roche Co., Branford, CT, USA) and one run using the Illumina HiSeq2000. The resulting 267,551 single-end reads from the GS Junior were assembled using Newbler v.2.5 (454 Life Sciences, Roche Co.), and three small gaps were closed by PCR. A dGTP BigDye Terminator Cycle Sequencing Kit (Applied Biosystems, Life Technologies, Waltham, MA, USA) was used to analyze some PCR frgments that could not be sequenced by a BigDye Terminator v3.1 Kit. The 49,531,305 paired-end reads from the Illumina HiSeq2000 run

were used to correct GS Junior pyrosequencing errors. The plastid genome sequence of *Lotharella vacuolata* was obtained in the previous sequencing project of its nucleomorph genome (see section 3).

2.3.2. Gene annotation

Open reading flames (ORFs) longer than 50 nucleotides were manually predicted as protein-coding genes using Artemis 13.2 (Rutherford et al. 2000). Functional annotation of the ORFs was carried out by BLASTX and BLASTP searches (Altschul et al. 1997). Ribosomal RNA (rRNA) genes were predicted using RNAmmer 1.2 (Lagesen et al. 2007) as well as BLASTN searches against the rRNAs of *Bigelowiella natans*. Transfer RNA (tRNA) genes were predicted using tRNAscan-SE 1.21 (Schattner et al. 2005). Group I and group II introns were predicted using the RNAweasel web server (Lang et al. 2007). The plastid genomes of *G. stellata*, *L. vacuolata*, and *P. glossopodia* are deposited in the GenBank/EMBL/DDBJ databases under the accession numbers, AP014947, AP014949, and AP014948 respectively.

2.3.3. Estimation of rearrangement scenarios among plastid genomes

Possible rearrangement scenarios between two plastid genomes were estimated using UniMoG 1.0 (Hilker et al. 2012) with the double cut and join operation, which considers gene inversions, translocations, fissions, and fusions (Yancopoulos et al. 2005). Plastid genomes were compared within each group of five chlorarachniophytes, six Ulvophyceae species, 34 Trebouxiophyceae species, seven Chlorophyceae species, and three Pedinophyceae species (Table 2.1). Each dataset consists of conserved plastid genes encoding proteins, rRNAs, and tRNAs, except for duplicated genes in regions of inverted repeats (IRs).

2.3.4. Phylogenic analyses

To infer phylogenetic trees, available plastid genomes in chlorophytes (Ulvophyceae, Trebouxiophyceae, Chlorophyceae, Pedinophyceae, and Prasinophyceae) were collected from the GenBank database (Table 2.1). Plastid gene sequences of *Tetraselmis subcordiformis* were identified from transcriptome data deposited in GenBank (accession number: GANN00000000.1). Plastid genomes of two species in Streptophyta, *Mesostigma viride* and *Chlorokybus atmophyticus*, were used as the outgroup. The final dataset was composed of 55 plastid-encoded proteins, excluding highly divergent genes (*e.g., rpl19*, *rps9*, *ycf1*, and *maturase-like*), collected from 70 taxa. Their amino acid sequences were aligned using MAFFT 7.164b with the L-INS-i option (Katoh and Toh 2008). Poorly aligned regions were removed manually using MEGA 6.0 (Tamura et al. 2013). The final concatenated sequences consisted of 9,876 amino acid positions. Maximum likelihood (ML) analyses were carried out using RAxML v.8.0.20 (Stamatakis 2014) with the LG+GAMMA+F model that was the best-fit model selected by IQ-TREE multicore version 1.3.2 (Nguyen et al. 2015) using the Bayesian information criterion (BIC). The best-scoring ML tree was determined in multiple searches using 20 distinct randomized maximum-parsimony trees, and statistical support (BP) was evaluated by 500 rapid bootstrap replicates. Bayesian analyses were performed using MrBayes v3.2.6-svn (Ronquist et al. 2012) and the LG+GAMMA+F model. The inference consisted of 1,000,000 generations with sampling every 1,000 generations, starting from a random starting tree and using four Metropolis-coupled Markov chain Monte Carlo (MCMCMC) simulations. Two separate runs were performed, and Bayesian posterior probabilities (BPP) were calculated from the majority rule consensus of the tree sampled after the initial 250 burn-in trees.

2.4. Results and Discussion

2.4.1. Architecture of plastid genomes in three chlorarachniophytes

I obtained the complete sequences of the circular plastid genomes for *Gymnochlora stellata* and *Partenskyella glossopodia*, and the almost complete plastid sequence of *Lotharella vacuolata*. The sizes of the plastid genomes in *G. stellata*, *P. glossopodia*, and *L. vacuolata* were 67,451, 72,620, and 71,557 bp, respectively (Fig. 2.1a, b, c). In the *L. vacuolata* genome, a short sequence gap between *psbE* and *atpI* could not be filled by my PCR-based sequencing analyses, and the corresponding region of *Bigelowiella natans* could not be sequenced either, presumably due to their secondary structures (Rogers et al. 2007). In the other two plastid genomes, corresponding regions consisted of two pairs of inverted repeats (IRs), likely constructing stem-loop structures (71 bp and 132 bp in *G. stellata* and 176 bp and 206 bp in *P. glossopodia*). All three plastid genomes had a canonical quadripartite structure consisting of two IRs, dividing the circular genome into a short and a large single-copy (SSC and LSC) region (Fig. 2.1a, b, c). Each of the IRs encoded three ribosomal RNA genes (*rns*, *rnl*, and *rrn5*), the same set of 4 tRNAs, and two or four proteins; the IRs of *G. stellata* and *L. vacuolata* consisted of *psbM* and *petB*, and the *P. glossopodia* IRs carried *psbM*, *petA*, *petB*, and *petD*, and a duplicated *psbM* is pseudogenized in *L. vacuolata* and *P. glossopodia* (Fig. 2.1a, b, c). The three genomes were predicted to encode the same set of 59 plastid proteins, 6 rRNAs, and almost the same number of tRNAs (29 for *L. vacuolata* and 31 for *G. stellata* and *P. glossopodia*), and some genes had duplicated copies in IRs (Fig. 2.1d, Table 2.2). Although *trnH* (GUG) and *trnG* (GCC) were not found in *L. vacuolata*, *trnH* (GUG) is expected to be present in the unsequenced region between *psbE* and *atpI* because it was detected in the corresponding region in the other plastid genomes.

The plastid genomes of five chlorarachniophytes lack several genes that conserved in core

chlorophytes, e.g., *petL*, *psaM*, *psbZ*, *rpl12*, *rpl32*, *rps9*, *infA*, *ccsA*, *cemA*, *chlB*, *chlL*, *chlN*, and *ftsH* (*ycf2*). Some of those homologs (*rpl12*, *rpl32*, *rps9*, *infA*, and *ftsH*) were found to be encoded in the nuclear genome of *B. natans* (Curtis et al. 2012), which suggested that a part of plastid genes were lost or transferred to the nuclear genome through the secondary endosymbiosis in chlorarachniophytes. Multiple gene losses for three subunits of a light-independent protochlorophyllide reductase (*chlB*, *chlL*, and *chlN*) were also reported in plastid genomes of some land plants (Wicke et al. 2011), implying that *chl* genes somehow tend to disappear from plastid genomes. The *ycf1* genes of chlorarachniophyte plastid genomes are homologous to those of chlorophytes, but they have no clear sequence homology to streptophyte *ycf1* genes that currently have been identified to encode a component of translocons at the inner envelop membrane of chloroplasts (Kikuchi et al. 2013). Although function of chlorarachniophyte *ycf1* genes remains unknown, it is interesting to note that the *ycf1* coding sequences show a large variation in both sequence and length (885 to 1,695 amino acids) within chlorarachniophytes, and such variation also has been observed among chlorophytes and streptophytes (de Vries et al. 2015).

2.4.2. High conservation of chlorarachniophyte plastid genomes

Genome organization was highly conserved among the plastid genomes of five chlorarachniophytes, including *B. natans* and *Lotharella oceanica*. In terms of gene content, almost all genes were shared among the five genomes (Table 2.2, 2.3). A remarkable difference was the presence/absence of an ORF (*maturase-like*) that encoded a protein that was roughly similar to bacterial reverse-transcriptase/maturase. No *maturase-like* genes were found in the *B. natans* plastid genome, whereas the other genomes had one. The five plastid genomes had slight differences in size, ranging from approximately 69 to 72 kbp. The size differences were mainly due to duplicated genes in IRs and variation in the size of the *ycf1* genes in SSC regions (Fig. 2.1d). The gene order was mostly identical among the five plastid genomes, except for

the duplicated genes and a couple of tRNA genes located near IR boundaries (Fig. 2.1d). The order of *petB* and *petD* was inverted between *P. glossopodia* and *L. vacuolata*, and coding strand switches were observed in *trnS* (UGA) among the five genomes (Fig. 2.1d).

There are extensive rearrangements in plastid genomes in general, even between closely related taxa (Brouard et al. 2011; Leliaert and Lopez-Bautista 2015; Turmel et al. 2015). I estimated rearrangement scenarios with gene translocations, inversions, fissions, and fusions between the plastid genomes of chlorarachniophytes using the double cut and join operation. The estimated number of rearrangement events was 2–8 among five plastid genomes of chlorarachniophytes. I also estimated rearrangement scenarios within chlorophyte groups, and determined that the number of rearrangement events were 22 to 83 in Chlorophyceae, 38 to 71 in Ulvophyceae, 1 to 75 in Trebouxiophyceae, and 4 to 53 in the group of Pedinophyceae and Chlorellales. Even for closely related species of chlorophytes, *Bryopsis hypnoides* and *Bryopsis plumosa*, and *Chlorella vulgaris* and *Chlorella variabilis*, the estimated numbers of rearrangements were 42 and 26, respectively. Thus, there were clearly fewer predicted rearrangement events in the chlorarachniophytes than in the chlorophyte groups. This may be explained by the higher gene density of chlorarachniophyte plastid genomes, which apparently increases the risk of gene disruption via frequent rearrangements; the coding regions represented 85.1% to 87.1% of the plastid genomes of chlorarachniophytes, and 19.5% to 81.8% for those of core chlorophytes. A similar pattern was found in cryptophytes with complex secondary plastids. Based on comparative analyses of plastid genomes in four cryptophytes, there are only a small number of inversion events, and the coding regions account for a high proportion of these genomes, i.e., between 80% and 87% (Kim et al. 2015) My findings revealed that chlorarachniophyte plastid genomes were highly conserved in size, gene content, and gene order among species. This suggests that the current architecture of chlorarachniophyte plastid

genomes evolved in a common ancestor, and changed very little during the subsequent diversification of chlorarachniophyte species.

2.4.3. Introns in chlorarachniophyte plastid genomes

Based on the *in silico* prediction by RNAweasel, some putative introns were found in the four plastid genomes of *G. stellata*, *L. vacuolata*, *L. oceanica*, and *P. glossopodia*. A self-splicing group I intron was predicted in the plastid *trnL* (UAA) of *L. oceanica* (212 nucleotides), *L. vacuolata* (227 nucleotides), and *P. glossopodia* (187 nucleotides) at identical positions within their tRNA anticodon loops (Fig. 2.2a). Group I introns have been reported in plastid *trnL* genes of diverse chlorophytes (Kuhsel et al. 1990; Besendahl et al. 2000), and the positions of *trnL* introns are conserved among chlorophytes and chlorarachniophytes (Fig. 2.2a). This suggests that the group I introns of chlorarachniophyte *trnL* genes were derived from a green algal endosymbiont, and were subsequently lost in the plastid genomes of *B. natans* and *G. stellata* during their reductive evolution. I also found that *ycf3* and/or *psbM* of four chlorarachniophytes carry group II introns, which were predicted by their secondary structures, whereas *G. stellata* lacks the *ycf3* intron and *B. natans* had no introns in either gene (Fig. 2.2b, c). Intron sizes ranging from 282 to 537 nucleotides were estimated based on alignments of *psbM* and *ycf3* sequences of chlorarachniophytes, including the intron-lacking species (Fig. 2.2b, c). The intron positions of *ycf3* and *psbM* were conserved among the chlorarachniophytes (Fig. 2.2b, c). In chlorophyte plastid gnomes, group II introns were detected in *ycf3* of *Picocystis salinarum* (Lemieux et al. 2014), *Bryopsis hypnoides* (Lü et al. 2011), and *Bryopsis plumose* (Leliaert and Lopez-Bautista 2015), and in *psbM* of *Oocystis solitaria* (Turmel et al. 2009) and *Schizomeris leibleinii* (Brouard et al. 2011). The positions of *ycf3* introns were not conserved between the chlorophytes and the chlorarachniophytes, whereas *psbM* intron positions

were identical between *O. solitaria* and the chlorarachniophytes. The *ycf3* introns of chlorarachniophytes appear to have been present in their common ancestor, and the two species *G. stellata* and *B. natans* lost the intron. Introns of chlorarachniophyte *psbM* might also be inherited from the common ancestor of chlorarachniophytes for the following reasons. First, introns identical to chlorarachniophyte *psbM* were not found in chlorophytes, other than *O. solitaria*. Second, the *O. solitaria* plastid was phylogenetically distinct from chlorarachniophyte plastids (see following section). Last, the *psbM* intron of *O. solitaria* included an ORF coding a putative reverse transcriptase (Turmel et al. 2009), whereas no ORFs were detected in the chlorarachniophyte *psbM* introns.

I found that the plastid genomes of four chlorarachniophytes, *G. stellata*, *L. vacuolata*, *L. oceanica*, and *P. glossopodia*, possess at least one group II intron, whereas the *B. natans* plastid genome had no introns. As described above, plastid genomes of the four chlorarachniophytes other than *B. natans* consisted of an ORF encoding a putative reverse transcriptase/intron maturase. Reverse transcriptase/intron maturase proteins are generally encoded within group II introns, and promote splicing by facilitating the formation of the catalytically active structure of the intron RNA (Lambowitz and Zimmerly 2004). This implies that the plastid genome of *B. natans* discarded group II introns as well as the splicing-related gene during the reductive evolution.

2.4.4. Origin of chlorarachniophyte plastids

The endosymbiotic origin of the chlorarachniophyte secondary plastids has previously been predicted based on molecular phylogenetic analyses (van de Peer et al. 1995; Ishida et al. 1997, 1999; Rogers et al. 2007; Takahashi et al. 2007; Tanifuji, Onodera, Brown et al. 2014). Phylogenetic analyses with particular gene types (*e.g.,* the plastid and nucleomorph SSU rRNA, and a nucleus-encoded plastid-targeted protein)

have resulted in different inferred trees indicating that chlorarachniophyte plastids are closely related to Trebouxiophyceae (van de Peer et al. 1995), Ulvophyceae (Ishida et al. 1997; 1999), and *Tetraselmis* (Takahashi et al. 2007). Phylogenetic trees inferred with approximately 50 plastid-encoded proteins and/or nucleomorph-encoded proteins suggest that chlorarachniophyte plastids are related to the so-called UTC group including Ulvophyceae, Trebouxiophyceae, and Chlorophyceae (Rogers et al. 2007; Tanifuji, Onodera, Brown et al. 2014), whereas the accurate position of the chlorarachniophyte plastids within the UTC group remains unclear owing to poor taxon sampling. To address this issue, I inferred phylogenetic trees using 55 plastid-encoded proteins from 63 operational taxonomic units (OTUs) within Chlorophytes, five chlorarachniophyte OTUs, and two OTUs in Streptophytes as the outgroup (Fig. 2.3). The trees suggested the robust monophyly of the core chlorophytes (the UTC group, Pedinophyceae, and Chlorodendrophyceae) with strong support (BP = 100, BPP = 1.00), and chlorarachniophyte OTUs were included in this clade (Fig. 2.3). Although the monophyly of each of the Chlorophyceae and the Pedinophyceae was strongly supported (BP = 100, BPP = 1.00), Trebouxiophyceae was divided into two well-supported clades, and one that consisted of the Chlorellales formed a sister clade with the Pedinophyceae (Fig. 2.3). The five chlorarachniophyte OTUs formed a robust monophyletic group (BP = 100, BPP = 1.00), and were predicted to be closely related to the Bryopsidales in Ulvophyceae with 84% bootstrap support (BPP = 1.00).

My phylogenetic analyses suggest that chlorarachniophyte plastids are derived from a green algal lineage closely related to Bryopsidales, which is composed of filamentous and branched multicellular marine algae. A previous phylogenetic study based on nucleus-encoded EF-Tu supported the close relationship between chlorarachniophytes and Bryopsidales (Ishida et al. 1997). Interestingly, the chlorarachniophyte *Cryptochlora perforans* was isolated from a sample of the filamentous green alga

Boodleopsis pusilla in Bryopsidales (Calderon-Saenz and Schnetter 1987), and amoeboid cells of *C. perforans* penetrated the algal filaments and engulfed part of their contents (Calderon-Saenz and Schnetter 1989). This implies that the secondary plastids of chlorarachniophytes might be acquired by the uptake of a filamentous green alga of Bryopsidales, similar to the feeding behavior of *C. perforans*. Furthermore, some sea slugs temporarily use the plastids of ingested green algae in Bryopsidales (Clark et al. 1990; de Vries et al. 2013) suggesting that the plastids of this algal group somehow tend to be integrated into diverse organisms.

2.4.5. Conclusion

In this study, I reported three plastid genomes of chlorarachniophytes. My comparative analyses indicated that the plastid genomes were highly conserved in size, gene content, and gene order among chlorarachniophyte species. The current architecture of chlorarachniophyte plastid genomes was present in a common ancestor and changed very little during the evolution of these species. The extreme conservation of the plastid genomes may be explained by their highly compacted genome structures, which is expected to increase the risk of gene disruption by frequent genomic rearrangements. Additionally, My phylogenetic analyses based on multiple plastid genes suggest that the endosymbiotic origin of chlorarachniophyte plastids is closely related to a green algal lineage of Bryopsidales.

3. Nucleomorph genome sequences of two chlorarachniophytes, *Amorphochlora amoebiformis* and *Lotharella vacuolata*

3.1. Abstract

Nucleomorph genomes are an interesting and suitable model to study the reductive evolution of endosymbiotically-derived genomes. To date, nucleomorph genomes have been sequenced in four cryptophyte species and two chlorarachniophyte species, including *Bigelowiella natans* (373 kbp) and *Lotharella oceanica* (612 kbp). In this study, I report complete nucleomorph genome sequences of two chlorarachniophytes, *Amorphochlora amoebiformis* and *Lotharella vacuolata*, to gain insight into the reductive evolution of nucleomorph genomes in the chlorarachniophytes. The nucleomorph genomes consist of three chromosomes totaling 374 and 432 kbp in size in *A. amoebiformis* and *L. vacuolata*, respectively. Comparative analyses among four chlorarachniophyte nucleomorph genomes revealed that these sequences share 171 function-predicted genes (86% of total 198 function-predicted nucleomorph genes), including the same set of genes encoding 17 plastid-associated proteins, and no evidence of a recent nucleomorph-to-nucleus gene transfer was found. This suggests that chlorarachniophyte nucleomorph genomes underwent most of their reductive evolution prior to the radiation of extent members of the group. However, there are slight variations in genome size, GC content, duplicated gene number, and subtelomeric regions among the four nucleomorph genomes, suggesting that the genomes might be undergoing changes that do not affect the core functions in each species.

3.2. Introduction

Chlorarachniophytes is of special interest because their complex plastids still harbor a relict nucleus of the endosymbiont, which has disappeared in most cases of secondary endosymbioses (Palmer 1997). The relic nucleus, so-called the nucleomorph, is localized in the periplastidal compartment, the space between the inner and outer pair of plastid membranes, which is the remnant cytoplasm of the endosymbiont (Hibberd and Norris 1984). Interestingly, nucleomorphs have also been found in cryptophyte plastids that originated from a red algal endosymbiont (Douglas et al. 1991; Douglas and Penny 1999). Therefore, two distant algal groups evolved highly reduced nucleomorph genomes via different routes from different starting points. Nucleomorph genomes offer an interesting opportunity to study the reductive evolution of endosymbiotically derived genomes.

To date, nucleomorph genomes have been sequenced in two chlorarachniophytes, *Bigelowiella natans* (Gilson et al. 2006) and *Lotharella oceanica* (Tanifuji, Onodera, Brown et al. 2014), and four cryptophytes, *Guillardia theta* (Douglas et al. 2001), *Hemiselmis andersenii* (Lane et al. 2007), *Cryptomonas paramecium* (Tanifuji et al. 2011), and *Chroomonas mesostigmatica* (Moore et al. 2012). Comparative investigations have revealed that the nucleomorph genomes share many conserved features even between chlorarachniophytes and cryptophytes (Archibald 2007; Archibald and Lane 2009). For instance, all of the nucleomorph genomes are composed of three small chromosomes, which generally possess ribosomal DNA (rDNA) operons in the subtelomeric regions at both ends. Recently, polyploidy of nucleomorph genomes has been reported in *B. natans* (diploid) and *G. theta* (tetraploid) (Hirakawa and Ishida 2014). The nucleomorph genomes (373 to 703 kbp in size) tightly encode only 284 – 610 proteins. Many genes encode housekeeping proteins for eukaryotic core functions (*e.g.,* translation, transcription, and protein folding) and others are nucleomorph-specific orphan genes (ORFans) that encode

hypothetical proteins showing no sequence similarity to any known proteins. Interestingly, conserved sets of plastid-associated proteins were found to be encoded by nucleomorph genomes. For example, 17 proteins are shared in the two chlorarachniophytes, and 31 proteins are shared in three cryptophytes, excluding the non-photosynthetic *C. paramecium*. However, nucleomorph-encoded genes are insufficient to maintain the nucleomorph function, and all nucleomorph genomes sequenced so far lack DNA polymerase genes. As a notable feature, a massive number of ultra-small introns has been found in chlorarachniophyte nucleomorph genes despite the extreme compaction of nucleomorph genomes. *B. natans* and *L. oceanica* nucleomorph genomes have 852 and 1,011 introns, respectively, ranging from 18 to 23 nucleotides (nt), with typical spliceosomal GT-AG boundaries (Gilson et al. 2006; Tanifuji, Onodera, Brown et al. 2014).

Pulsed-field gel electrophoresis analyses have revealed that the nucleomorph genomes vary in size (see general introduction). Several factors that contribute to the size variation of cryptophyte nucleomorph genomes have been identified (Lane et al. 2007; Tanifuji et al. 2011; Moore et al. 2012). The average length of protein-coding genes and the total number of genes are slightly different among the four cryptophyte nucleomorph genomes sequenced so far, and the most remarkable difference is found in the length of intergenic spacers. A comparison of chlorarachniophyte nucleomorph genomes between *B. natans* (373 kbp) and *L. oceanica* (610 kbp) revealed that the size variation is mainly caused by multiple duplicated genes (Tanifuji et al. 2014).

To gain further insight into nucleomorph genome evolutionary processes in chlorarachniophytes, I sequenced the nucleomorph genomes of two different species, *Amorphochlora amoebiformis* and *Lotharella vacuolata*. *L. vacuolata* is closely related to *L. oceanica*, and *A. amoebiformis* belongs to a phylogenetically distinct genus. The nucleomorph genomes of *A.*

amoebiformis and *L. oceanica* are 374 kbp and 432 kbp in size, respectively. My comparative analyses of four chlorarachniophyte nucleomorph genomes indicate that all sequences share 189 protein-coding genes, including the same set of genes encoding 17 plastid-associated proteins. The most remarkable difference among the four genomes was the existence of multiple duplicated regions across the nucleomorph genomes of *Lotharella* species, which mainly caused the variation in the size of nucleomorph genomes. My results suggest that chlorarachniophyte nucleomorph genomes have reached an end point in reductive evolution, whereas the increases in genome size occurred in some species individually.

3.3. Materials and Methods

3.3.1. Nucleomorph DNA extraction and sequencing

A. amoebiformis (CCMP2058) and *L. vacuolata* (CCMP240) were cultured at 20ºC under white light conditions (80–100 µmol photons·m^{-2}·s^{-2}) on a 12:12 hour light:dark cycle in ESM medium (Kasai et al. 2009). Cultivations and DNA extractions were performed by Dr. Hirakawa (University of Tsukuba). Cells were collected by general centrifugation from 2–3 week-old cultures. Nucleomorph DNA was separated by pulsed-field gel electrophoresis (PFGE), according to the conditions outlined by Silver et al. (2007). The separated nucleomorph DNA was purified from the gel slice by electroelution with dialysis membrane tubing (Moore et al. 2002). Shotgun libraries were generated and Sanger sequenced at the National Institution of Genetics in Japan. Additional sequencing of the *L. vacuolata* nucleomorph genome was carried out via the 454 GS Junior System (454 Life Sciences, a Roche Co., Branford, CT) with DNA extracted from isolated plastids. *L. vacuolata* cells were resuspended in 10 mL of modified isolation buffer (600 mM D-Sorbitol, 10 mM KCl, 5 mM EDTA, 1 mM $MgCl_2$, 1 mM $MnCl_2$, and 50 mM

HEPES-KOH, pH 7.6) (Hopkins et al. 2012) and disrupted by a Yeda press with 60 $kg \cdot cm^{-2}$ pressure at 4ºC. The resulting sample was loaded in a Percoll step gradient (20%, 30%, and 40% in gradient buffer containing 600 mM D-Sorbitol, 5 mM EDTA, and 50 mM HEPES-KOH, pH 7.6) and centrifuged at 3,300 *g* for 20 min at 4ºC. Plastids were enriched in interphase between 20% and 30%, and DNA was extracted from this fraction, using the CTAB method (Ishida et al. 1999).

3.3.2. Genome assembly and annotation

In total, 13,734 (10,174,889 bp) and 33,256 (24,900,190 bp) Sanger reads of *A. amoebiformis* and *L. vacuolata* were assembled using CodonCode Aligner (CodonCode Co., Centerville, MA), respectively. A total of 105,915 reads (44,084,934 bp) of *L. vacuolata* from the 454 GS Junior System were assembled using Newbler Assembler v. 2.5 (454 Life Sciences, a Roche Co.). The Sanger *L. vacuolata* contigs were reassembled with the 454 GS Junior contigs by using CodonCode Aligner. In total, 17 and 56 resulting nucleomorph contigs of *A. amoebiformis* and *L. vacuolata* were obtained, respectively, and gaps were filled by multiple polymerase chain reactions (PCR) with 14 and 53 sets of primers, respectively. The assembly of *A. amoebiformis* was collaborated with Mr. Shirato (The graduate university for advanced studies). To confirm the sequences of duplicated gene regions in the *L. vacuolata* nucleomorph genome, I amplified those regions by PCR and sequenced them with the ABI 3130 Genetic Analyzer (Applied Biosystems, Life Technologies, Carlsbad, CA.).

I manually identified open reading flames (ORFs) longer than 50 amino acids in the nucleomorph genomes using the Artemis Genome Browser 13.2.0 (Rutherford et al. 2000). Ultra-small introns were initially assumed to be 18–23 nt with a typical spliceosomal boundary (5'–GT and AG–3')

based on the chlorarachniophyte nucleomorph genes sequenced so far. To presume the function of protein-coding genes, I performed homology searches with BLASTx and BLASTp against sequence databases in NCBI (Altchul et al. 1997) with a cut-off e-value of 0.001. Based on the BLAST surveys, ORFs coding hypothetical proteins that have no similarity with any sequences in other organisms are defined as orphan genes (ORFans). rRNAs were identified using RNAmmer 1.2 (Lagesen et al. 2007) and BLASTn against rRNA sequences of *B. natans*. tRNAs and permuted tRNAs were predicted by tRNAscan-SE v. 2.1 (Schattner et al. 2005) and SPLITS (Sugahara et al. 2006), and the following parameters were applied: -c -p 0.55 -F -3 -h -3 (Soma et al. 2007) and -c -p 0.6 -F -1 (Maruyama et al. 2010). snRNAs were detected using fRNAdb with an option (word size = 7) (Kin et al. 2007). Simple repeat sequences in nucleomorph genomes were identified by the RepeatMasker (http://www.repeatmasker.org/). For comparative analyses, I also reconsidered ORFs of *B. natans* and *L. oceanica* and altered the number of protein-coding genes and introns (Table 3.1). Nucleomorph genome sequences of *A. amoebiformis* and *L. vacuolata* were deposited in GenBank/DDBJ/EMBL, and the accession numbers are AB996602-AB996604 and AB996599-AB996601, respectively.

3.3.3. Comparative analyses

In total, 188 of the shared proteins were used to calculate the average size of nucleomorph-encoded proteins. To determine the statistical significance of size differences in gene-coding and intergenic regions among nucleomorph genomes, I employed a one-way analysis of variance (ANOVA) with StatPlus:mac (http://www.analystsoft.com/en/). Homologous genes among four nucleomorph genomes of chlorarachniophytes were searched using MCScanX (Wang et al. 2012), based on their amino acid sequence homology (e-value < 0.001) (listed in Table 3.2). Positions of homologous genes were manually

compared among nucleomorph chromosomes, which are shown in line images created by MCScanX. Syntenic blocks consisting of at least four homologous genes in the same order were identified using the same definition as that used by Moore et al. (2012).

To examine the possibility of a recent gene transfer from the nucleomorph to the nucleus after the divergence of chlorarachniophyte species, I searched nuclear genes for genes missing from individual nucleomorph genomes. Seven and 18 genes were absent from the *B. natans* and *A. amoebiformis* nucleomorph genomes compared with the other three nucleomorph genomes, respectively. These genes were searched in the nuclear genome of *B. natans* (Curtis et al. 2012) or in the nuclear transcriptome of *A. amoebiformis* by BLASTx with a cutoff e-value of $1E^{-5}$. The *A. amoebiformis* RNA-seq transcriptome data were generated by the National Center for Genome Resources (NCGR) as a part of the Marine Microbial Eukaryotic Transcriptome Sequencing Project (Keeling et al. 2014) (the sample ID is MMETSP0042).

3.4. Results and discussion

3.4.1. Architectures of two nucleomorph genomes in chlorarachniophytes

The general characteristics of nucleomorph genomes in *A. amoebiformis* and *L. vacuolata* are summarized in Table 3.1. Both nucleomorph genomes are composed of three chromosomes totaling 374.0 kbp (131.9 kbp, 124.0 kbp, and 118.0 kbp) in *A. amoebiformis* and 431.9 kbp (166.2 kbp, 141.6 kbp, and 124.1 kbp) in *L. vacuolata*. The actual genome sizes were slightly different from the predicted sizes, according to the PFGE analyses, ~330 kbp and ~450 kbp (Silver et al. 2007). The GC content is 30.0% and 24.7% in *A. amoebiformis* and *L. vacuolata*, respectively. The number of total genes is predicted to be 340, including 300 protein-coding genes, 21 transfer RNAs (tRNAs), 18 rRNAs, and three non-coding

RNAs (ncRNAs) in the *A. amoebiformis* nucleomorph genome, and 359 genes, including 319 protein-coding genes, 19 tRNAs, 18 rRNAs, and three ncRNAs, in *L. vacuolata* (Table 3.1). The gene density is 0.91 and 0.83 genes/kbp in each nucleomorph genome (Table 3.1). All three chromosomes of *A. amoebiformis* and *L. vacuolata* carry identical sequences in the six subtelomeric regions comprised of an rDNA operon (SSU rDNA, 5.8S rDNA, and LSU rDNA) and the *dnaK* gene (Fig. 3.1, 3.2). When these genomes were compared with the nucleomorph genomes of two other chlorarachniophytes, *B. natans* and *L. oceanica*, all four nucleomorph genomes generally showed similar architectures; however, several notable variations were found.

The most remarkable difference is in the genome sizes. The *L. vacuolata* nucleomorph genome is approximately 50 kbp larger than those of *A. amoebiformis* and *B. natans* (Table 3.1). The primary factor leading to this size variation is the existence of multiple duplicated genes spreading in the *L. vacuolata* nucleomorph genome. Although no duplicated gene exists in the *A. amoebiformis* and *B. natans* nucleomorph genomes, excluding the subtelomeric repeats, *L. vacuolata* has 13 duplicated regions, including 35 complete and seven partial genes, totaling 37.8 kbp in size (Fig. 3.2; Table 3.3). The duplicated gene sequences were exactly identical, and 12 and 23 of those genes have intra- and inter-chromosomal copies, respectively. Interestingly, multiple gene duplications have also been found in the large nucleomorph genome of *L. oceanica* (~612 kbp) (Tanifuji, Onodera, Brown et al. 2014) that is closely related to *L. vacuolata*, and some of the duplicated genes (*e.g., rpl27*, *rpl12*, *secY*, *gsp2*, and *clpP-3*) are shared between these two *Lotharella* species (Fig. 3.2). This suggests that the genome size increase occurred via gene duplication before the divergence of *Lotharella* species. Furthermore, the *L. oceanica* nucleomorph genome carries a long subtelomeric sequence consisting of 45 ORFs between the SSU rDNA and *dnaK* of all six chromosome ends (total ~210 kbp) (Tanifuji, Onodera, Brown et al. 2014),

which is not seen in *L. vacuolata*. The *L. oceanica* nucleomorph genome appears to have acquired these subtelomeric sequences after the divergence of these two species. The length of intergenic regions also contributes to the size variation of the chlorarachniophyte nucleomorph genomes. The average length of intergenic regions is 134.5 bp (n = 356), 163.0 bp (n = 633), 112.5 (n = 329), and 86.6 (n = 339) in *L. vacuolata*, *L. oceanica*, *B. natans*, and *A. amoebiformis*, respectively, which are significantly different ($p < 0.001$, ANOVA) (Table 3.3). It has been reported that cryptophyte nucleomorph genomes also have size variations, and multiple duplicated genes and slightly longer intergenic spacers contribute to increases in the size of the relatively large nucleomorph genome in *C. mesostigmatica* (Moore et al. 2012). Interestingly, similar factors contribute to the size variation of nucleomorph genomes in both chlorarachniophytes and cryptophytes, despite their independent origins.

The telomeric and subtelomeric regions were found to have slight variations among the chlorarachniophyte nucleomorph genomes. The telomere sequence of *A. amoebiformis* is composed of [TCCTGGG] repeats, whereas other species typically carry [TCTAGGG]n. Moreover, the typical telomere sequence of chlorophytes is [TTTAGGG]n (Fulnečková et al. 2012), suggesting that an ancestral chlorarachniophyte had telomeric repeats of [TCTAGGG], and *A. amoebiformis* acquired the substitutions in the telomeric sequence. Subtelomeric regions consisted of an rDNA operon (SSU rDNA, 5.8S rDNA, and LSU rDNA), which is highly conserved in all nucleomorph genomes; however, there is variation in the gene order. The *A. amoebiformis*, *L. vacuolata*, and *L. oceanica* nucleomorph genomes carry the sequence of SSU-5.8S-LSU-telemere in this order, whereas the *B. natans* and *Chlorarachnion reptans* rDNA operons lie in the opposite direction (LSU-5.8S-SSU-telemere) (Silver et al. 2010). An inversion event of the rDNA operon is assumed to have occurred in a common ancestor shared between *B. natans* and *C. reptans,* based on phylogeny (Silver et al. 2010). Pseudogenes that partially encode the

3'-end of *myb1* were found to reside on each of the LSU rDNA downstream regions in *L. vacuolata* (Fig. 3.2), and similar *dnaK* pseudogenes exist in the LSU rDNA downstream regions in *B. natans* (Gilson et al. 2006). The functional *myb1* and *dnaK* genes are located near the subtelomeric region in one of the *L. vacuolata* and *B. natans* chromosomes, respectively. These pseudogenes would be unexpected products of inter-chromosomal recombination because nucleomorph subtelomeric regions are thought to have undergone frequent gene conversions via inter-chromosomal recombination to maintain the nearly identical rDNA sequences (Tanifuji, Onodera, Brown et al. 2014).

3.4.2. Gene content of nucleomorph genomes

Similar numbers of tRNAs, rRNAs, and ncRNAs are found in the chlorarachniophyte nucleomorph genomes sequenced thus far; however, these genomes have a remarkable variation in the number of protein coding genes (Table 3.1). The nucleomorph genomes of *A. amoebiformis*, *B. natans*, *L. vacuolata*, and *L. oceanica* have 300, 288, 319, and 596 protein-coding genes, respectively (Table 3.1). This variation is mainly caused by duplicated genes; thus, the number of non-redundant protein genes is almost identical among three species except for *L. oceanica* (295, 288, 294, and 338, respectively). Approximately 60% of nucleomorph-encoded proteins are annotated by homology with sequences of other organisms, and 40% are nucleomorph-specific hypothetical proteins, the so-called ORFans, that have no sequence similarity with any genes in databases.

The nucleomorph genomes of four chlorarachniophytes, *A. amoebiformis*, *B. natans*, *L. vacuolata*, and *L. oceanica* share 189 protein-coding genes, including 171 function-predicted genes and 18 ORFans (Fig. 3.3a, Table 3.2). Total 198 function-predicted genes have been annotated in chlorarachniophyte nucleomorph genomes, and 86% (171/198) of them are overlapped among four

genomes (Fig. 3.3a). In cryptophytes, 69% (216/311) function-predicted genes are shared among four nucleomorph genomes (Moore et al. 2012), suggesting that chlorarachniophyte nucleomorph genomes are less diverse than cryptophyte ones in term of gene content. Interestingly, all four chlorarachniophyte nucleomorphs possess the same set of genes encoding 17 plastid-associated proteins. The other annotated genes mainly encode housekeeping proteins for transcription, translation, DNA/RNA metabolism, and protein folding/degradation, and these genes would remain in the nucleomorph genomes for expression of the 17 plastid-associated proteins. When content of nucleomorph conserved core genes in four chlorarachniophytes was compared with those in four cryptophytes, 93 of 171 chlorarachniophyte core genes (54%) were overlapped with the cryptophyte core genes (Fig. 3.3b). Shared genes in the both groups were found in multiple categories of eukaryotic housekeeping functions (*e.g.,* translation, transcription, DNA/RNA metabolism, and protein fate and degradation), and half of shared genes (49/93 genes) were categorized as translation (Fig 3.3b). These data suggest that there are similar reductive pressures on nucleomorph retained genes in both chlorarachniophytes and cryptophytes.

Although annotated genes are mostly conserved among the four nucleomorph genomes of chlorarachniophytes, several genes have been lost independently in each species. For instance, *A. amoebiformis*, *B. natans*, *L. vacuolata*, and *L. oceanica* lacked 18, 7, 14, and 11 annotated genes, respectively. Tanifuji, Onodera, Brown et al. (2014) surveyed recent gene transfers of nucleomorph missing genes from nucleomorph to nuclear genomes by using the nuclear genome and transcriptome data of *B. natans* and *L. oceanica*; however, no evidence of gene transfer was found. I also searched for the lineage-specific nucleomorph missing genes in the nuclear genome of *B. natans* and the transcriptome data of *A. amoebiformis*, but did not detect any of them. The nucleomorph-to-nucleus gene transfer presumably did not occur after the divergence of chlorarachniophyte species. One possible explanation is

that the difference in GC content between nucleomorph and nuclear genomes (average 29% and 45%, respectively, in *B. natans*) would be a barrier for the expression of transferred genes and successful gene transfer. Overall, these data suggest that chlorarachniophyte nucleomorph genomes would have almost reached an end point in reductive evolution; however, they maintain some room for further reduction. Although conserved genes among different chlorarachniophyte nucleomorphs have been mostly annotated by homology searches, many hypothetical protein-coding genes (ORFans) are found to be lineage-specific genes. Even when closely related *Lotharella* species are compared, they have 59 lineage-specific ORFans (52.7% and 39.1% of the total ORFans in *L. vacuolata* and *L. oceanica*, respectively) (Fig. 3.3a). This suggests that loss and gain of many ORFans occurred independently after the divergence of chlorarachniophyte species. The function of ORFans is unclear, and it is hypothesized that ORFans may have taken over the function of lineage-specific missing genes such as those described above (Tanifuji, Onodera, Brown et al. 2014).

It has been reported that the nucleomorph genome of *B. natans* has two permuted $tRNA^{Ser}$ genes, *trnS* (AGA) and *trnS* (CGA), which have also been found in the nuclear genomes of several green algae, including *Ostreococcus* and *Micromonas* (Maruyama et al. 2010). I found those two permuted $tRNA^{Ser}$ genes in the *A. amoebiformis* nucleomorph genome, but no permuted tRNA was detected in *L. vacuolata* and *L. oceanica*. Thus, the green algal ancestor of chlorarachniophyte plastids is postulated to have permuted $tRNA^{Ser}$ genes; however, *L. vacuolata* and *L. oceanica* would have lost these genes after the divergence of chlorarachniophyte species.

The four nucleomorph genomes of chlorarachniophytes lack 5S rRNA gene, which is common in cryptophyte nucleomorph genomes. It has been known that yeast 5S rRNA recruits two ribosomal proteins, Rpl5 and Rpl11, to form 5S ribonucleoprotein particle, which is incorporated into

eukaryotic 60S preribosomes (Staley and Woolford 2009), and the C-terminal basic region of Rpl5 is important in the binding to 5S rRNA (Deshmukh et al. 1995). Although homologous genes for Rpl5 and Rpl11 were found in nucleomorph genomes of all four chlorarachniophytes, the C-terminal regions of Rpl5 were highly divergent compared to homologs of other organisms. Additionally, several genes for PPC-targeted 60S ribosome components are absent from both nucleomorph and nuclear genomes in the chlorarachniophyte *B. natans*, while almost complete set of PPC-targeted ribosome genes have been found in genomes of the cryptophyte *G. theta* (Curtis et al. 2012). Sequence divergence of the key ribosomal protein and partially lacking of 60S ribosome components might be related to the missing 5S rRNA gene in chlorarachniophyte nucleomorph genomes.

3.4.3. Ultra-small introns of nucleomorph genes

The chlorarachniophyte nucleomorph genomes are known to have numerous ultra-small spliceosomal introns ranging from 18 to 23 nucleotide (nt) (865 and 1,021 in *B. natans* and *L. oceanica*, respectively), whereas cryptophyte nucleomorphs have a small number of introns (0–24) (Lane et al. 2007; Moore et al. 2012). In this study, I predicted 793 and 1,028 of ultra-small introns in *A. amoebiformis* and *L. vacuolata*, respectively (Table 3.1). Most of these introns are 18–23 nt in size and possess a typical spliceosomal boundary (5'–GT and AG–3'), which is similar to that observed in other chlorarachniophytes (Fig. 3.4). A remarkable difference of introns among four chlorarachniophyte nucleomorph genomes is the size distribution (Fig. 3.4). The proportion of 19-nt introns is the highest in *B. natans* (70.3%) and *L. oceanica* (49.3%), whereas 18-nt and 20-nt introns are abundant in *A. amoebiformis* (42.1%) and *L. vacuolata* (35.8%), respectively. Total intron sizes are 14.9, 16.5, 20.0, and 20.9 kbp in *A. amoebiformis*, *B. natans*, *L. oceanica*, and *L. vacuolata*, respectively. A positive correlation between the nucleomorph genome size

and the intron length was assumed (Tanifuji, Onodera, Brown et al. 2014), but my data does not support this hypothesis. Although most of the introns are 18–23 nt in size, I found two exceptional introns that were 40 and 41 nt in size in the different positions of *A. amoebiformis* and *L. vacuolata prp43-2* genes, respectively (Fig. 3.5). These introns could be derived from the fusion of two ultra-small introns because a relict AG boundary exists at the center of the introns. *L. oceanica prp43-2* also has a 32-nt intron (Tanifuji, Onodera, Brown et al. 2014) at the same position as that of the 40-nt intron in *A. amoebiformis*, which would be the result of size reduction, following the intron fusion.

It has been reported that the positions of ultra-small introns are mostly conserved among homologous genes of *B. natans*, *L. oceanica*, and *Gymnochlora stellata* (Tanifuji, Onodera, Brown et al. 2014; Slamovits and Keeling 2009). I compared 290 introns within 55 conservative homologous genes among four chlorarachniophyte nucleomorph genomes. The positions of 38.3% introns were identical in the four species, and 86.6% introns were conserved in at least two species (Table 3.4). Many ultra-small introns were established in the current style before the divergence of chlorarachniophytes, and lineage-specific intron gain and loss seems infrequent.

In terms of splicing machinery for nucleomorph transcripts, I identified several spliceosomal protein genes and three small nuclear RNA (snRNA) genes (*U2*, *U5,* and *U6*) in the *A. amoebiformis* and *L. vacuolata* nucleomorph genomes. The *B. natans* nucleomorph genome encodes *U1*, *U2*, *U5*, and *U6* snRNA, and the *L. oceanica* nucleomorph genome has *U2*, *U5*, and *U6* snRNA genes. Tanifuji, Onodera, Brown et al. (2014) reported that the *L. oceanica* nucleomorph completely lacked all of the snRNAs, but my survey detected three snRNAs. The *U4* snRNA is absent from all nucleomorph genomes, which is consistent with the ultra-small size of introns, because *U4* snRNA is generally used to bring two remote splice sites together (Staley and Guthrie 1998). *U1* snRNA has a function to identify a 5' splice site, but

three nucleomorph genomes unexpectedly lack this gene. It is likely that I simply could not find several snRNA genes due to the low sequence similarity with canonical snRNAs. However, I could not exclude the possibility that snRNAs are transported from the nucleus to the nucleomorph across multiple plastid envelope membranes.

3.4.4. Rearrangement of nucleomorph genomes

Comparative analyses of nucleomorph genomes have revealed the existence of gene order conservation, so-called synteny, among distantly related species (Moore et al. 2012; Tanifuji, Onodera, Brown et al. 2014). Lane et al. (2007) suggested that non-homologous recombination events are likely to disrupt coding sequences in extremely reduced and compacted nucleomorph genomes. Therefore, recombination frequency is decreased, resulting in the retention of many syntenic blocks in nucleomorph genomes. In cryptophyte nucleomorphs, the average number of genes within a syntenic block consisting of four or more homologous genes, excluding ORFans, between two of *G. theta*, *H. andersenii*, *C. paramecium*, and *C. mesostigmatica*, is 6.7–19.4 (Moore et al. 2012). My comparative analysis of four chlorarachniophyte nucleomorphs indicated that syntenic blocks were composed of 6.2 (n = 17), 5.9 (n = 13), and 5.8 (n = 14) genes between *B. natans* and *A. amoebiformis*, *L. vacuolata*, and *L. oceanica*, 6.5 (n = 11) genes between *A. amoebiformis* and *L. vacuolata*, 6.9 (n = 11) genes between *A. amoebiformis* and *L. oceanica*, and 11.5 (n = 21) genes between *L. vacuolata* and *L oceanica,* on average when the same definition as that used by Moore et al. (2012) was applied (Fig. 3.6, 3.7, and Table 3.5). Nucleomorph genomes appear to be more scrambled in chlorarachniophytes than in cryptophytes. Even when two closely related *Lotharella* species (nucleus- and nucleomorph-encoded small subunit rDNAs are 95% and 99% identical, respectively, between *L. vacuolata* and *L. oceanica*) were compared, approximately 20% of the total genes (61/319

genes) were excluded from syntenic blocks in the *L. vacuolata* nucleomorph genome. Many of the syntenic blocks between *L. vacuolata* and *L. oceanica* are disrupted by duplicated gene regions (Fig. 3.2). These data suggest that genomic rearrangement of chlorarachniophyte nucleomorphs seems to be under progression at the species level, and recombination frequency would be higher in duplicated regions.

Nucleomorph syntenic blocks contain several ORFans, the so-called syntenic ORFans, encoding nucleomorph specific hypothetical proteins. It is difficult to predict the origins and functions of ORFans because of their high sequence diversity. However, syntenic ORFans have the potential for estimating homologous genes via comparison of positions and coding gene sizes among different nucleomorph genomes. In cryptophytes, a portion of ORFans are located at the same syntenic position as functional annotated genes found in the other nucleomorph genomes, suggesting that those ORFans originated by diversification of the annotated genes (Lane et al. 2007; Moore et al. 2012). My comparative analysis detected several syntenic ORFans in chlorarachniophyte nucleomorph genomes (Fig. 3.8). The 594-amino acid coding ORFan (*orf594*) of *A. amoebiformis* is located between *rad25* and *eif6*, and the *mcm-like* gene composed of 606 amino acids occupies the same syntenic position of *B. natans* (Fig. 3.8a). *A. amoebiformis* possesses the 486-amino acid ORFan (*orf486*) next to *rpl2*, while the other three chlorarachniophytes have *tcpG* genes (480–506 amino acids) at the same position (Fig. 3.8b). The *L. vacuolata orf776* between *mce* and *U5 snRNP* (116 kDa) is located in the same syntenic position of *L. oceanica tbl3* (780 amino acids) (Fig. 3.8c). These data suggest that nucleomorph-encoded ORFans are generated by sequence divarication of functional annotated protein genes in both chlorarachniophytes and cryptophytes. However, it remains unclear whether ORFans still have the original function of homologous genes or not.

I also found disrupted syntenic ORFans in the *L. oceanica* nucleomorph genome. Two ORFans (*orf76* and *orf269*) of *L. oceanica* occupy the same syntenic position of *L. vacuolata nop56* (414 amino acids) between *eif2G* and *cwc22* (Fig. 3.8d). Interestingly, these two ORFans have similarity in the 5' and 3' partial coding sequence of *nop56*, and the coding region is divided into two ORFans by a stop codon in the first exon. Similar phenomena have been reported in cryptophyte nucleomorph genomes. For instance, *sf3b3-like* of *C. mesostigmatica* corresponds to three syntenic ORFans of *C. paramecium*. These data imply another generating system of nucleomorph ORFans by the splitting of gene coding regions. In fact, the average protein size of ORFans (237.2 amino acids) was significantly smaller than that of functional annotated genes (348.6 amino acids, $p < 0.01$, *t*-test) in the chlorarachniophytes.

3.4.5. Conclusion

Nucleomorph genomes are highly reduced through the achievement of secondary endosymbiosis. In this study, I sequenced two complete nucleomorph genomes of chlorarachniophytes, *A. amoebiformis* and *L. vacuolata*. Our comparative analyses of nucleomorph genomes in four chlorarachniophyte species proposed that most of the functionally annotated genes were shared between them, and a small number of core gene losses was observed in each nucleomorph genome individually. This suggests that reductive evolution of the nucleomorph genomes in chlorarachniophytes has mostly reached an endpoint, and that the genome reduction of chlorarachniophyte nucleomorphs has progressed more than that of cryptophytes that are undergoing gene losses associated with core eukaryotic housekeeping functions (Moore et al. 2012). My data also revealed that size increases of nucleomorph genomes occurred via multiple gene duplications in *Lotharella* species.

4. Overexpression of molecular chaperone genes in nucleomorph genomes

4.1. Abstract

Nucleomorphs of chlorarachniophytes and cryptophytes contain a greatly reduced genome that possesses only several hundred genes with high evolutionary rates. I examined the relative transcription levels of the genes of all proteins encoded by the nucleomorph genomes of two chlorarachniophytes and three cryptophytes using an RNA-seq transcriptomic approach. The genes of two heat-shock proteins, Hsp70 and Hsp90, were highly expressed under normal conditions. It has been shown that molecular chaperone overexpression allows an accumulation of genetic mutations in bacteria. My results suggest that overexpression of heat-shock proteins in nucleomorph genomes may play a role in buffering the mutational destabilization of proteins, which might allow the high evolutionary rates of nucleomorph-encoded proteins.

Introduction

Nucleomorph genomes are characterized by a highly compact structure with very short intergenic regions (Keeling and Slamovits 2005; Williams et al. 2005). The impact of this gene-dense structure on the regulation of nucleomorph gene expression is poorly understood. Recently, Tanifuji, Onodera, Brown et al. (2014) reported the transcription patterns of nucleomorph genes using genome mapping analyses with RNA-seq datasets in the chlorarachniophyte *B. natans* and three cryptophytes. In all four species, RNA transcripts covered over 99% of the entire nucleomorph genomes including intergenic regions, and global transcript levels were equal or higher for nucleomorph genes than for nuclear homologs. I have studied

the nucleomorph mRNA expression in these four species as well as another chlorarachniophyte *Amorphochlora amoebiformis*, and discovered that nucleomorph genes for two heat-shock proteins (NmHsp70 and NmHsp90) are transcribed to a remarkable degree in four of these species. Our results imply a relationship between higher levels of Hsp transcripts and higher evolutionary rates of nucleomorph genes.

4.2. Materials and Methods

4.2.1. Genome mapping analysis

In genome mapping analysis with transcriptome data, RNA-seq reads (50 or 100 bases in each length) were collected by Illumina sequencing for two chlorarachniophytes and three cryptophytes. The quality of all reads was checked by FastQC v. 0.10.1 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and subsequently the first 14 nucleotides of all reads and the last 50 nucleotides of 100-bases reads were trimmed by Fastx-Trimmer program in FASTX-Toolkit package (http://hannonlab.cshl.edu/fastx_toolkit/index.html). The resulting reads of 36 bases were mapped onto the concatenated nucleomorph chromosomes by BWA-0.5.9 program (Li and Durbin 2009). An option (-e=21, maximum gap length) was apply to chlorarachniophytes to allow most of the ultra-short introns. A small number of introns were manually removed from the nucleomorph sequences of cryptophytes. Total 1,264,727, 6,542,159, 4,192,729, 2,042,331, and 2,382,754 reads were successfully mapped onto gene coding regions for *Amorphochlora amoebiformis*, *Bigelowiella natans*, *Chroomonas mesostigmatica*, *Guillardia theta*, and *Cryptomonas paramecium*, respectively. The number of reads per each gene was counted following an annotation file by BEDTools-2.17.0 (Quinlan and Hall 2010). The counted data was normalized as Reads Per Kilobase of exon per Million mapped reads (RPKM) (Mortazavi et al. 2008).

4.2.2. Real time quantitative PCR

This real time quantitative PCR was performed by Dr. Hirakawa (university of Tsukuba). *Amorphochlora amoebiformis* CCMP2058 cells were maintained at 20ºC under white illumination (80 μmol photons·m–2·s–1) on 12: 12 h light: dark cycle in ESM medium. Total RNA was purified by Trizol Reagent (Invitrogen) from the 1.2×107 cells of *A. amoebiformis* culture in mid-light and mid-dark phases. The amount of RNA was quantified by NanoDrop-1000 (Thermo Scientific). The cDNA was synthesized by ReverTra Ace qPCR RT Kit (Toyobo) with 3.0 μg total RNA in total 40 μl reaction mix. Primers for quantification of the nucleus- and nucleomorph-encoded heat-shock protein genes (hsp70, hsp90, nmhsp70, and nmhsp90), the nucleomorph-encoded plastid translocon gene (nmtoc75), and the small subunit ribosomal RNA (18S rRNA) were designed using the Primer3Plus online software (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/primer3plus.cgi): nmhsp70-F: 5'–TGTTGAAGTTGTGGTTACTGTTCC–3', nmhsp70-R: 5'–CAATACAAGCAGCCGTAGGTTC–3', nmhsp90-F: 5'–GCAGCAGTTAAAGGAATTTGAAGG–3', nmhsp90-R: 5'–CGCATAGCAGTGAATAAGAAGCAAC–3', hsp70-F: 5'–CGTCACCTTCGATCTGGATG–3', hsp70-R: 5'–TCCCTTGTCGTTTGTGATGG–3', hsp90-F: 5'–TCATTGGTTTCCCCATCTCTC–3', hsp90-R: 5'–TCGTCATCGTCATCAACATCC–3', nmtoc75-F: 5'–TGCTTTATAACGACCACGAAGAAG–3', nmtoc75-R: 5'–GCACCACGGATGGGTTTTAC–3', 18SrDNA-F: 5'–ATAGAGGGACTATCGGACCCAAC–3', 18SrDNA: 5'–CGAGCCAAGGACAAAACAAAC–3'. Each fragment was amplified and cloned into a pGEM-T easy vector (Promega), serial dilutions of which were used to create the standard curve. RT-qPCR was carried out by the Thermal Cycler Dice Real Time System II (Takara) under following condition: 0.5μl of cDNA,

0.4 μM of each primer, 12.5 μl of SYBR Premix Ex Taq II (Takara), and DNase/RNase-free water up to 25μl. The cycling condition comprised 3 min of denaturation at 95ºC followed by 40 cycles of 10 sec at 95ºC, 30 sec at 60ºC, and a melting curve programme. Relative expressions were calculated by the Ct values (2nd derivative maximum) and the standard curves of serial dilutions, and normalized to 18S rRNA between light and dark samples.

4.2.3. Calculation of evolutionary rates of homologous protein sequences.

To calculate evolutionary rates (substitution rates) of nucleus- and nucleomorph-encoded proteins, I used 26 ribosomal protein sequences (rps2, 3, 4, 5, 6, 7, 10, 11, 12, 13, 14, 15, 16, 17, 26, and rpl3, 5, 8, 10, 11, 15, 17, 19, 23, 27, and rpl32) of chlorarachniophytes, *Amorphochlora amoebiformis*, *Bigelowiella natans*, and *Lotherlla globosa/vacuolata*, and 28 ribosomal proteins (rps5, 6, 8, 9, 11, 13, 14, 19, 20, 23, 25, 26, and rpl1, 3, 5, 8, 12, 14, 15, 17, 19, 21, 26, 30, 31, 32, 34, and rpl36) of cryptophytes, *Guillardia theta*, *Cryptomonas paramecium*, and *Chroomonas mesostigmatica*. These sequences were collected from the databases (GenBank and JGI) and the NCGR transcriptome data (sample ID are MMETSP0038, 0041, 0042, and 0047; Keeling et al. 2014), and part of sequences came from nucleomorph genome data of *A. amoebiformis* and *L. vacuolata*, which were sequenced in this study. Homologous sequences of each protein were automatically aligned with the L-INS-I method of the MAFFT package (Katoh, Toh 2008), and all gaps were deleted. The amino acid substitution rates among three species were calculated by the codeml with an empirical model in PAML 4.6 package (Yang 2007), using 4,971 and 4,992 amino acid (AA) sites of nucleus-encoded ribosomal proteins, and 4,065 and 4,236 AA sites of nucleomorph-encoded ribosomal proteins for chlorarachniophytes and cryptophytes, respectively.

To calculate amino acid substitution rates of highly and lowly expressed genes (HEG and LEG), I picked up ten most highly/lowly expressed genes from chlorarachniophytes (*A. amoebiformis* and *B. natans*) and cryptophytes (*C. mesostigmatica* and *G. theta*) based on their RPKM values; unique open reading frames (ORFs) were excluded from the selection. The chlorarachniophyte HEG are clpC, gsp2, hsp70, hsp90, rnabp1, rpl23, rps27, sufB, u2AF, and ycf16, and LEG are cdc48-like, mce, mcm2, nbp1, prp43-2, rhel1, rpoF, sbp1, tcpB, and tcpT. The cryptophyte HEG are cbbX, cdc48, clpP1, clpP2, cpn60, gsp2, hsp70, hsp90, rps15, and tha4, and LEG are cdc28, cycB, ggt, mcm2, pab2, rad3, sbp1, smc1, trf, and u5snRNP. Homologous sequences of each protein were automatically aligned with the L-INS-I method, and all gaps were deleted. The amino acid substitution rates between homologs from two species were calculated by the codeml in PAML 4.6 package.

4.3. Results and Discussion

4.3.1. High expression levels of nucleomorph chaperone genes

In order to calculate the relative mRNA expression levels of nucleomorph-encoded proteins, I analyzed transcriptome data from two chlorarachniophytes, *Amorphochlora amoebiformis* CCMP2058 and *Bigelowiella natans* CCMP2755, and three cryptophytes, *Guillardia theta* CCMP2712, *Cryptomonas paramecium* CCAP977/2A, and *Chroomonas mesostigmatica* CCMP1168, that were generated by the NCGR Marine Microbial Eukaryotic Transcriptome Sequencing Project: sample ID are MMETSP0038, 0042, 0045, 0046, and 0047 (Keeling et al. 2014). Several million RNA-seq reads were mapped onto the nucleomorph genome sequences, and the depth of coverage in each protein-coding region was measured. Relative transcript levels were estimated by calculating RPKM values (Reads Per Kilobase of exon per Million mapped reads) (Mortazavi et al. 2008). In two chlorarachniophytes, the average transcript levels

among all nucleomorph genes were found to be 3,032 and 3,691 in *A. amoebiformis* and *B. natans*, respectively. Interestingly, the transcript levels of two heat-shock proteins, NmHsp70 and NmHsp90, were significantly higher, at 54,375 and 45,761 in *A. amoebiformis*, and at 35,738 and 38,000 in *B. natans*, respectively (Fig. 4.1a, b and Table 4.1, 4.2); none of the genes for co-chaperones of NmHsp70 and NmHsp90 have been detected in the nucleomorph genomes. The mRNA derived from these two genes are predicted to comprise 11 and 7% of the total nucleomorph-derived mRNA in *A. amoebiformis* and *B. natans*, respectively. To compare the mRNA expression levels of nucleus- and nucleomorph-encoded heat-shock proteins, real time quantitative PCR (qPCR) with gene specific primers for *A. amoebiformis* was carried out by Dr. Hirakawa (university of Tsukuba). The NmHsp70 and NmHsp90 genes were highly transcribed in comparison with the nucleomorph-encoded plastid translocon NmToc75 showing the average transcript level (Fig. 4.1c), which was in agreement with the transcriptome data. The transcript levels of the NmHsp genes were equal to or greater than those of nucleus-encoded cytoplasmic homologs (Fig. 4.1c). Since translation rates of nuclear and nucleomorph genes are unknown, I was not able to compare the amount of proteins based on mRNA transcript levels. However, it would be interesting to see whether levels of nucleomorph-encoded heat-shock proteins are as abundant as their transcripts, because the volume of the periplastidal compartment (PPC) in which the nucleomorph resides is much smaller than the cytoplasm. Although the transcript levels of cytoplasmic Hsp70 and Hsp90 changed dramatically under the light and dark conditions, the NmHsp70 and NmHsp90 genes were constantly highly expressed in the PPC (Fig. 4.1c). Light induction of Hsp transcripts has been reported in photosynthetic organisms, the cyanobacterium *Synechocystis* (Hihara et al. 2001), the green alga *Chlamydomonas* (von Gromoff et al. 1989), and land plants (Li et al. 2000; Rossel et al. 2002). Hence, it

is highly possible that the transcript levels of nucleus-encoded Hsps in the chlorarachniophyte would also increase in response to light.

High levels of Hsp gene transcription were also observed in two cryptophyte nucleomorph genomes. The transcript levels of NmHsp70 and NmHsp90 were found to be 57,544 and 43,706 in *C. mesostigmatica*, 58,270 and 31,472 in *G. theta*, and 25,571 and 15,252 in *C. paramecium*, while the average transcript levels among all nucleomorph genes of these species were 2,021, 2,333, and 2,768, respectively (Fig. 4.2a, b, c Table 4.3, 4.4, 4.5). The mRNA transcripts from NmHsp genes were predicted to comprise 10, 8, and 3% of the total nucleomorph-derived mRNA in *C. mesostigmatica*, *G. theta*, and *C. paramecium*, respectively. The NmHsp70 and NmHsp90 were the most highly expressed genes in the nucleomorph genomes of *C. mesostigmatica* and two chlorarachniophytes. However, in *G. theta* and *C. paramecium*, the mRNA levels of a few genes for plastid stroma proteins (CbbX and ClpP1) and nuclear proteins (Gsp2 and H2B) were greater than those of NmHsp genes (Fig. 4.2b, c Table 4.4, 4.5).

4.3.2. Rapid evolutionary rates in nucleomorphs

Molecular chaperones, including heat-shock proteins, are essential proteins that play an important role in the folding, disaggregation, and intracellular transport of proteins in cells (Saibil 2013). The PPC is predicted to contain fewer proteins than the cytoplasm, since the volume of PPC and the number of nucleomorph genes are much smaller than that of cytoplasm/nuclear genes. This raises the question: Why are the Hsp70 and Hsp90 molecular chaperones so highly expressed in the nucleomorph genomes? One intriguing possibility is that it is related to the evolutionary rate of nucleomorph genomes. In *Escherichia coli*, it has been reported that the overexpression of GroEL/GroES chaperonins mask the deleterious

effects of mutated proteins, which allows for an increase in the number of accumulating mutations (Maisnier-Patin et al. 2005; Tokuriki and Tawfik 2009). In eukaryotic cells, Hsp90 is thought to act as a buffer for genetic variations (Taipale et al. 2010). Furthermore, members of the bacterial genus *Buchnera*, which are endosymbionts of aphids, have greatly reduced genomes, which evolve faster than their homologs in closely related free-living bacteria (Moran 1996). Most of the heat-shock proteins in *Buchnera aphidicola* are overexpressed even under non-stress conditions (Wilcox et al. 2003), implying that these chaperones mask the destabilizing effects of mutations and allow the genetic variation in the endosymbiotic genome (McCutcheon and Moran 2011). A similar situation is seen in the nucleomorph genomes of chlorarachniophytes and cryptophytes. To estimate the sequence divergence of nucleomorph genomes, I used 26 and 28 homologous sequences of nucleus- and nucleomorph-encoded ribosomal proteins in three chlorarachniophytes (*Amorphochlora amoebiformis*, *Bigelowiella natans*, and *Lotharella globosa*/*vacuolata*) and three cryptophytes (*Chroomonas mesostigmatica*, *Guillardia theta*, and *Cryptomonas paramecium*), respectively. Amino acid substitution rates of the nucleomorph-encoded proteins were calculated to be between 0.5611 and 0.5665, and 0.4292 and 0.4882 per site among the three species of chlorarachniophytes and cryptophytes, respectively, while the substitution rates in nucleus-encoded proteins were between 0.1664 and 0.1766, and 0.1997 and 0.2055 (Fig. 4.1d, 4.2d). The nucleomorph-encoded proteins were clearly evolving faster than nuclear ones in both chlorarachniophytes and cryptophytes, and a high evolutionary rate of nucleomorph genes has been reported in *B. natans*, previously (Patron et al. 2006). These data provide a possibility that the higher expression of molecular chaperones might compensate for the quick evolution of nucleomorph genes by buffering the deleterious effects of mutated proteins in the PPC. In addition, the quick evolution could also be attributed to the loss

of genes involved in nucleomorph DNA replication and repair (Curtis et al. 2012), like bacterial endosymbionts (McCutcheon and Moran 2011).

In my analysis, the NmHsp70 and NmHsp90 transcript levels of *C. mesostigmatica* and two chlorarachniophytes were significantly higher than the other transcripts, while those of *G. theta* and *C. paramecium* were inconspicuous (Fig. 4.1, 4.2). Interestingly, although the nucleomorph genomes of these two cryptophytes possess genes that encode proteasome subunits (Douglas et al. 2001; Tanifuji et al. 2011, Stork et al. 2012), *C. mesostigmatica* and two chlorarachniophytes completely lack these genes (Gilson et al. 2006; Moore et al. 2012). This implies that *C. mesostigmatica* and two chlorarachniophytes are more susceptible to protein misfolding because of the lack of a proteasomal degradation system in the PPC. Therefore, these three species might express the NmHsp genes at higher levels to mitigate these risks.

4.3.3. GC percentages of highly expressed genes in nucleomorphs

I also found an interesting feature of base composition in nucleomorph genes, which is worth mentioning. It has been shown that highly expressed genes are less divergent and have an amino acid compositional bias leading to a higher GC content in AT-rich genomes of bacterial endosymbionts (Schaber et al. 2005) and the eukaryotic parasite, *Plasmodium* (Chanda et al. 2005). Nucleomorph genomes are also extremely AT-rich genomes. I calculated the GC content of each gene in five nucleomorph genomes of chlorarachniophytes and cryptophytes, and found a positive correlation between the GC content and the relative transcription level in all five genomes; the GC content tends to be increased depending on the transcription level (Fig. 4.3a - e). Furthermore, I identified the 10 most highly and lowly expressed genes (HEG and LEG) from two species of chlorarachniophytes/cryptophytes, and calculated their amino acid

substitution rates between homologous sequences. Average substitution rates of the HEG/LEG between *A. amoebiformis* and *B. natans*, and *C. mesostigmatica* and *G. theta*, were found to be 0.39/0.68 and 0.23/0.66, respectively (Fig. 4.3f), suggesting that the sequence conservation of HEG was clearly higher than that of LEG. These results suggest that the highly transcribed genes with higher GC content are more conserved in nucleomorph genomes. Therefore, there would be a similar selective pressure in highly expressed genes for nucleomorph genomes as well as other endosymbiotic/parasitic genomes.

4.3.4. Conclusion

Nucleomorph genomes have been specialized in function and structure in the secondary endosymbiosis. I performed RNA-seq of nucleomorphs, and calculated expression levels, evolution rates, and GC% of nucleomorph-encoded protein genes. The genes of two heat-shock proteins, Hsp70 and Hsp90, were highly expressed under normal conditions. The evolutionary rates of the nucleomorph-encoded proteins were faster than those of nuclear-encoded proteins. These suggest that overexpression of heat-shock proteins in nucleomorph genomes, which are highly conservative in base composition, may play a role in buffering the mutational destabilization of proteins, which might allow the high evolutionary rates of nucleomorph-encoded proteins.

5. General discussion

5.1. Reductive evolution of the endosymbiotically-derived genomes during secondary endosymbiosis

In this study, I reveal that the plastid and nucleomorph genomes of chlorarachniophytes have been highly reduced before the radiation of most chlorarachniophyte lineages (Chapter 2 and 3). Similar aspects are predicted for the plastid and nucleomorph genomes of the cryptophytes. These findings indicate that the genomes of endosymbionts are highly minimized during the early phase of secondary endosymbiosis (Fig. 6.1). Genome reduction of endosymbionts is generally accompanied by EGT to the host nuclei, and with the construction of the protein transport system in the plastid. It is important to understand how the EGT and protein transport system were rapidly constructed in the early phase of secondary endosymbiosis. During the primary endosymbiosis, genes maintaining the endosymbiotic relationships (*e.g.*, sugar metabolic proteins) are thought to have been transferred from a chlamydial bacterium to the host nucleus by LGT during primary endosymbiosis (Ball et al. 2013). In secondary endosymbiosis, the LGT before engulfing secondary endosymbionts might be important for the establishment of plastids. A non-photosynthetic phagocytic euglenophyte, *Peranema trichophorum*, which belongs to a different lineage than photosynthetic euglenophytes, acquired several genes by LGT from "red lineage" organisms (Maruyama et al. 2011). Euglenophytes might have ingested the "red lineage" organism before engulfing a green alga, and they acquired the "red lineage" genes by LGT. Although Archibald et al. (2003) discussed that the phylogenetically mosaic genes of the chlorarachniophyte *B. natans* were derived from the prey after secondary endosymbiosis, it can be interpreted that *B. natans* had acquired the phylogenetically mosaic genes before endosymbiosis. To understand the timing of this EGT/LGT, genome sequencing of a non-photosynthetic chlorarachniophyte, like *Minorisa minuta*, is needed.

B. natans has no ongoing EGT from the nucleomorph or plastid to the nucleus (Curtis et al. 2012; Chapter 3). One explanation for the reduced ongoing EGT is the "limited transfer window" hypothesis (Curtis et al. 2012). It posits that species with a single plastid per cell has a lower possibility to transfer organelle genes to the host cell by EGT than with multiple organelles in a cell, because the species with a single organelle cannot survive the lysis of the plastid (Barbrook et al. 2006; Smith et al. 2011). *B. natans* possesses a single plastid and nucleomorph per cell; however, another chlorarachniophyte has multiple plastids and nucleomorphs per cell. In this study, I reveal that there is no EGT in *A. ameobiformis* after the radiation of the main lineage of chlorarachniophytes (Chapter 3). Because *A. amoebiformis* possesses multiple plastids and nucleomorphs per cell (Ishida et al. 2000), the absence of EGT in *A. amoebiformis* cannot be explained by the "limited transfer window" hypothesis. My analysis is based on transcriptome data; therefore, the nuclear genome of *A. amoebiformis* might have un-transcribed DNA fragments of the plastids and nucleomorphs. To confirm such presence/absence of recent EGT, genome sequencing of another chlorarachniophyte with multiple plastids is needed.

For the reductive evolution of the plastid and nucleomorph genomes of chlorarachniophytes, frequency of genome rearrangements is one of the main differences (Chapter 2 and 3). The plastid genomes have few rearrangements because of their compact structure, *i.e.,* short intergenic regions. In contrast, the nucleomorph genomes have frequent genome recombination despite their short intergenic regions, similar to the plastid genomes. This difference might be explained by the gene content. The plastid genomes possess only essential genes for photosynthesis; however, the nucleomorph genomes have many ORFans, which are not shared by the chlorarachniophytes, as well as well-conserved function-predicted genes. Although the importance of the ORFans remains unclear, it is more likely to perform recombination in the ORFans than the conserved function-predicted genes. In cryptophytes, more

frequent recombination is also detected in the nucleomorph genomes, with more ORFans than the plastid genomes, which are composed of few genes with unknown function.

5.2. Regulation of endosymbiont gene expression during the secondary endosymbiosis

I reveal that rapid evolutionary rates of the nucleomorph genomes might be allowed through high an expression level of the genes for molecular chaperones (Chapter 4; Fig. 6.1). The overexpression of chaperones might be one of driving forces of nucleomorph genome reduction, as well as loss of the DNA repair system for the nucleomorphs (Chapter 5; Gilson et al. 2006; Curtis et al. 2012). The nucleomorph genome reduction, which minimized intergenic regions, introns, and protein sizes, causes loss of gene expression patterns (Chapter 5), and overlaps transcripts with multiple genes (Williams et al. 2005). In chlorarachniophytes, genome reduction also results in low splicing efficiency (Gilson et al. 2006; Slamovits and Keeling 2009; Tanifuji, Onodera, Moore et al. 2014), which is considered to contribute toward higher expression levels of nucleomorph-encoded genes as opposed to nuclear-encoded ones (Tanifuji, Onodera, Moore et al. 2014).

In this study, I reveal that the host nucleus transcriptionally regulates some components of the nucleomorph- or plastid-targeted genes encoded in the nuclear genome throughout the cell cycle (Chapter 5; Fig. 6.1). The host cell had probably acquired the transcriptional regulation during EGT to the host genome in the early stage of endosymbiosis because gene expression without regulation might endanger the cell. Therefore, the EGT might have been intermediated with DNA, including promoters and transcriptional regulatory regions, but not RNA. During the secondary endosymbiotic process, expression patterns of the photosynthesis-related protein genes have been changed (chapter 5). This is beneficial,

because expression patterns of the oscillator genes, which periodically regulate gene expressions as transcription factors, have been changed. In land plants, several oscillator genes (*e.g.*, *LHY* and *CCA1*) control periodical gene expression (Hsu and Harmer 2014); however, I could not clearly detect such oscillator genes in the chlorarachniophytes in the present study. To identify the oscillator genes in chlorarachniophytes, novel methods to construct knockdown or knockout cells are required.

5.3. Nucleomorph fate

Previous studies suggest that the nucleomorph genome of chlorarachniophytes is decreasing, and that it is in an intermediate stage of its reductive evolution (Gilson et al. 2006). However, the present study shows that the nucleomorph genome of chlorarachniophytes have almost reached an endpoint of reductive evolution, and the nucleomorph genomes of some lineages have experienced a secondary expansion due to gene duplication (Chapter 3). This conclusion is supported by evidence that the expression of nucleomorph-targeted protein genes is intricately regulated by the host cell, and that expression patterns of the near-complete genes of the nucleomorph genome are missing (Chapter 5).

6. Acknowledgements

I would like to thank:

Dr. Ken-ichiro Ishida (university of Tsukuba) my supervisor for providing me insightful comments about this research.

Dr. Yoshihisa Hirakawa (university of Tsukuba) for providing me many invaluable ideas on this study, allowing me to use his data in this study and kindly editing this thesis.

Dr. Mamoru Sugita (Nagoya university) for sequencing the plastid genome of *G. stellata*, and giving me kind comments.

Dr. Rumiko Kofuji (Kanazawa university) for sequencing the plastid genome of *G. stellata*, and giving me kind comments.

Dr. Goro Tanifuji (university of Tsukuba) for kindly providing me RNA-seq data of *B. natans* (Chapter 4).

Mr. Shu Shirato (graduate university for advanced studies) for assembling the nucleomorph genome of *A. amoebiformis*, and encouraging me.

Finally, I would like to thank all members in my laboratory of university of Tsukuba for providing me great environment for this study. I am a recipient of the Japan Society for the Promotion of Science (JSPS) Research Fellowships for Young Scientists 26-572.

7. References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res. 25:3389–3402.

Anders S, Pyl PT, Huber W. 2014. HTSeq a python framework to work with high-throughput sequencing data. Bioinformatics 31:166–169.

Archibald JM. 2007. Nucleomorph genomes: structure, function, origin and evolution. BioEssays 29:392–402.

Archibald JM. 2009. The Puzzle of Plastid Evolution. Curr. Biol. 19:R81–R88.

Archibald JM, Lane CE. 2009. Going, going, not quite gone: Nucleomorphs as a case study in nuclear genome reduction. J. Hered. 100:582–590.

Archibald JM, Rogers MB, Toop M, Ishida K, Keeling PJ. 2003. Lateral gene transfer and the evolution of plastid-targeted proteins in the secondary plastid-containing alga *Bigelowiella natans*. Proc. Natl. Acad. Sci. U. S. A. 100:7678–7683.

Asamizu E, Ichihara H, Nakaya A, Nakamura Y, Hirakawa H, Ishii T, et al. 2014. Plant Genome DataBase Japan (PGDBj): A portal website for the integration of plant genome-related databases. Plant Cell Physiol. 55:e8–e8.

Ball SG, Subtil A, Bhattacharya D, Moustafa A, Weber APM, Gehre L, et al. 2013. Metabolic effectors secreted by bacterial pathogens: essential facilitators of plastid endosymbiosis? Plant Cell 25:7–21.

Barbrook AC, Howe CJ, Purton S. 2006. Why are plastid genomes retained in non-photosynthetic organisms? Trends Plant Sci. 11:101–108.

Besendahl A, Qiu YL, Lee J, Palmer JD, Bhattacharya D. 2000. The cyanobacterial origin and vertical transmission of the plastid tRNA(Leu) group-I intron. Curr. Genet. 37:12–23.

Blanc G, Gallot-Lavallée L, Maumus F. 2015. Provirophages in the *Bigelowiella* genome bear testimony to past encounters with giant viruses. Proc. Natl. Acad. Sci. 112:E5318–E5326.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.

Brouard JS, Otis C, Lemieux C, Turmel M. 2011. The chloroplast genome of the green alga *Schizomeris*

leibleinii (Chlorophyceae) provides evidence for bidirectional DNA replication from a single origin in the Chaetophorales. Genome Biol. Evol. 3:505–515.

Calderon-Saenz E, Schnetter R. 1987. *Cryptochlora perforans*, a new genus and species of algae (Chlorarachniophyta), capable of penetrating dead algal filaments. Plant Syst. Evol. 158:69–71.

Calderon-Saenz E, Schnetter R. 1989. Morphology, biology, and systematics of *Cryptochlora perforans* (Chlorarachniophyta), a phagotrophic marine alga. Plant Syst. Evol. 163:165–176.

Chanda I, Pan A, Dutta C. 2005. Proteome Composition in *Plasmodium falciparum*: Higher Usage of GC-Rich Nonsynonymous Codons in Highly Expressed Genes. J. Mol. Evol. 61:513–523.

Clark KB, Jensen KR, Stirts HM. 1990. Survey for functional kleptoplasty among West Atlantic *Ascoglossa* (equals *Sacoglossa*) (Mollusca: Opisthobranchia). Veliger 33: 339–345.

Curtis BA, Archibald JM. 2010. A spliceosomal intron of mitochondrial DNA origin. Curr. Biol. 20:R919–20.

Curtis BA, Tanifuji G, Burki F, Gruber A, Irimia M, Maruyama S, et al. 2012. Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. Nature 492:59–65.

de Hoon MJL, Imoto S, Nolan J, Miyano S. 2004. Open source clustering software. Bioinformatics 20:1453–1454.

de Vries J, Habicht J, Woehle C, Huang C, Christa G, Wagele H, et al. 2013. Is ftsH the key to plastid longevity in sacoglossan slugs? Genome Biol. Evol. 5:2540–2548.

de Vries J, Sousa FL, Bölter B, Soll J, Gould SB. 2015. YCF1: A green TIC? Plant Cell 27:1827–1833.

del Campo J, Not F, Forn I, Sieracki ME, Massana R. 2013. Taming the smallest predators of the oceans. ISME J. 7:351–358.

Delwiche CF, Palmer JD. 1996. Rampant horizontal transfer and duplication of rubisco genes in eubacteria and plastids. Mol. Biol. Evol. 13:873–882.

Deshmukh M, Stark J, Yeh LC, Lee JC, Woolford JL. 1995. Multiple regions of yeast ribosomal protein L1 are important for its interaction with 5 S rRNA and assembly into ribosomes. J. Biol. Chem. 270:30148–56.

Donaher N, Tanifuji G, Onodera NT, Malfatti SA, Chain PSG, Hara Y, et al. 2009. The complete plastid genome sequence of the secondarily nonphotosynthetic alga *Cryptomonas paramecium*: reduction,

compaction, and accelerated evolutionary rate. Genome Biol. Evol. 1:439–448.

Douglas S, Zauner S, Fraunholz M, Beaton M, Penny S, Deng LT, et al. 2001. The highly reduced genome of an enslaved algal nucleus. Nature 410:1091–1096.

Douglas SE, Murphy CA, Spencer DF, Gray MW. 1991. Cryptomonad algae are evolutionary chimaeras of two phylogenetically distinct unicellular eukaryotes. Nature 350:148–151.

Douglas SE, Penny SL. 1999. The plastid genome of the cryptophyte alga, *Guillardia theta*: Complete sequence and conserved synteny groups confirm its common ancestry with red algae. J. Mol. Evol. 48:236–244.

Eberhard S, Finazzi G, Wollman F-A. 2008. The dynamics of photosynthesis. Annu. Rev. Genet. 42:463–515.

Eschbach S, Hofmann CJ, Maier UG, Sitte P, Hansmann P. 1991. A eukaryotic genome of 660 kb: electrophoretic karyotype of nucleomorph and cell nucleus of the cryptomonad alga, *Pyrenomonas salina*. Nucleic Acids Res. 19:1779–1781.

Farinas B, Mary C, De O Manes CL, Bhaud Y, Peaucellier G, Moreau H. 2006. Natural synchronisation for the study of cell division in the green unicellular alga *Ostreococcus tauri*. Plant Mol. Biol. 60:277–292.

Filée J, Forterre P. 2005. Viral proteins functioning in organelles: a cryptic origin? Trends Microbiol. 13:510–513.

Filée J, Forterre P, Sen-Lin T, Laurent J. 2002. Evolution of DNA polymerase families: evidences for multiple gene exchange between cellular and viral proteins. J. Mol. Evol. 54:763–73.

Fortunato AE, Annunziata R, Jaubert M, Bouly JP, Falciatore A. 2014. Dealing with light: The widespread and multitasking cryptochrome/photolyase family in photosynthetic organisms. Journal of Plant Physiology.

Fujiwara T, Misumi O, Tashiro K, Yoshida Y, Nishida K, Yagisawa F, et al. 2009. Periodic gene expression patterns during the highly synchronized cell nucleus and organelle division cycles in the unicellular red alga *Cyanidioschyzon merolae*. DNA Res. 16:59–72.

Fulnečková J, Hasíková T, Fajkus J, Lukešová A, Eliáš M, Sýkorová E. 2012. Dynamic evolution of telomeric sequences in the green algal order Chlamydomonadales. Genome Biol. Evol. 4:248–264.

Gile GH, Keeling PJ. 2008. Nucleus-encoded periplastid-targeted EFL in Chlorarachniophytes. Mol. Biol. Evol. 25:1967–1977.

Gillott MA, Gibbs SP. 1980. Cryptomonad nucleomorph: its ultrastructure and evolutionary significance. J. Phycol. 16:558–568.

Gilson P, McFadden G. 1999. Molecular, morphological and phylogenetic characterization of six chlorarachniophyte strains. Phycol. Res. 47: 7–19.

Gilson PR, Su V, Slamovits CH, Reith ME, Keeling PJ, McFadden GI. 2006. Complete nucleotide sequence of the chlorarachniophyte nucleomorph: Nature's smallest nucleus. Proc. Natl. Acad. Sci. 103:9566–9571.

Gomord V, Denmat L-A, Fitchette-Laine A-C, Satiat-Jeunemaitre B, Hawes C, Faye L. 1997. The C-terminal HDEL sequence is sufficient for retention of secretory proteins in the endoplasmic reticulum (ER) but promotes vacuolar targeting of proteins that escape the ER. Plant J. 11:313–325.

Gould SB, Sommer MS, Hadfi K, Zauner S, Kroth PG, Maier UG. 2006a. Protein targeting into the complex plastid of cryptophytes. J. Mol. Evol. 62:674–681.

Gould SB, Sommer MS, Kroth PG, Gile GH, Keeling PJ, Maier UG. 2006b. Nucleus-to-nucleus gene transfer and protein retargeting into a remnant cytoplasm of cryptophytes and diatoms. Mol. Biol. Evol. 23:2413–2422.

Gould SB, Waller RF, McFadden GI. 2008. Plastid evolution. Annu. Rev. Plant Biol. 59:491–517.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29:644–652.

Green BR. 2011. Chloroplast genomes of photosynthetic eukaryotes. Plant J. 66:34–44.

Greenwood AD. 1974. The Cryptophyta in relation to phylogeny and photosynthesis. In Electron microscopy (J. V. Sanders and D.J. Goodchildeds.), pp. 566–567, Australian academy of sciences, Canbera.

Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith Jr RK, Hannick LI, et al. 2003. Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res. 31:5654–5666.

Hansmann P, Falk H, Sitte P. 1985. DNA in the nucleomorph of cryptomonas demonstrated by DAPI fluorescence. Zeitschrift Fur Naturforsch. C-A J. Biosci. 40:933–935.

Hardcastle TJ, Kelly KA. 2010. baySeq: empirical Bayesian methods for identifying differential expression in sequence count data. BMC Bioinformatics 11:422.

Helt GA, Nicol JW, Erwin E, Blossom E, Blanchard SG, Chervitz SA, et al. 2009. Genoviz Software Development Kit: Java tool kit for building genomics visualization applications. BMC Bioinformatics 10:266.

Hempel F, Bullmann L, Lau J, Zauner S, Maier UG. 2009. ERAD-derived preprotein transport across the second outermost plastid membrane of diatoms. Mol. Biol. Evol. 26:1781–1790.

Hibberd DJ, Norris RE. 1984. Cytology and ultrastructure of *Chlorarachnion reptans* (Chlorarachniophyta divisio nova, Chlorarachniophyceae classis nova). J. Phycol. 20:310–330.

Hihara Y, Kamei A, Kanehisa M, Kaplan A, Ikeuchib M. 2001. DNA microarray analysis of cyanobacterial gene expression during acclimation to high light DNA microarray analysis of cyanobacterial gene expression. Plant Cell 13:793–806.

Hilker R, Sickinger C, Pedersen CNS, Stoye J. 2012. UniMoG-a unifying framework for genomic distance calculation and sorting based on DCJ. Bioinformatics 28:2509–2511.

Hirakawa Y. 2014. Complex plastids of chlorarachniophyte algae. Perspect. Phycol. 1:87–92.

Hirakawa Y, Burki F, Keeling PJ. 2012. Genome-based reconstruction of the protein import machinery in the secondary plastid of a chlorarachniophyte alga. Eukaryot. Cell 11:324–333.

Hirakawa Y, Burki F, Keeling PJ. 2011. Nucleus- and nucleomorph-targeted histone proteins in a chlorarachniophyte alga. Mol. Microbiol. 80:1439–1449.

Hirakawa Y, Gile GH, Ota S, Keeling PJ, Ishida KI. 2010. Characterization of periplastidal compartment-targeting signals in chlorarachniophytes. Mol. Biol. Evol. 27:1538–1545.

Hirakawa Y, Ishida K. 2015. Prospective function of FtsZ proteins in the secondary plastid of chlorarachniophyte algae. BMC Plant Biol. 15:276.

Hirakawa Y, Ishida K. 2014. Polyploidy of endosymbiotically derived genomes in complex algae. Genome Biol. Evol. 6:974–80.

Hirakawa Y, Kofuji R, Ishida K. 2008. Transient transformation of a chlorarachniophyte alga, *Lotharella*

amoebiformis (Chlorarachniophyceae), with *uid* and *egfp* reporter genes. J. Phycol. 44:814–820.

Hirakawa Y, Nagamune K, Ishida K. 2009. Protein targeting into secondary plastids of chlorarachniophytes. Proc. Natl. Acad. Sci. 106:12820–12825.

Hopkins JF, Spencer DF, Laboissiere S, Neilson JAD, Eveleigh RJM, Durnford DG, et al. 2012. Proteomics reveals plastid- and periplastid-targeted proteins in the chlorarachniophyte alga *Bigelowiella natans*. Genome Biol. Evol. 4:1391–406.

Hovde BT, Deodato CR, Hunsperger HM, Ryken SA, Barlow B, Starkenburg SR, et al. 2015. Genome sequence and transcriptome analyses of *Chrysochromulina tobin*: metabolic tools for enhanced algal fitness in the prominent order Prymnesiales (Haptophyceae). PLoS Genet. 11:1–31.

Hsu PY, Harmer SL. 2014. Wheels within wheels: the plant circadian system. Trends Plant Sci. 19:240–249.

Idoine AD, Boulouis A, Rupprecht J, Bock R. 2014. The diurnal logic of the expression of the chloroplast genome in *Chlamydomonas reinhardtii*. PLoS One 9:e108760.

Imoto Y, Yoshida Y, Yagisawa F, Kuroiwa H, Kuroiwa T. 2011. The cell cycle, including the mitotic cycle and organelle division cycles, as revealed by cytological observations. Microscopy 60:S117–S136.

Ishida K. 2005. Protein targeting into plastids: a key to understanding the symbiogenetic acquisitions of plastids. J. Plant Res. 118:237–245.

Ishida K, Cao Y, Hasegawa M, Okada N, Hara Y. 1997. The origin of chlorarachniophyte plastids, as inferred from phylogenetic comparisons of amino acid sequences of EF-Tu. J. Mol. Evol. 45:682–687.

Ishida K, Endo H, Koike S. 2011. *Partenskyella glossopodia* (Chlorarachniophyceae) possesses a nucleomorph genome of approximately 1 Mbp. Phycol. Res. 59:120–122.

Ishida K, Green BR, Cavalier-Smith T. 1999. Diversification of a chimaeric algal group, the chlorarachniophytes: phylogeny of nuclear and nucleomorph small-subunit rRNA genes. Mol. Biol. Evol. 16: 321–331.

Ishida KI, Ishida N, Hara Y. 2000. *Lotharella amoeboformis* sp. nov.: A new species of chlorarachniophytes from Japan. Phycol. Res. 48:221–229.

Janouškovec J, Liu S-L, Martone PT, Carré W, Leblanc C, Collén J, et al. 2013. Evolution of red algal plastid genomes: ancient architectures, introns, horizontal gene transfer, and taxonomic utility of plastid markers. PLoS One 8:e59001.

Kanesaki Y, Imamura S, Minoda A, Tanaka K. 2012. External light conditions and internal cell cycle phases coordinate accumulation of chloroplast and mitochondrial transcripts in the red alga *Cyanidioschyzon merolae*. DNA Res. 19:289–303.

Kasai F, Kawachi M, Erata M, Yumoto K, Sato M. 2009. NIES-collection list of strains, 8th edition. Jpn J Phycol 57: 220.

Katoh K, Toh H. 2008. Recent developments in the MAFFT multiple sequence alignment program. Brief. Bioinform. 9:286–298.

Kaundal R, Saini R, Zhao PX. 2010. Combining machine learning and homology-based approaches to accurately predict subcellular localization in *Arabidopsis*. PLANT Physiol. 154:36–54.

Keeling PJ. 2010. The endosymbiotic origin, diversification and fate of plastids. Philos. Trans. R. Soc. B Biol. Sci. 365:729–748.

Keeling PJ, Burki F, Wilcox HM, Allam B, Allen EE, Amaral-Zettler LA, et al. 2014. The marine microbial eukaryote transcriptome sequencing project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. PLoS Biol. 12:e1001889.

Keeling PJ, Slamovits CH. 2005. Causes and effects of nuclear genome reduction. Curr. Opin. Genet. Dev. 15:601–608.

Khan H, Parks N, Kozera C, Curtis BA, Parsons BJ, Bowman S, et al. 2007. Plastid genome sequence of the cryptophyte alga *Rhodomonas salina* CCMP1319: Lateral transfer of putative DNA replication machinery and a test of chromist plastid phylogeny. Mol. Biol. Evol. 24:1832–1842.

Kikuchi S, Bédard J, Hirano M, Hirabayashi Y, Oishi M, Imai M, et al. 2013. Uncovering the protein translocon at the chloroplast inner envelope membrane. Science 339:571–4.

Kim JI, Yoon HS, Yi G, Kim HS, Yih W, Shin W. 2015. The plastid genome of the cryptomonad *Teleaulax amphioxeia*. PLoS One 10:e0129284.

Kin T, Yamada K, Terai G, Okida H, Yoshinari Y, Ono Y, et al. 2007. fRNAdb: a platform for

mining/annotating functional RNA candidates from non-coding RNA sequences. Nucleic Acids Res. 35:D145–D148.

Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. J. Mol. Biol. 305:567–580.

Kucho KI, Okamoto K, Tabata S, Fukuzawa H, Ishiura M. 2005. Identification of novel clock-controlled genes by cDNA macroarray analysis in *Chlamydomonas reinhardtii*. Plant Mol. Biol. 57:889–906.

Kuhsel MG, Strickland R, Palmer JD. 1990. An ancient group I intron shared by eubacteria and chloroplasts. Science 250: 1570–3.

Lagesen K, Hallin P, Rodland EA, Staerfeldt H-H, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 35:3100–3108.

Lambowitz AM, Zimmerly S. 2004. Mobile group II introns. Annu. Rev. Genet. 38:1–35.

Lane CE, Archibald JM. 2006. Novel nucleomorph genome architecture in the cryptomonad genus *Hemiselmis*. J. Eukaryot. Microbiol. 53:515–21.

Lane CE, van den Heuvel K, Kozera C, Curtis BA, Parsons BJ, Bowman S, et al. 2007. Nucleomorph genome of *Hemiselmis andersenii* reveals complete intron loss and compaction as a driver of protein structure and function. Proc. Natl. Acad. Sci. U. S. A. 104:19908–19913.

Lang BF, Laforest M-J, Burger G. 2007. Mitochondrial introns: a critical view. Trends Genet. 23:119–125.

Leliaert F, Lopez-Bautista JM. 2015. The chloroplast genomes of *Bryopsis plumosa* and *Tydemania expeditiones* (Bryopsidales, Chlorophyta): compact genomes and genes of bacterial origin. BMC Genomics 16:204.

Lemieux C, Otis C, Turmel M. 2014. Six newly sequenced chloroplast genomes from prasinophyte green algae provide insights into the relationships among prasinophyte lineages and the diversity of streamlined genome architecture in picoplanktonic species. BMC Genomics 15:857.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–60.

Li QB, Haskell D, Zhang C, Sung DY, Guy C. 2000. Diurnal regulation of Hsp70s in leaf tissue. Plant J. 21:373–378.

Lien T, Knutsen G. 1979. Synchronous growth of *Chlamydomonas reinhardtii* (Chlorophyceae): a review of optimal conditions. J. Phycol. 15:191–200.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15:550.

Lü F, Xü W, Tian C, Wang G, Niu J, Pan G, et al. 2011. The *Bryopsis hypnoides* plastid genome: multimeric forms and complete nucleotide sequence. PLoS One 6:e14663.

Ludwig M, Gibbs S. 1987. Are the nucleomorphs of cryptomonads and *Chlorarachnion* the vestigial nuclei of eukaryotic endosymbionts? Ann. N. Y. Acad. Sci. 503:198–211.

Ludwig M, Gibbs SP. 1985. DNA is present in the nucleomorph of cryptomonads: further evidence that the chloroplast evolved from a eukaryotie endosymbiont. Protoplasma 20:9–20.

Ma L, Li J, Qu L, Hager J, Chen Z, Zhao H, et al. 2001. Light control of *Arabidopsis* development entails coordinated regulation of genome expression and cellular pathways. Plant Cell 13:2589–2607.

Maisnier-Patin S, Roth JR, Fredriksson A, Nyström T, Berg OG, Andersson DI. 2005. Genomic buffering mitigates the effects of deleterious mutations in bacteria. Nat. Genet. 37:1376–1379.

Martin W, Herrmann RG. 1998. Gene transfer from organelles to the nucleus: how much, what happens, and why? Plant Physiol. 118:9–17.

Maruyama S, Sugahara J, Kanai A, Nozaki H. 2010. Permuted tRNA genes in the nuclear and nucleomorph genomes of photosynthetic eukaryotes. Mol. Biol. Evol. 27:1070–1076.

Maruyama S, Suzaki T, Weber APM, Archibald JM, Nozaki H. 2011. Eukaryote-to-eukaryote gene transfer gives rise to genome mosaicism in euglenids. BMC Evol. Biol. 11:105.

McCutcheon JP, Moran NA. 2011. Extreme genome reduction in symbiotic bacteria. Nat. Rev. Microbiol. 10:13–26.

McFadden GI, Gilson PR, Hofmann CJ, Adcock GJ, Maier UG. 1994. Evidence that an amoeba acquired a chloroplast by retaining part of an engulfed eukaryotic alga. Proc. Natl. Acad. Sci. 91:3690–3694.

Miyagishima S. 2011. Mechanism of plastid division: from a bacterium to an organelle. Plant Physiol. 155:1533–44.

Miyagishima S-Y, Suzuki K, Okazaki K, Kabeya Y. 2012. Expression of the nucleus-encoded chloroplast division genes and proteins regulated by the algal cell cycle. Mol. Biol. Evol. 29:2957–70.

Moestrup O, Sengco M. 2001. Ultrastructural studies on *Bigelowiella natans*, gen. et sp. nov., A chlorarachniophyte flagellate. J. Phycol. 37:624–646.

Monnier A, Liverani S, Bouvet R, Jesson B, Smith JQ, Mosser J, et al. 2010. Orchestrated transcription of biological processes in the marine picoeukaryote *Ostreococcus* exposed to light/dark cycles. BMC Genomics 11:192.

Moore CE, Curtis B, Mills T, Tanifuji G, Archibald JM. 2012. Nucleomorph genome sequence of the cryptophyte alga *Chroomonas mesostigmatica* CCMP1168 reveals lineage-specific gene loss and genome complexity. Genome Biol. Evol. 4:1162–1175.

Moore D, Dowhan D, Chory J, Ribaudo RK. 2002. Isolation and purification of large DNA restriction fragments from agarose gels. Curr. Protoc. Mol. Biol. Chapter 2:Unit 2.6.

Moran NA. 1996. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. Proc. Natl. Acad. Sci. U. S. A. 93:2873–2878.

Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. 2007. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 35:W182–W185.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods 5:621–628.

Nefissi R, Natsui Y, Miyata K, Oda A, Hase Y, Nakagawa M, et al. 2011. Double loss-of-function mutation in *EARLY FLOWERING 3* and *CRYPTOCHROME 2* genes delays flowering under continuous light but accelerates it under long days and short days: an important role for *Arabidopsis CRY2* to accelerate flowering time in continuous light. J. Exp. Bot. 62:2731–2744.

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: A Fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32:268–274.

Noordally ZB, Millar AJ. 2015. Clocks in algae. Biochemistry 54:171–183.

Ota S, Ueda K, Ishida K. 2005. *Lotharella vacuolata* sp. nov., a new species of chlorarachniophyte algae, and time-lapse video observations on its unique post-cell division behavior. Phycol. Res. 53:275–286.

Palmer JD. 1997. Organelle genomes-going, going, gone! Science 275:790–790.

Patron NJ, Rogers MB, Keeling PJ. 2006. Comparative rates of evolution in endosymbiotic nuclear

genomes. BMC Evol. Biol. 6:46.

Phipps KD, Donaher NA., Lane CE, Archibald JM. 2008. Nucleomorph karyotype diversity in the freshwater cryptophyte genus Cryptomonas. J. Phycol. 44:11–14.

Price DC, Chan CX, Yoon HS, Yang EC, Qiu H, Weber APM, et al. 2012. *Cyanophora paradoxa* genome elucidates origin of photosynthesis in algae and plants. Science 335:843–847.

Pudasaini A, Zoltowski BD. 2013. Zeitlupe senses blue-light fluence to mediate circadian timing in *Arabidopsis thaliana*. Biochemistry 52:7150–7158.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842.

Rensing SA, Goddemeier M, Hofmann CJ, Maier UG. 1994. The presence of a nucleomorph *hsp70* gene is a common feature of Cryptophyta and Chlorarachniophyta. Curr. Genet. 26:451–455.

Rice DW, Palmer JD. 2006. An exceptional horizontal gene transfer in plastids: gene replacement by a distant bacterial paralog and evidence that haptophyte and cryptophyte plastids are sisters. BMC Biol. 4:31.

Rodríguez-Ezpeleta N, Brinkmann H, Burey SC, Roure B, Burger G, Löffelhardt W, et al. 2005. Monophyly of primary photosynthetic eukaryotes: Green plants, red algae, and glaucophytes. Curr. Biol. 15:1325–1330.

Rogers MB, Archibald JM, Field MA, Li C, Striepen B, Keeling PJ. 2004. Plastid-targeting peptides from the chlorarachniophyte *Bigelowiella natans*. J. Eukaryot. Microbiol. 51:529–535.

Rogers MB, Gilson PR, Su V, McFadden GI, Keeling PJ. 2007. The complete chloroplast genome of the chlorarachniophyte *Bigelowiella natans*: Evidence for independent origins of chlorarachniophyte and euglenid secondary endosymbionts. Mol. Biol. Evol. 24:54–62.

Ronquist F, Teslenko M, van Der Mark P, Ayres DL, Darling A, Höhna S, et al. 2012. Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61:539–542.

Rossel JB, Wilson IW, Pogson BJ. 2002. Global changes in gene expression in response to high light in *Arabidopsis*. Plant Physiol. 130:1109–1120.

Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, et al. 2000. Artemis: sequence

visualization and annotation. Bioinformatics 16:944–945.

Saibil H. 2013. Chaperone machines for protein folding, unfolding and disaggregation. Nat. Rev. Mol. Cell Biol. 14:630–642.

Saldanha AJ. 2004. Java Treeview–extensible visualization of microarray data. Bioinformatics 20:3246–8.

Schaber J, Rispe C, Wernegreen J, Buness A, Delmotte F, Silva FJ, et al. 2005. Gene expression levels influence amino acid usage and evolutionary rates in endosymbiotic bacteria. Gene 352:109–17.

Schaffer R, Landgraf J, Accerbi M, Simon V, Larson M, Wisman E. 2001. Microarray analysis of diurnal and circadian-regulated genes in *Arabidopsis*. Plant Cell 13:113–123.

Schattner P, Brooks AN, Lowe TM. 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res. 33:W686–W689.

Silver TD, Koike S, Yabuki A, Kofuji R, Archibald JM, Ishida K-I. 2007. Phylogeny and nucleomorph karyotype diversity of chlorarachniophyte algae. J. Eukaryot. Microbiol. 54:403–10.

Silver TD, Moore CE, Archibald JM. 2010. Nucleomorph ribosomal DNA and telomere dynamics in chlorarachniophyte algae. J. Eukaryot. Microbiol. 57:453–9.

Slamovits CH, Keeling PJ. 2009. Evolution of ultrasmall spliceosomal introns in highly reduced nuclear genomes. Mol. Biol. Evol. 26:1699–1705.

Smith DR, Crosby K, Lee RW. 2011. Correlation between nuclear plastid DNA abundance and plastid number supports the limited transfer window hypothesis. Genome Biol. Evol. 3:365–71.

Soma A, Onodera A, Sugahara J, Kanai A, Yachie N, Tomita M, et al. 2007. Permuted tRNA genes expressed via a circular RNA intermediate in *Cyanidioschyzon merolae*. Science 318:450–453.

Sommer MS, Gould SB, Lehmann P, Gruber A, Przyborski JM, Maier UG. 2007. Der1-mediated preprotein import into the periplastid compartment of chromalveolates? Mol. Biol. Evol. 24:918–928.

Staley JP, Guthrie C. 1998. Mechanical devices of the spliceosome: motors, clocks, springs, and things. Cell 92: 315–26.

Staley JP, Woolford JL. 2009. Assembly of ribosomes and spliceosomes: complex ribonucleoprotein machines. Curr. Opin. Cell Biol. 21:109–118.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313.

Standage DS, Brendel VP. 2012. ParsEval: parallel comparison and analysis of gene structure annotations. BMC Bioinformatics 13:187.

Stork S, Lau J, Moog D, Maier UG. 2013. Three old and one new: Protein import into red algal-derived plastids surrounded by four membranes. Protoplasma 250:1013–1023.

Stork S, Moog D, Przyborski JM, Wilhelmi I, Zauner S, Maier UG. 2012. Distribution of the SELMA translocon in secondary plastids of red algal origin and predicted uncoupling of ubiquitin-dependent translocation from degradation. Eukaryot. Cell 11:1472–1481.

Sugahara J, Yachie N, Sekine Y, Soma A. 2006. SPLITS: a new program for predicting split and intron-containing tRNA genes at the genome level. In Silico Biol. 6: 411–8.

Sun J, Nishiyama T, Shimizu K, Kadota K. 2013. TCC: an R package for comparing tag count data with robust normalization strategies. BMC Bioinformatics 14:219.

Taipale M, Jarosz DF, Lindquist S. 2010. HSP90 at the hub of protein homeostasis: emerging mechanistic insights. Nat. Rev. Mol. Cell Biol. 11:515–528.

Takahashi F, Okabe Y, Nakada T, Sekimoto H, Ito M, Kataoka H, et al. 2007. Origins of the secondary plastids of euglenophyta and chlorarachniophyta as revealed by an analysis of the plastid-targeting, nuclear-encoded gene *psbO*. J. Phycol. 43:1302–1309.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol. Biol. Evol. 30:2725–2729.

Tanifuji G, Onodera NT, Brown MW, Curtis BA, Roger AJ, Ka-Shu Wong G, et al. 2014a. Nucleomorph and plastid genome sequences of the chlorarachniophyte *Lotharella oceanica*: convergent reductive evolution and frequent recombination in nucleomorph-bearing algae.

Tanifuji G, Onodera NT, Hara Y. 2010. Nucleomorph genome diversity and its phylogenetic implications in cryptomonad algae. Phycol. Res. 58:230–237.

Tanifuji G, Onodera NT, Moore CE, Archibald JM. 2014b. Reduced nuclear genomes maintain high gene transcription levels. Mol. Biol. Evol. 31:625–635.

Tanifuji G, Onodera NT, Wheeler TJ, Dlutek M, Donaher N, Archibald JM. 2011. Complete

nucleomorph genome sequence of the nonphotosynthetic alga *Cryptomonas paramecium* reveals a core nucleomorph gene set. Genome Biol. Evol. 3:44–54.

Tardif M, Atteia A, Specht M, Cogne G, Rolland N, Brugiere S, et al. 2012. PredAlgo: a new subcellular localization prediction tool dedicated to green algae. Mol. Biol. Evol. 29:3625–3639.

Timmis JN, Ayliffe MA, Huang CY, Martin W. 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat. Rev. Genet. 5:123–135.

Tokuriki N, Tawfik DS. 2009. Chaperonin overexpression promotes genetic variation and enzyme evolution. Nature 459:668–673.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25:1105–11.

Turmel M, Otis C, Lemieux C. 2015. Dynamic evolution of the chloroplast genome in the green algal classes Pedinophyceae and Trebouxiophyceae. Genome Biol. Evol. 7:2062–2082.

Turmel M, Otis C, Lemieux C. 2009. The chloroplast genomes of the green algae *Pedinomonas minor*, *Parachlorella kessleri*, and *Oocystis solitaria* reveal a shared ancestry between the Pedinomonadales and Chlorellales. Mol. Biol. Evol. 26:2317–2331.

Van de Peer Y, Rensing SA, Maier UG, de Wachter R. 1996. Substitution rate calibration of small subunit ribosomal RNA identifies chlorarachniophyte endosymbionts as remnants of green algae. Proc. Natl. Acad. Sci. U. S. A. 93:7732–7736.

von Gromoff ED, Treier U, Beck CF. 1989. Three light-inducible heat shock genes of *Chlamydomonas reinhardtii*. Mol. Cell. Biol. 9:3911–8.

Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, et al. 2012. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 40:e49–e49.

Wastl J, Maier UG. 2000. Transport of proteins into cryptomonads complex plastids. J. Biol. Chem. 275:23194–8.

Wenden B, Kozma-Bognár L, Edwards KD, Hall AJW, Locke JCW, Millar AJ. 2011. Light inputs shape the *Arabidopsis* circadian system. Plant J. 66:480–491.

Wicke S, Schneeweiss GM, DePamphilis CW, Müller KF, Quandt D. 2011. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol. Biol. 76:273–297.

Wilcox JL, Dunbar HE, Wolfinger RD, Moran NA. 2003. Consequences of reductive evolution for gene expression in an obligate endosymbiont. Mol. Microbiol. 48:1491–1500.

Williams BAP, Slamovits CH, Patron NJ, Fast NM, Keeling PJ. 2005. A high frequency of overlapping gene expression in compacted eukaryotic genomes. Proc. Natl. Acad. Sci. 102:10936–10941.

Yancopoulos S, Attie O, Friedberg R. 2005. Efficient sorting of genomic permutations by translocation, inversion and block interchange. Bioinformatics 21:3340–3346.

Yang Y, Matsuzaki M, Takahashi F, Qu L, Nozaki H. 2014. Phylogenomic analysis of "red" genes from two divergent species of the "green" secondary phototrophs, the chlorarachniophytes, suggests multiple horizontal gene transfers from the red lineage before the divergence of extant chlorarachniophytes. PLoS One 9:e101158.

Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. 13:555–556.

Yoshida Y, Kuroiwa H, Misumi O, Yoshida M, Ohnuma M, Fujiwara T, et al. 2010. Chloroplasts divide by contraction of a bundle of nanofilaments consisting of polyglucan. Science 329:949–953.

Zones JM, Blaby IK, Merchant SS, Umen JG. 2015. High-resolution profiling of a synchronized diurnal transcriptome from *Chlamydomonas reinhardtii* reveals continuous cell and metabolic differentiation. 27:1–28.

8. Table and Figures

8.1. Figures

Figure 1.1. Diagram of the secondary endosymbioses in chlorarachniophytes and cryptophytes.

The numbers of protein-coding genes and plastid or nucleomorph-targeted proteins are shown. During establishment of secondary endosymbiosis, the nuclear genomes of the endosymbiont were reduced by gene loss and EGT to the host nucleus. Plastid- or nucleomorph-targeted nuclear-encoded protein genes are transcribed and translated by the host cell, and the proteins are transported into the plastid or nucleomorph.

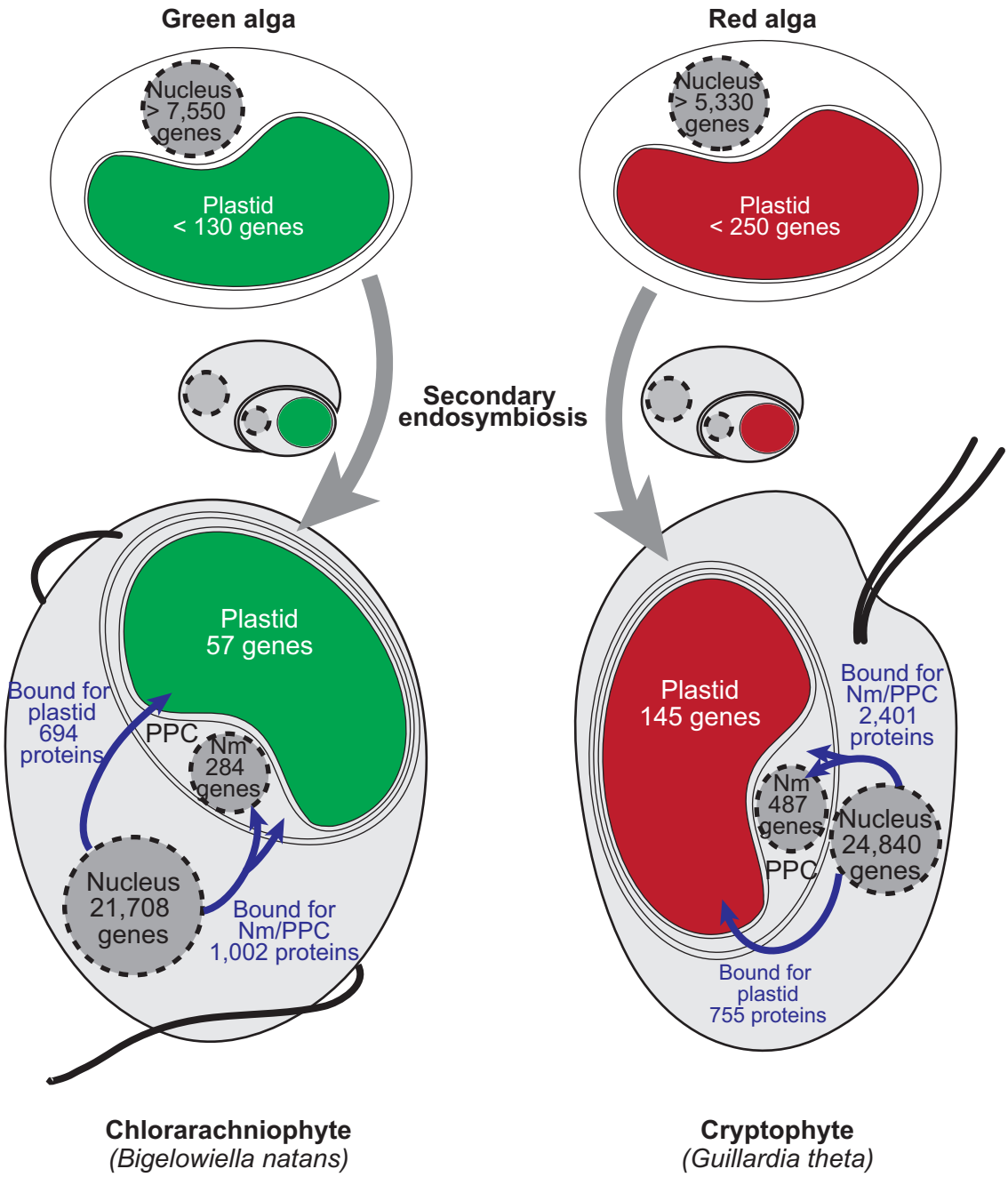

Green alga
Nucleus
> 7,550
genes
Plastid
< 130 genes
Red alga
Nucleus
> 5,330
genes
Plastid
< 250 genes
Secondary
endosymbiosis
Plastid
57 genes
Bound for
plastid
694
proteins
PPC
Nm
284
genes
Nucleus
21,708
genes
Bound for
Nm/PPC
1,002 proteins
Plastid
145 genes
Bound for
Nm/PPC
2,401
proteins
Nm
487
genes
Nucleus
24,840
genes
PPC
Bound for
plastid
755 proteins
Chlorarachniophyte
(Bigelowiella natans)
Cryptophyte
(Guillardia theta)

Figure 2.1. Genome map of the plastid genomes of three chlorarachniophytes.

Plastid genomes of *Gymnochlora stellata* (a), *Partenskyella glossopodia* (b), and *Lotharella vacuolata* (c). Genes on the outside are transcribed in the clockwise direction, and inner genes are transcribed in the counterclockwise direction. Genes are colored according to their function as follows: photosynthesis (green), transcription/translation (pink), ribosomal/transfer RNAs (blue), and miscellaneous (yellow). Introns are showed by black boxes. Inverted repeats (IR) are indicated by thick lines outside the circle. (d) Gene conservation and rearrangement of plastid genomes among five chlorarachniophytes. Thick lines indicate IRs, and shaded regions represent the rearranged genes (*e.g.*, insertion/deletion and coding strand switch) among the plastid genomes.

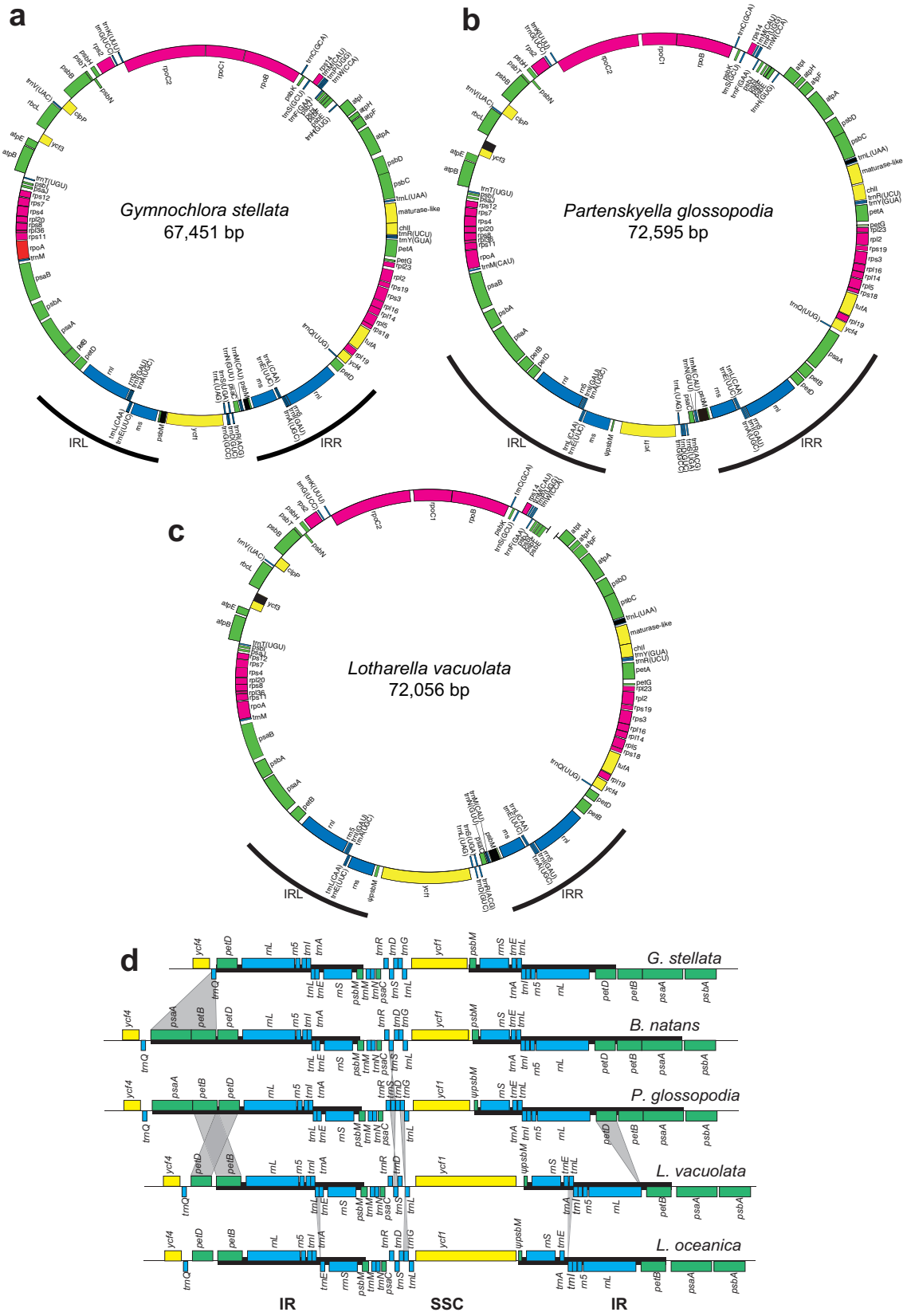

a
Gymnochlora stellata
67,451 bp
IRL
IRR
b
Partenskyella glossopodia
72,595 bp
IRL
IRR
c
Lotharella vacuolata
72,056 bp
IRL
IRR
d
G. stellata
B. natans
P. glossopodia
L. vacuolata
L. oceanica
IR
SSC
IR

Figure 2.2. Intron positions of three plastid genes of chlorarachniophytes.

(a) Schematic image and alignment of plastid *trnL* (UAA) genes in chlorarachniophytes and chlorophytes. The *trnL* genes of *Bigelowiella natans* and *Lotharella vacuolata* lack a group I intron. (b) Alignment of 5′ partial sequences of chlorarachniophyte *ycf3* genes including a group II intron. (c) Alignment of 5′ sequences of *psbM* genes in chlorarachniophytes and *Oocystis solitaria*, showing the conserved position of group II introns. Abbreviations:

Bn, *Bigelowiella natans*; Gs, *Gymnochlora stellata*; Lo, *Lotharella oceanica*; Lv, *Lotharella vacuolata*; Pg, *Partenskyella glossopodia*; Te, *Tydemania expeditiones*; Bp, *Bryopsis plumosa*; Da, *Dicloster acuatus*; Os, *Oocysits solitaria*; Sh, *Stigeoclonium helveticum*; Ao, *Actodesmus obliquus*.

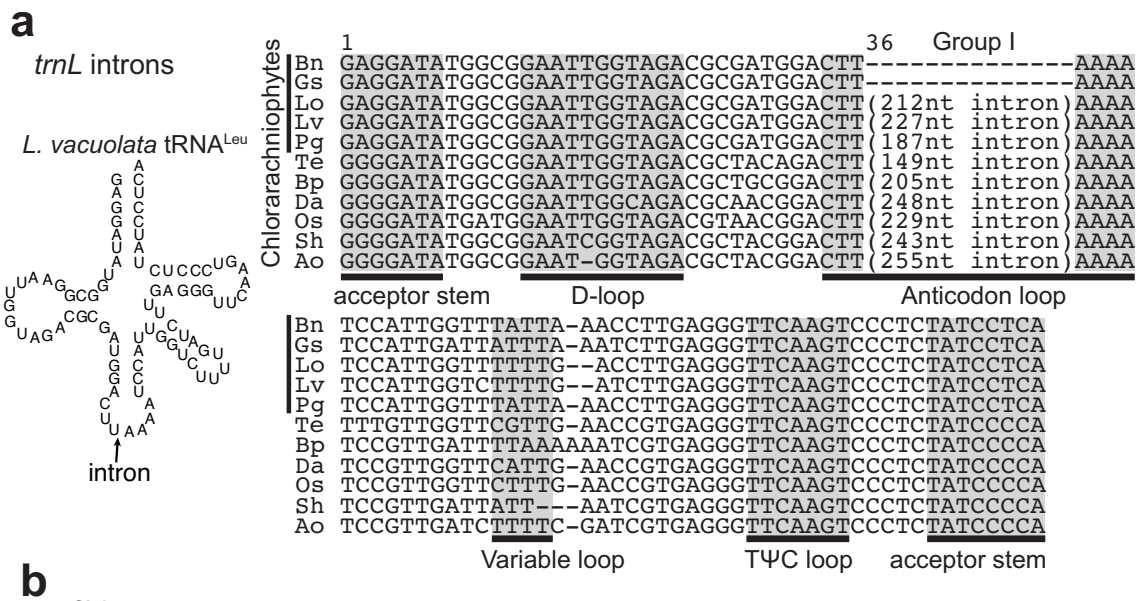

a
trnL introns
L. vacuolata tRNALeu
intron
Chlorarachniophytes
1
36
Group I
Bn GAGGATATGGCGGAATTGGTAGACGCGATGGACTT----------AAAA
Gs GAGGATATGGCGGAATTGGTAGACGCGATGGACTT----------AAAA
Lo GAGGATATGGCGGAATTGGTAGACGCGATGGACTT(212nt intron)AAAA
Lv GAGGATATGGCGGAATTGGTAGACGCGATGGACTT(227nt intron)AAAA
Pg GAGGATATGGCGGAATTGGTAGACGCGATGGACTT(187nt intron)AAAA
Te GGGGATATGGCGGAATTGGTAGACGCTACAGACTT(149nt intron)AAAA
Bp GGGGATATGGCGGAATTGGTAGACGCTGCGGACTT(205nt intron)AAAA
Da GGGGATATGGCGGAATTGGCAGACGCAACGGACTT(248nt intron)AAAA
Os GGGGATATGATGGAATTGGTAGACGTAACGGACTT(229nt intron)AAAA
Sh GGGGATATGGCGGAATCGGTAGACGCTACGGACTT(243nt intron)AAAA
Ao GGGGATATGGCGGAAT-GGTAGACGCTACGGACTT(255nt intron)AAAA
acceptor stem
D-loop
Anticodon loop
Bn TCCATTGGTTTATTA-AACCTTGAGGGTTCAAGTCCCTCTATCCTCA
Gs TCCATTGATTATTTA-AATCTTGAGGGTTCAAGTCCCTCTATCCTCA
Lo TCCATTGGTTTTTTG--ACCTTGAGGGTTCAAGTCCCTCTATCCTCA
Lv TCCATTGGTCTTTTG--ATCTTGAGGGTTCAAGTCCCTCTATCCTCA
Pg TCCATTGGTTTATTA-AACCTTGAGGGTTCAAGTCCCTCTATCCTCA
Te TTTGTTGGTTCGTTG-AACCGTGAGGGTTCAAGTCCCTCTATCCCCA
Bp TCCGTTGATTTTAAAAAATCGTGAGGGTTCAAGTCCCTCTATCCCCA
Da TCCGTTGGTTCATTG-AACCGTGAGGGTTCAAGTCCCTCTATCCCCA
Os TCCGTTGGTTCTTTG-AACCGTGAGGGTTCAAGTCCCTCTATCCCCA
Sh TCCGTTGATTATT---AATCGTGAGGGTTCAAGTCCCTCTATCCCCA
Ao TCCGTTGATCTTTTC-GATCGTGAGGGTTCAAGTCCCTCTATCCCCA
Variable loop
TΨC loop
acceptor stem
b
ycf3 introns
1
34
Group II
Bn ATGCCTAGATCTCAAAAAAATGATAACTTTATC----------GATAAAACATTTACA...
Gs ATGCCTCGATCTCAAAAAAACGATAATTTTATA----------GACAAAACATTTACC...
Lo ATGCCGAGATCTCAAAAAAATGACAATTTTATA(490nt intron)GATAAAACGTTTACG...
Lv ATGCCAAGATCCCAGAAAAATGATAATTTTATA(469nt intron)GATAAAACTTTTACT...
Pg ATGCCAAGATTTCAAAAAAATGACAATTTTATT(435nt intron)GATAAAACATTTACA...
M P R X Q K N D N F I D K T F T
c
psbM introns
1
25
Group II
Bn ATGGAAGTAAATATTTTAGGAGTC----------ATTGCTGTAGCGCTTTTTATA...
Gs ATGGAAGTAAATATTTTAGGATTA(282nt intron)ATTGCAGTAGCACTTTTTATT...
Lo ATGGAAACTAATACTTTAGGATTA(508nt intron)ATAGCAGTAGCACTTTTTATT...
Lv ATGGAAACTAATACTTTAGGATTA(495nt intron)ATAGCAGTAGCACTTTTTATT...
Pg ATGGAAGTAAATACTTTAGGACTA(537nt intron)ATCGCTGTAGCACTTTTTATA...
M E X N X L G X I A V A L F I

Figure 2.3. Maximum likelihood (ML) phylogenic tree of 55 plastid-encoded proteins in chlorarachniophytes and diverse chlorophyte species.

The best tree was reconstructed using the concatenated dataset of 9,876 amino acids. The values at nodes represent bootstrap support that are higher than 50%. Bayesian posterior probabilities (BPP) were calculated by MrBayes and those of >0.5 are shown below each node. Thick lines show BP = 100 and BPP = 1.00. Bar represents 0.2 substitutions per site.

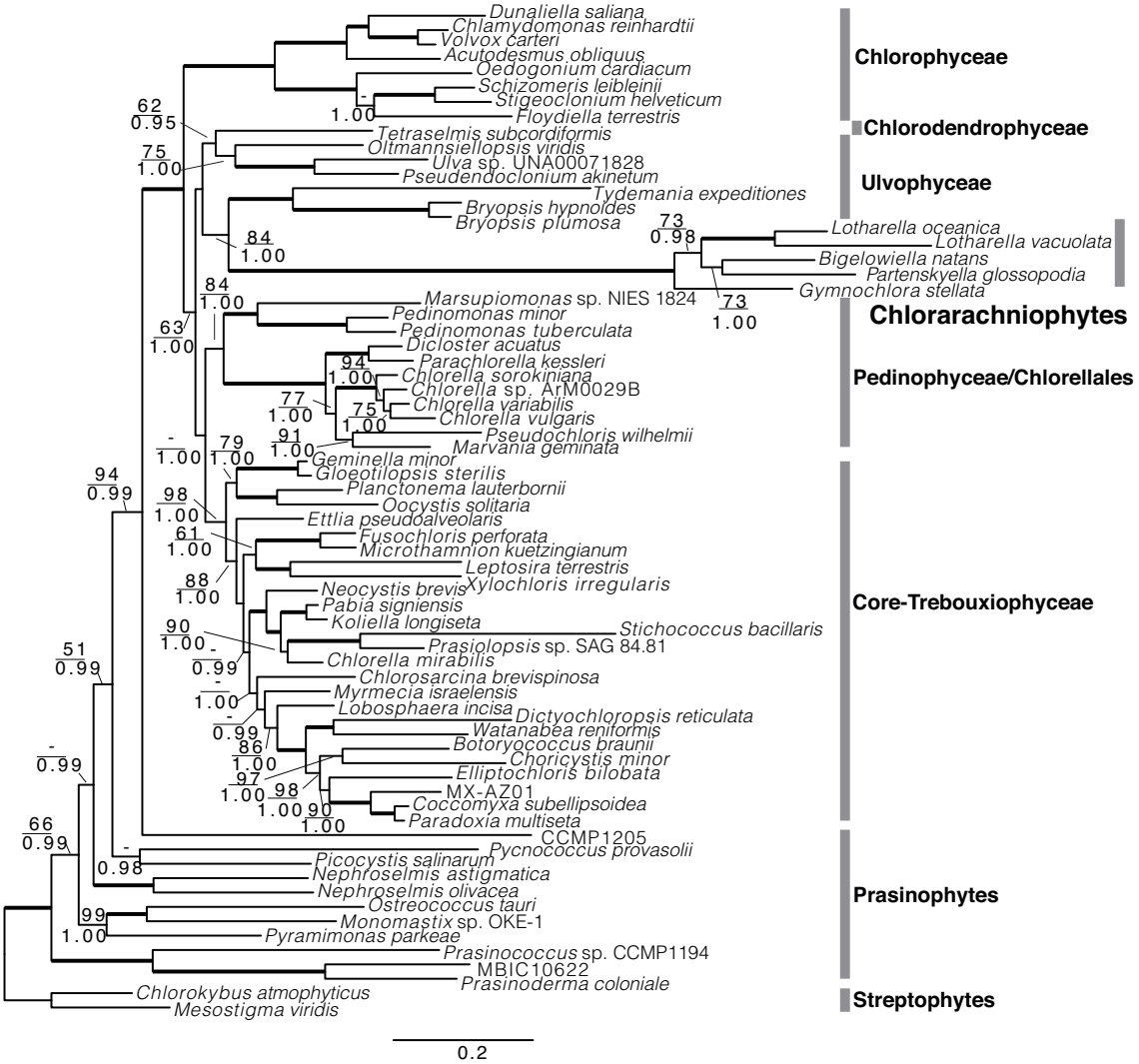

Dunaliella saliana
Chlamydomonas reinhardtii
Volvox carteri
Acutodesmus obliquus
Oedogonium cardiacum
Schizomeris leibleinii
Stigeoclonium helveticum
Floydiella terrestris
Tetraselmis subcordiformis
Oltmannsiellopsis viridis
Ulva sp. UNA00071828
Pseudendoclonium akinetum
Tydemania expeditiones
Bryopsis hypnoides
Bryopsis plumosa
Lotharella oceanica
Lotharella vacuolata
Bigelowiella natans
Partenskyella glossopodia
Gymnochlora stellata
Marsupiomonas sp. NIES 1824
Pedinomonas minor
Pedinomonas tuberculata
Dicloster acuatus
Parachlorella kessleri
Chlorella sorokiniana
Chlorella sp. ArM0029B
Chlorella variabilis
Chlorella vulgaris
Pseudochloris wilhelmii
Marvania geminata
Geminella minor
Gloeotilopsis sterilis
Planctonema lauterbornii
Oocystis solitaria
Ettlia pseudoalveolaris
Fusochloris perforata
Microthamnion kuetzingianum
Leptosira terrestris
Xylochloris irregularis
Neocystis brevis
Pabia signiensis
Koliella longiseta
Stichococcus bacillaris
Prasiolopsis sp. SAG 84.81
Chlorella mirabilis
Chlorosarcina brevispinosa
Myrmecia israelensis
Lobosphaera incisa
Dictyochloropsis reticulata
Watanabea reniformis
Botryococcus braunii
Choricystis minor
Elliptochloris bilobata
MX-AZ01
Coccomyxa subellipsoidea
Paradoxia multiseta
CCMP1205
Pycnococcus provasolii
Picocystis salinarum
Nephroselmis astigmatica
Nephroselmis olivacea
Ostreococcus tauri
Monomastix sp. OKE-1
Pyramimonas parkeae
Prasinococcus sp. CCMP1194
MBIC10622
Prasinoderma coloniale
Chlorokybus atmophyticus
Mesostigma viridis
0.2
Chlorophyceae
Chlorodendrophyceae
Ulvophyceae
Chlorarachniophytes
Pedinophyceae/Chlorellales
Core-Trebouxiophyceae
Prasinophytes
Streptophytes

Figure 3.1. Nucleomorph genome map of the chlorarachniophyte *Amorphochlora amoebiformis*. The genome is comprised of three chromosomes, which are shown as being artificially broken at their midpoint. Genes indicated on the right side are transcribed from top to bottom, and the genes on the left side are transcribed in the opposite direction. Colors of gene blocks correspond to predicted functional categories in the box. Syntenic regions with *B. natans* (blue), *L. vacuolata* (red), *L. oceanica* (gray), and both *L. vacuolata* and *L. oceanica* (green) are shaded by color gradations.

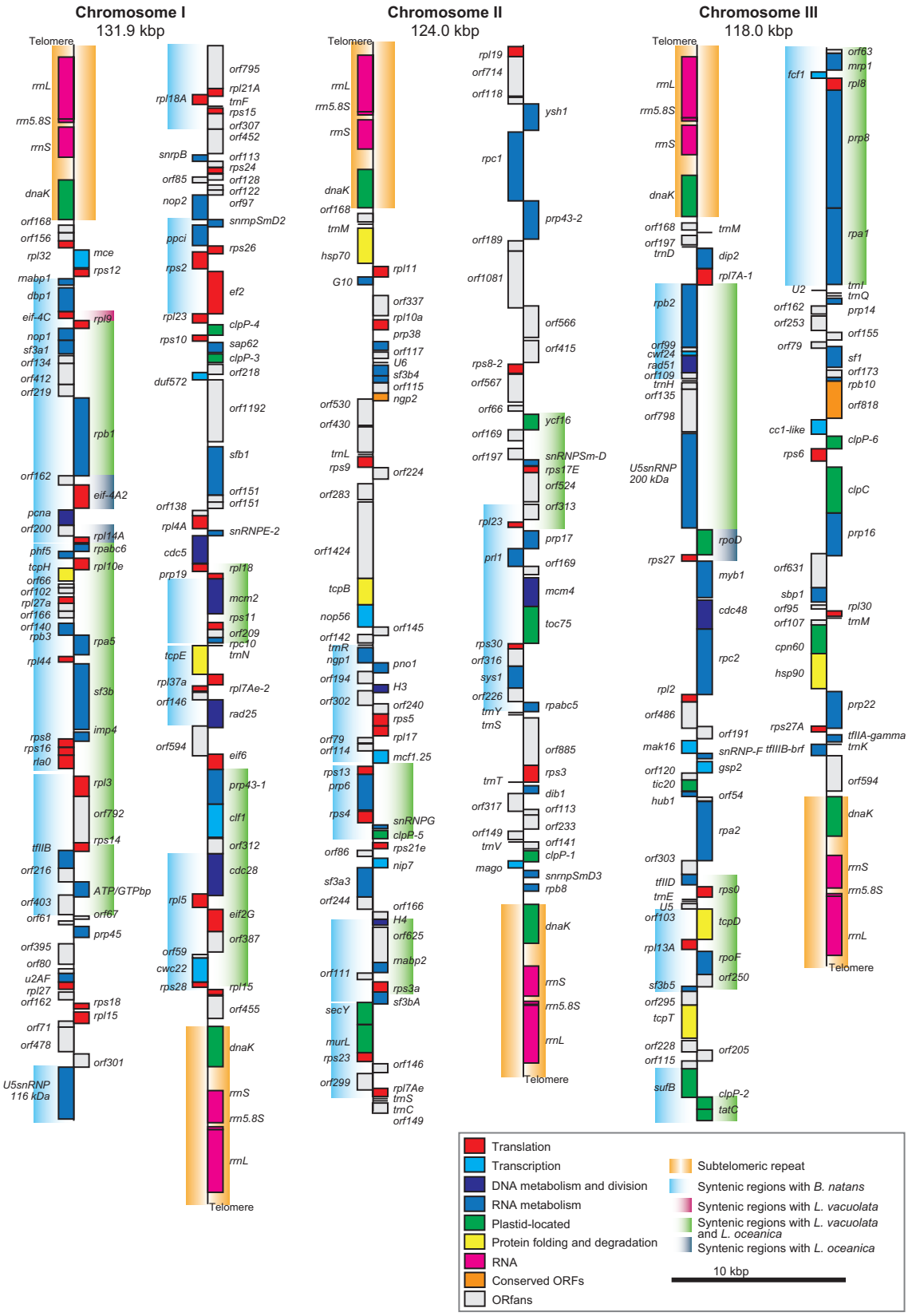

Chromosome I
131.9 kbp
Chromosome II
124.0 kbp
Chromosome III
118.0 kbp
Telomere
rrnL
rrn5.8S
rrnS
dnaK
orf168
orf156
rpl32
rnabp1
dbp1
eif-4C
nop1
sf3a1
orf134
orf412
orf219
orf162
pcna
orf200
phf5
tcpH
orf66
orf102
rpl27a
orf166
orf140
rpb3
rpl44
rps8
rps16
rla0
tfIIB
orf216
orf403
orf61
orf395
orf80
u2AF
rpl27
orf162
orf71
orf478
U5snRNP 116 kDa
mce
rps12
rpl9
rpb1
eif-4A2
rpl14A
rpabc6
rpl10e
rpa5
sf3b
imp4
rpl3
orf792
rps14
ATP/GTPbp
orf67
prp45
rps18
rpl15
orf301
rpl18A
orf795
rpl21A
trnF
rps15
orf307
orf452
snrpB
orf113
rps24
orf85
orf128
orf122
nop2
orf97
snrnpSmD2
ppci
rps26
rps2
ef2
rpl23
clpP-4
rps10
sap62
clpP-3
duf572
orf218
orf1192
sfb1
orf151
orf138
orf151
rpl4A
snRNPE-2
cdc5
prp19
rpl18
mcm2
rps11
orf209
rpc10
trnN
tcpE
rpl37a
orf146
rpl7Ae-2
rad25
orf594
eif6
prp43-1
clf1
orf312
cdc28
rpl5
eif2G
orf387
orf59
cwc22
rps28
rpl15
orf455
dnaK
rrnS
rrn5.8S
rrnL
Telomere
rrnL
rrn5.8S
rrnS
dnaK
orf168
tmM
hsp70
G10
orf530
orf430
trnL
rps9
orf283
orf1424
tcpB
nop56
orf142
tmR
ngp1
orf194
orf302
orf79
orf114
rps13
prp6
rps4
orf86
sf3a3
orf244
orf111
secY
murL
rps23
orf299
rpl11
orf337
rpl10a
prp38
orf117
U6
sf3b4
orf115
ngp2
orf224
orf145
pno1
H3
orf240
rps5
rpl17
mcf1.25
snRNPG
clpP-5
rps21e
nip7
orf166
H4
orf625
rnabp2
rps3a
sf3bA
orf146
rpl7Ae
trnS
trnC
orf149
rpl19
orf714
orf118
ysh1
rpc1
prp43-2
orf189
orf1081
orf566
orf415
rps8-2
orf567
orf66
orf169
orf197
ycf16
snRNPSm-D
rps17E
orf524
orf313
rpl23
prp17
prl1
orf169
mcm4
toc75
rps30
orf316
sys1
orf226
trnY
trnS
rpabc5
orf885
rps3
trnT
dib1
orf317
orf113
orf233
orf149
trnV
orf141
clpP-1
mago
snrnpSmD3
rpb8
dnaK
rrnS
rrn5.8S
rrnL
Telomere
rrnL
rrn5.8S
rrnS
dnaK
orf168
orf197
trnD
rpb2
orf99
cwf24
rad51
orf109
trnH
orf135
orf798
U5snRNP 200 kDa
rps27
rps27
orf486
mak16
orf120
tic20
hub1
orf303
tfIID
trnE
U5
orf103
rpl13A
sf3b5
orf295
tcpT
orf228
orf115
sufB
trnM
dip2
rpl7A-1
rpoD
myb1
cdc48
rpc2
orf191
snRNP-F
gsp2
orf54
rpa2
rps0
tcpD
rpoF
orf250
orf205
clpP-2
tatC
fcf1
orf63
mrp1
rpl8
prp8
rpa1
U2
trnI
trnQ
orf162
orf253
prp14
orf155
orf79
sf1
orf173
rpb10
orf818
cc1-like
clpP-6
rps6
clpC
prp16
orf631
sbp1
orf95
orf107
cpn60
hsp90
rpl30
trnM
prp22
rps27A
tfIIA-gamma
tfIIIB-brf
trnK
orf594
dnaK
rrnS
rrn5.8S
rrnL
Telomere
Translation
Transcription
DNA metabolism and division
RNA metabolism
Plastid-located
Protein folding and degradation
RNA
Conserved ORFs
ORfans
Subtelomeric repeat
Syntenic regions with B. natans
Syntenic regions with L. vacuolata
Syntenic regions with L. vacuolata and L. oceanica
Syntenic regions with L. oceanica
10 kbp

Figure 3.2. Nucleomorph genome map of the chlorarachniophyte *Lotharella vacuolata*.

The genome is comprised of three chromosomes, which are shown artificially broken at their midpoint. Genes indicated on the right side are transcribed from top to bottom, and the genes on the left side are transcribed in the opposite direction. Colors of gene blocks correspond to predicted functional categories in the box. Duplicated gene regions (green) and syntenic regions with *B. natans* (blue), *A. amoebiformis* (red), *L. oceanica* (gray), and both *B. natans* and *A. amoebiformis* (yellow) are shaded by color gradations.

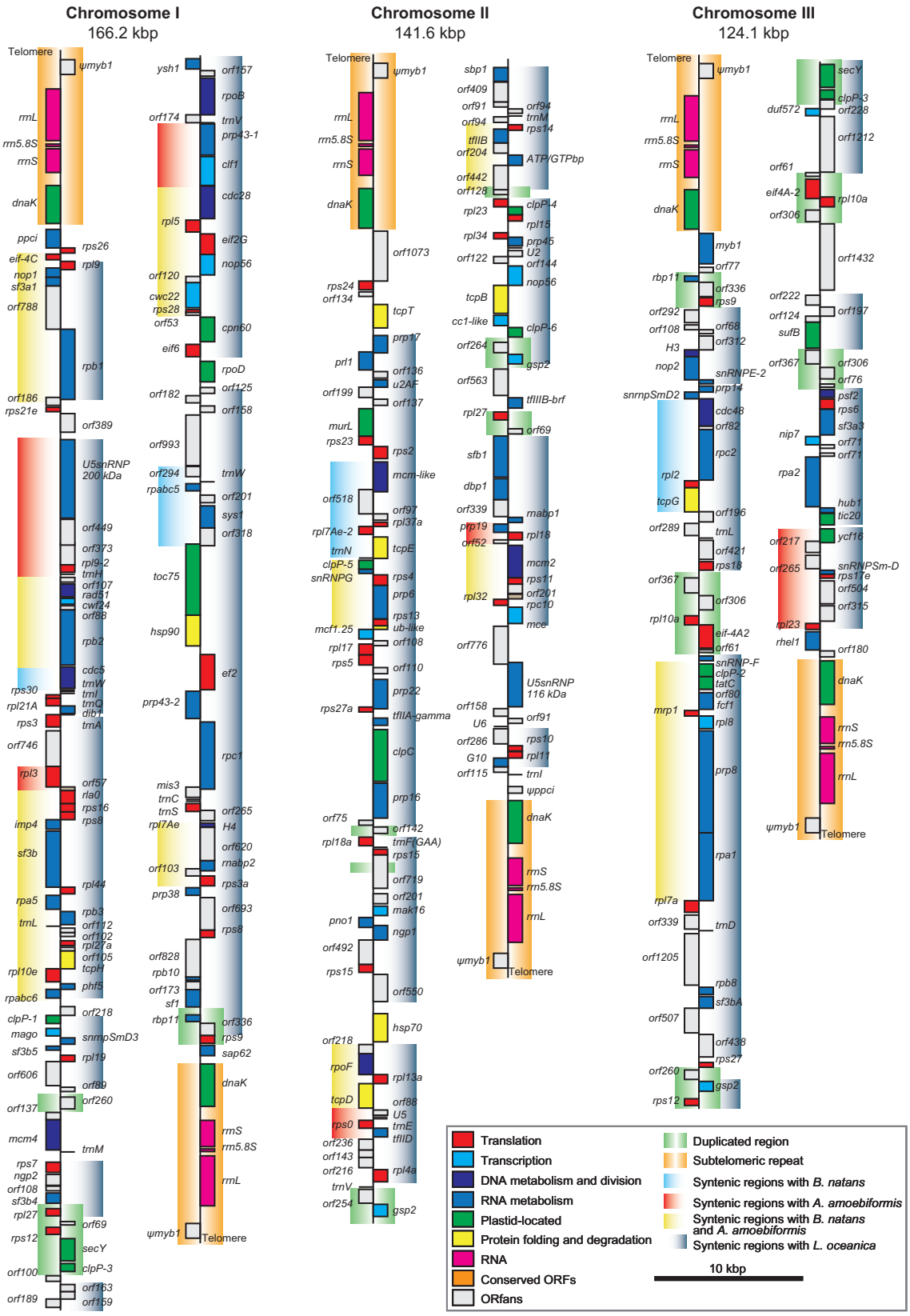

Chromosome I
166.2 kbp
Chromosome II
141.6 kbp
Chromosome III
124.1 kbp
Translation
Transcription
DNA metabolism and division
RNA metabolism
Plastid-located
Protein folding and degradation
RNA
Conserved ORFs
ORfans
Duplicated region
Subtelomeric repeat
Syntenic regions with B. natans
Syntenic regions with A. amoebiformis
Syntenic regions with B. natans and A. amoebiformis
Syntenic regions with L. oceanica
10 kbp

Figure 3.3. Comparison of gene content among nucleomorph genomes.

(a) Comparison of gene content among four chlorarachniophyte nucleomorph genomes. Venn diagrams indicate the number of shared and/or unique genes categorized as total protein-coding genes, function-predicted protein genes, and hypothetical protein genes (ORFans). (b) Comparison of conserved core genes between four chlorarachniophytes and four cryptophytes. Total 93 function-predicted genes are overlapped among eight nucleomorph genomes of chlorarachniophytes and cryptophytes. Light Venn diagrams show the number of shared and/or unique genes in each functional category.

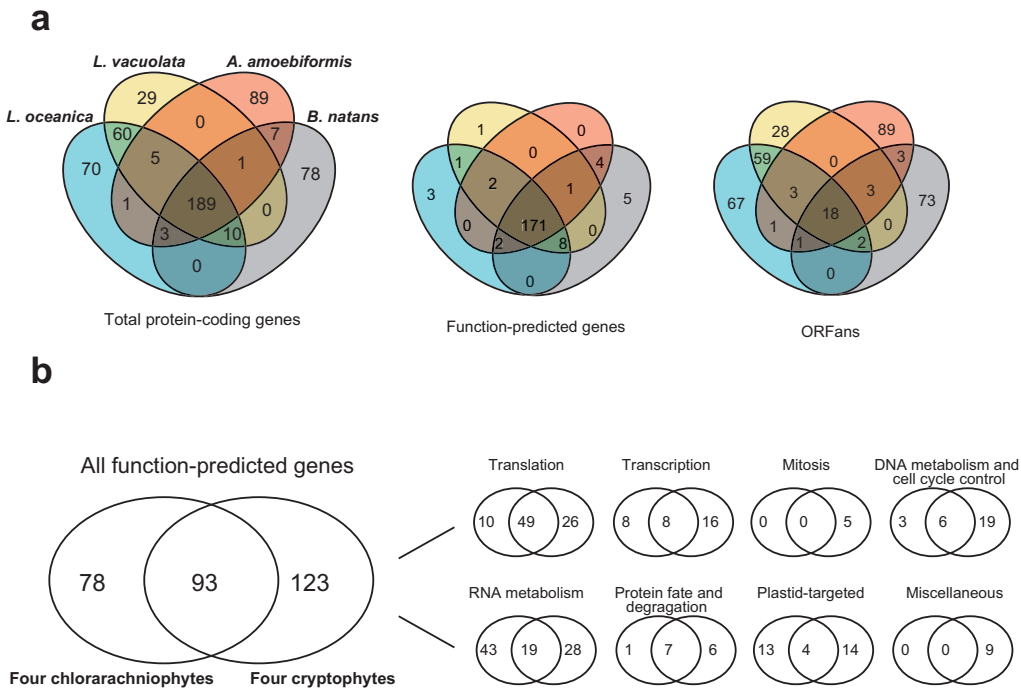

a
L. vacuolata
A. amoebiformis
L. oceanica
B. natans
29
89
0
60
7
5
1
70
78
1
189
0
3
10
0
Total protein-coding genes
1
0
1
0
4
2
1
3
5
0
171
0
2
8
0
Function-predicted genes
28
89
59
0
3
3
3
67
73
18
1
0
1
2
0
ORFans
b
All function-predicted genes
78
93
123
Four chlorarachniophytes
Four cryptophytes
Translation
10
49
26
Transcription
8
8
16
Mitosis
0
0
5
DNA metabolism and cell cycle control
3
6
19
RNA metabolism
43
19
28
Protein fate and degragation
1
7
6
Plastid-targeted
13
4
14
Miscellaneous
0
0
9

Figure 3.4. Size distribution of ultra-small introns in chlorarachniophyte nucleomorph genes.

The total number of introns in each size category is indicated by color bars: *B. natans* (pink), *A. amoebiformis* (red), *L. vacuolata* (blue), and *L. oceanica* (green).

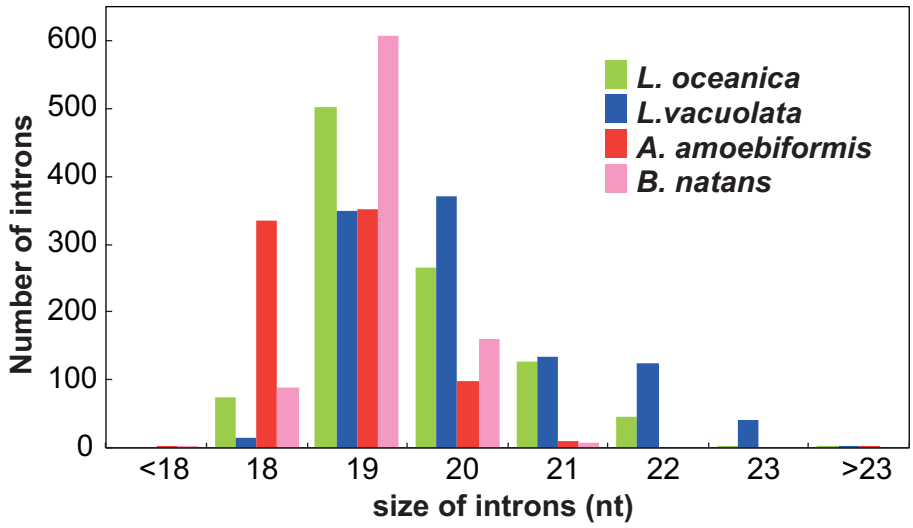
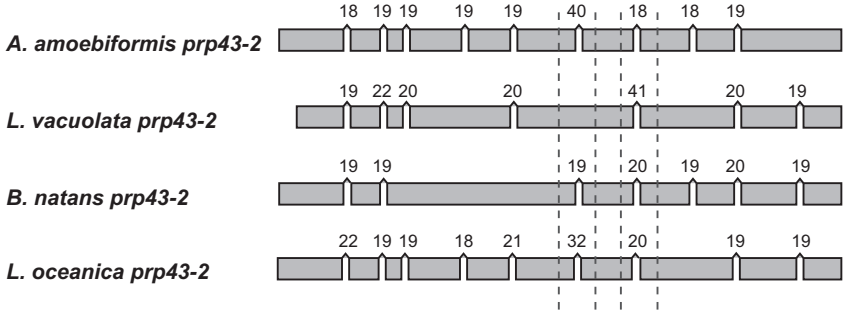

Number of introns
600
500
400
300
200
100
0
<18
18
19
20
21
22
23
>23
size of introns (nt)
L. oceanica
L.vacuolata
A. amoebiformis
B. natans

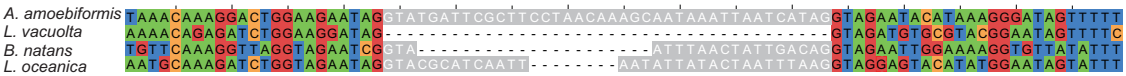
Figure 3.5. Comparison of intron positions in chlorarachniophyte *prp43-2* genes.

(a) Gray blocks show exons, and each number on the gaps indicates intron size (nt). Relatively longer introns of 40, 41, and 32 nt in *A. amoebiformis*, *L. vacuolata*, and *L. oceanica*, respectively, are shown. (b, c) Nucleotide alignments of longer introns containing domains. Gray nucleotides show introns.

a


18 19 19 19 19 40 18 18 19
A. amoebiformis prp43-2
19 22 20 20 41 20 19
L. vacuolata prp43-2
19 19 19 20 19 20 19
B. natans prp43-2
22 19 19 18 21 32 20 19 19
L. oceanica prp43-2

b


A. amoebiformis
L. vacuolta
B. natans
L. oceanica

c

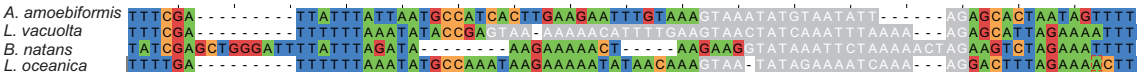

A. amoebiformis
L. vacuolta
B. natans
L. oceanica

Figure 3.6. Comparison of homologous gene positions on nucleomorph genomes in four chlorarachniophytes.

Colored blocks show three chromosomes in each nucleomorph genome: *B. natans* (green), *A. amoebiformis* (yellow), *L. vacuolata* (red), and *L. oceanica* (blue). The internal lines indicate paired homologous genes. Subtelomeric regions comprised of an rDNA operon are excluded from this comparative analysis.

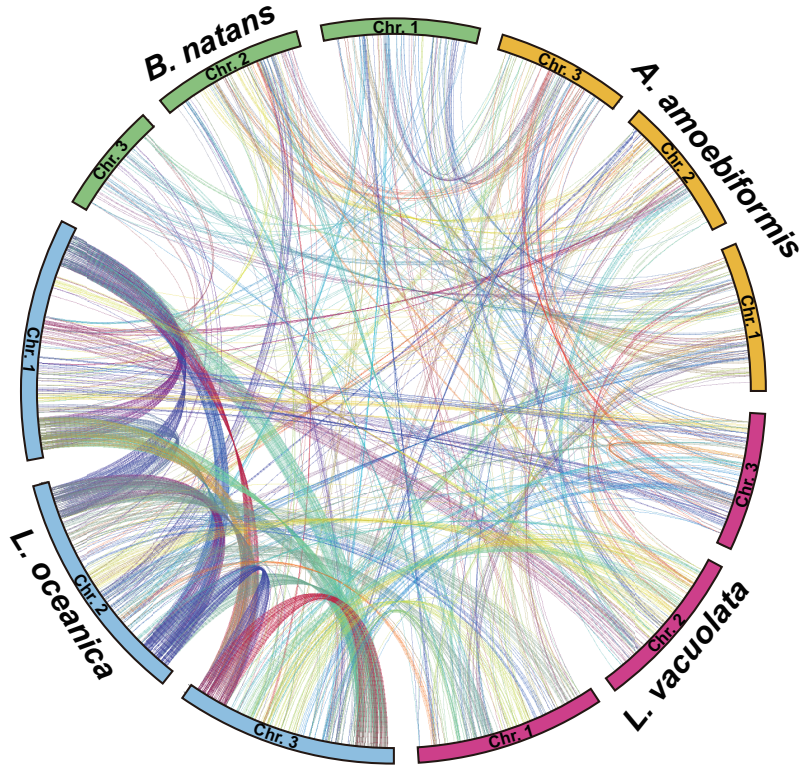

B. natans
Chr. 1
Chr. 2
Chr. 3
A. amoebiformis
Chr. 3
Chr. 2
Chr. 1
L. vacuolata
Chr. 3
Chr. 2
Chr. 1
L. oceanica
Chr. 3
Chr. 2
Chr. 1

Figure 3.7. Comparison of homologous gene positions on nucleomorph genomes in two *Lotharella* species.

Colored blocks show three chromosomes in each nucleomorph genome: *L. vacuolata* (green) and *L. oceanica* (yellow). The internal lines indicate paired homologous genes. Subtelomeric regions comprised of an rDNA operon are excluded from this comparative analysis.

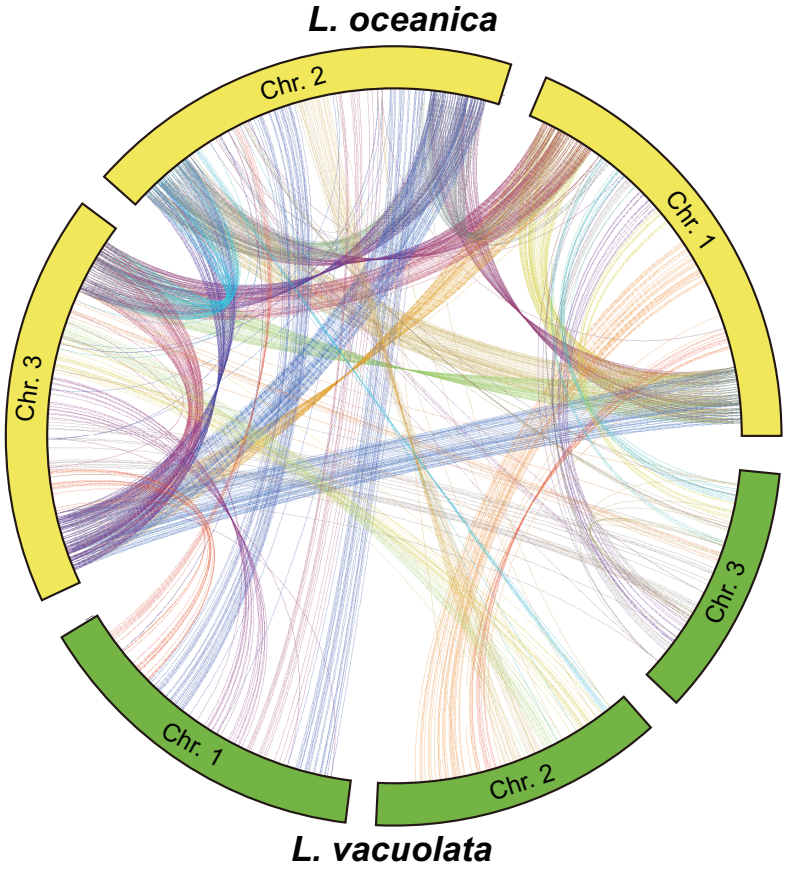

L. oceanica
Chr. 2
Chr. 1
Chr. 3
Chr. 3
Chr. 1
Chr. 2
L. vacuolata

Figure 3.8. Degraded ORFans in syntenic blocks of chlorarachniophyte nucleomorph genomes.

Functions of annotated genes and ORFan genes are shown in gray and black boxes, respectively. Gray highlights indicate syntenic positions between different nucleomorph genomes. Red highlights show the correspondence of syntenic ORFans and functional annotated genes that occupy the same syntenic positions. (a) The *mcm-like* gene of *B. natans* corresponds to an ORFan of *A. amoebiformis*. (b) *tcpG* genes of three chlorarachniophytes occupy the same syntenic position as that of an ORFan in *A. amoebiformis*. (c) The *L. oceanica tbl3* gene corresponds to an ORFan in *L. vacuolata*. (d) The *L. vacuolata nop56* gene corresponds to two ORFans in *L. oceanica*.

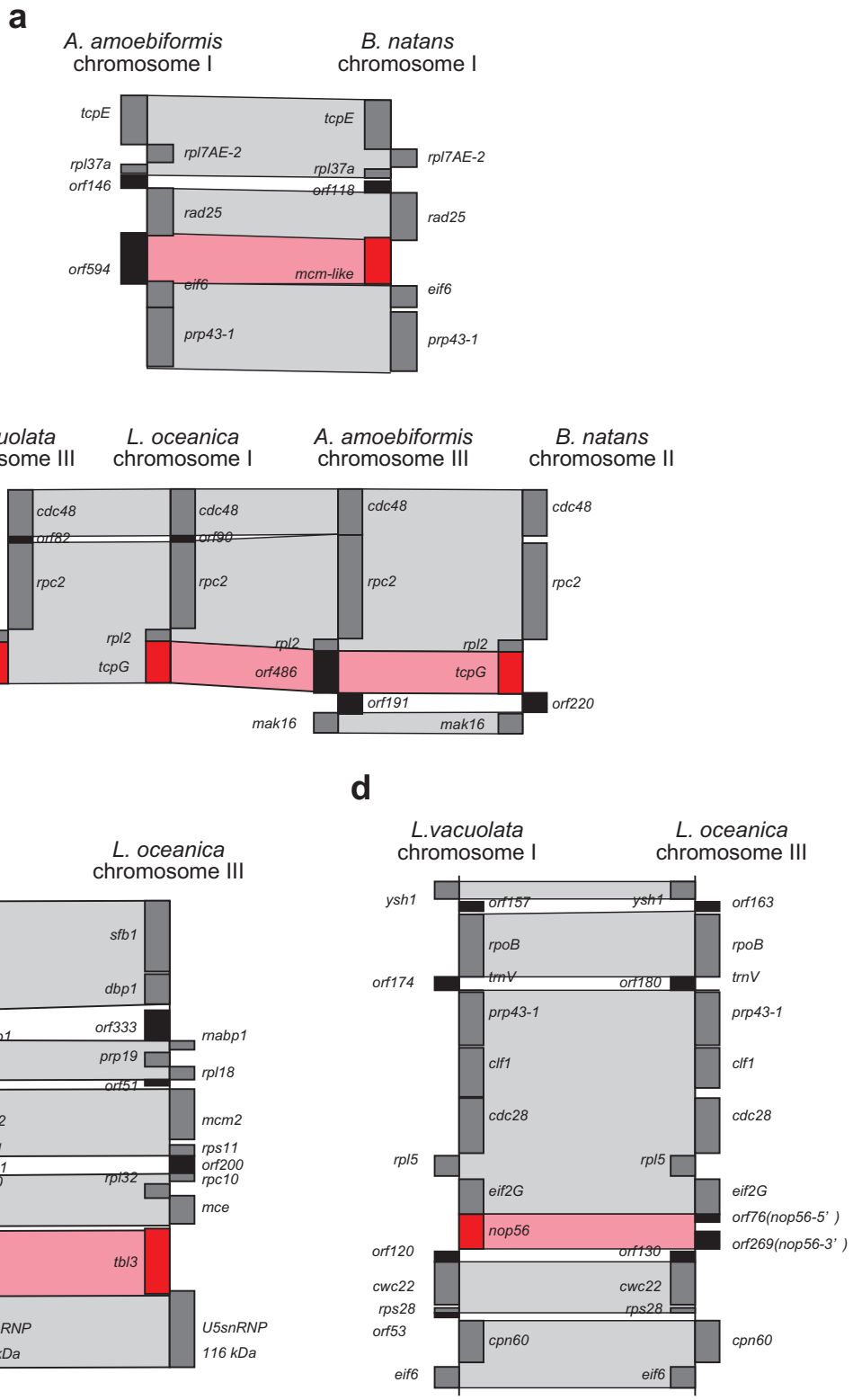

a
A. amoebiformis
chromosome I
B. natans
chromosome I
tcpE
rpl7AE-2
rpl37a
orf146
rad25
orf594
mcm-like
eif6
prp43-1
orf118
b
L.vacuolata
chromosome III
L. oceanica
chromosome I
A. amoebiformis
chromosome III
B. natans
chromosome II
cdc48
orf82
orf90
rpc2
rpl2
tcpG
orf486
orf191
mak16
orf220
c
L.vacuolata
chromosome II
L. oceanica
chromosome III
sfb1
dbp1
orf339
orf333
rnabp1
prp19
rpl18
orf52
orf51
mcm2
rps11
orf201
orf200
rpc10
rpl32
mce
orf776
tbl3
U5snRNP
116 kDa
d
L.vacuolata
chromosome I
L. oceanica
chromosome III
ysh1
orf157
orf163
rpoB
trnV
orf174
orf180
prp43-1
clf1
cdc28
rpl5
eif2G
nop56
orf76(nop56-5')
orf269(nop56-3')
orf120
orf130
cwc22
rps28
orf53
cpn60
eif6

Figure 4.1 Relative gene expression and evolutionary rates of nucleomorph-encoded proteins in chlorarachniophytes.

(a, b) Bars show relative mRNA transcript levels (RPKM: Reads Per Kilobase of exon per Million mapped reads) for all nucleomorph genes in the chlorarachniophyte *Amorphochlora amoebiformis* and *Bigelowiella natans*. (c) Relative transcription levels of nucleus- and nucleomorph-encoded heat-shock proteins and NmToc75estimated by real time quantitative PCR in a dark- (black bars) and a light-phase (white bars) culture. Error bars represent the SD of triplicate experiments. (d) Evolutionary rates of nuclear (N) and nucleomorph (Nm) ribosomal protein sequences among three chlorarachniophytes. The length of dotted lines indicates the evolutionary distances estimated by amino acid substitution rates (the numbers on lines).

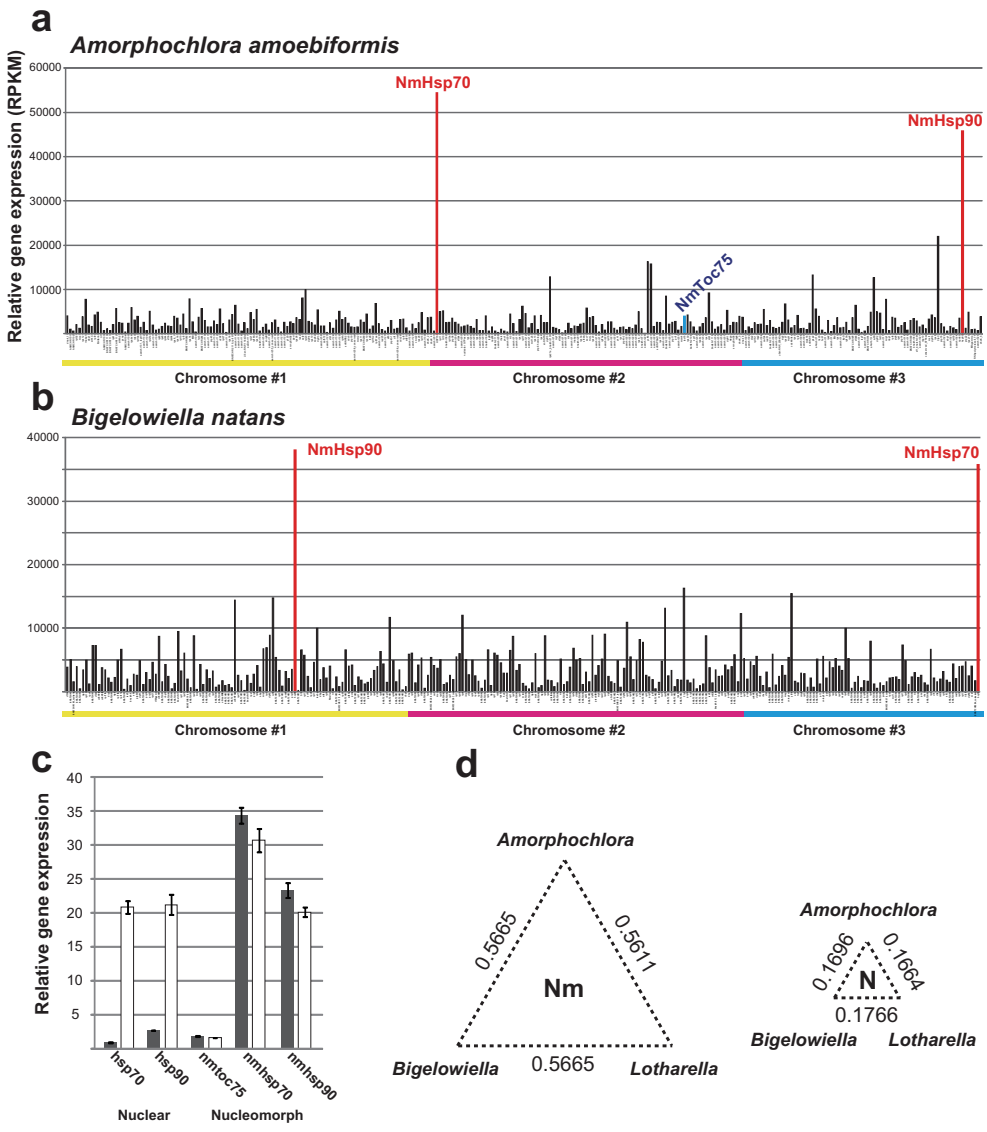

a
Amorphochlora amoebiformis
Relative gene expression (RPKM)
60000
50000
40000
30000
20000
10000
NmHsp70
NmHsp90
NmToc75
Chromosome #1
Chromosome #2
Chromosome #3
b
Bigelowiella natans
40000
30000
20000
10000
NmHsp90
NmHsp70
Chromosome #1
Chromosome #2
Chromosome #3
c
Relative gene expression
40
35
30
25
20
15
10
5
hsp70
hsp90
nmtoc75
nmhsp70
nmhsp90
Nuclear
Nucleomorph
d
Amorphochlora
0.5665
0.5611
Nm
Bigelowiella
0.5665
Lotharella
Amorphochlora
0.1696
0.1664
N
0.1766
Bigelowiella
Lotharella

Figure 4.2. Relative gene expression and evolutionary rates of nucleomorph-encoded proteins in cryptophytes.

(a, b, c) Bars show relative mRNA transcript levels (RPKM) for all nucleomorph genes in three cryptophyte species. (d) Evolutionary rates of nuclear (N) and nucleomorph (Nm) ribosomal protein sequences among three cryptophytes. The length of dotted lines indicates the evolutionary distances estimated by amino acid substitution rates (the numbers on lines).

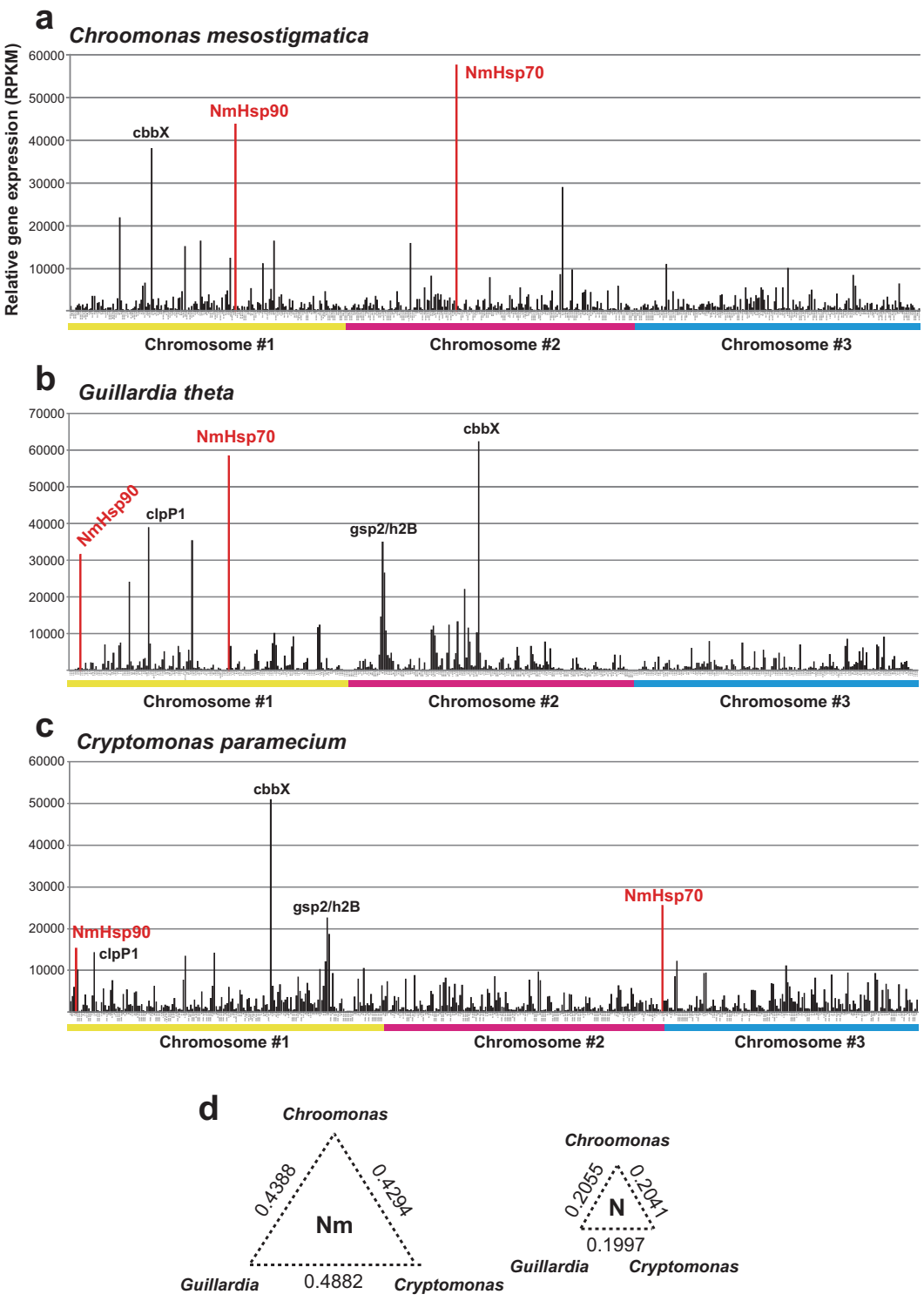

a
Chroomonas mesostigmatica
Relative gene expression (RPKM)
60000
50000
40000
30000
20000
10000
NmHsp70
NmHsp90
cbbX
Chromosome #1
Chromosome #2
Chromosome #3
b
Guillardia theta
70000
60000
50000
40000
30000
20000
10000
NmHsp70
cbbX
NmHsp90
clpP1
gsp2/h2B
Chromosome #1
Chromosome #2
Chromosome #3
c
Cryptomonas paramecium
60000
50000
40000
30000
20000
10000
cbbX
gsp2/h2B
NmHsp70
NmHsp90
clpP1
Chromosome #1
Chromosome #2
Chromosome #3
d
Chroomonas
0.4388
0.4294
Nm
Guillardia
0.4882
Cryptomonas
Chroomonas
0.2055
0.2041
N
0.1997
Guillardia
Cryptomonas

Figure 4.3. Correlation between relative transcription level and GC content of nucleomorph genes.

(a–e) Each bar indicates the GC content of nucleomorph genes ordered by their relative transcription levels, RPKM values, in three cryptophytes (orange) and two chlorarachniophytes (blue). Highly expressed genes reside in the left side of graphs. (f) Amino acid substitution rates of highly/lowly expressed genes (HEG/LEG) between homologs from two species of cryptophytes, *C. mesostigmatica* and *G. theta* (orange) and chlorarachniophytes, *A. amoebiformis* and *B. natans* (blue). Error bars indicate the SD of substitution rates from 10 HEG/LEG.

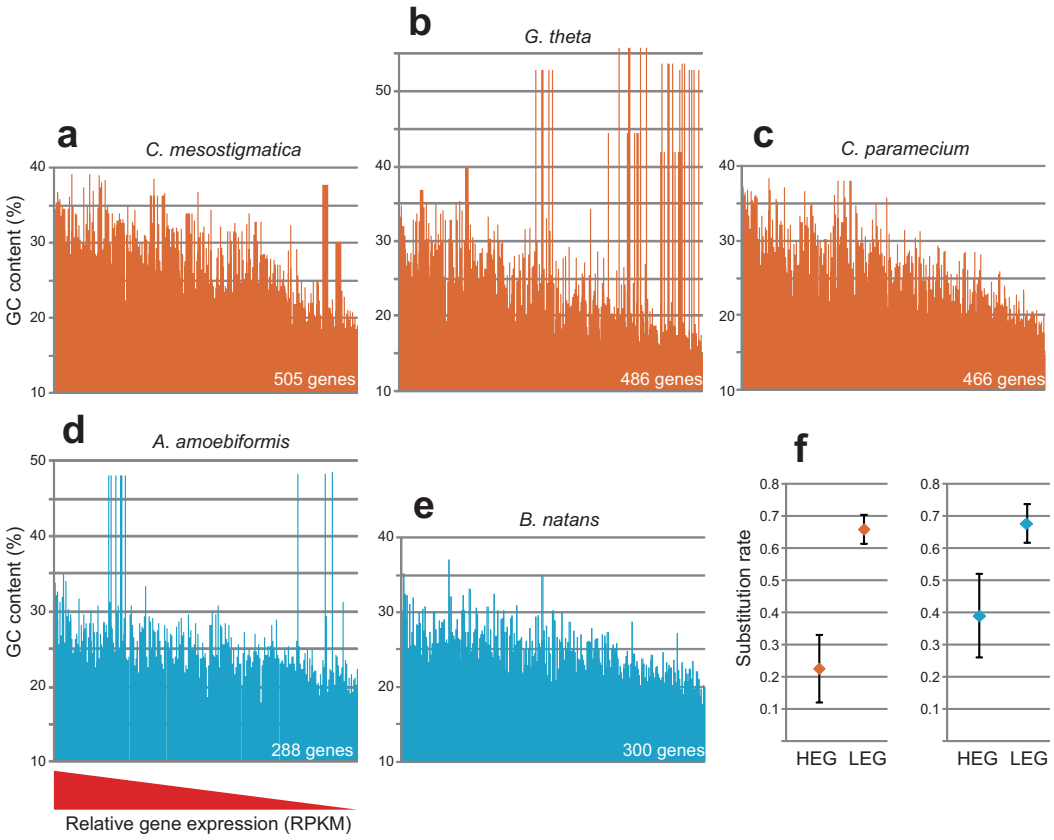
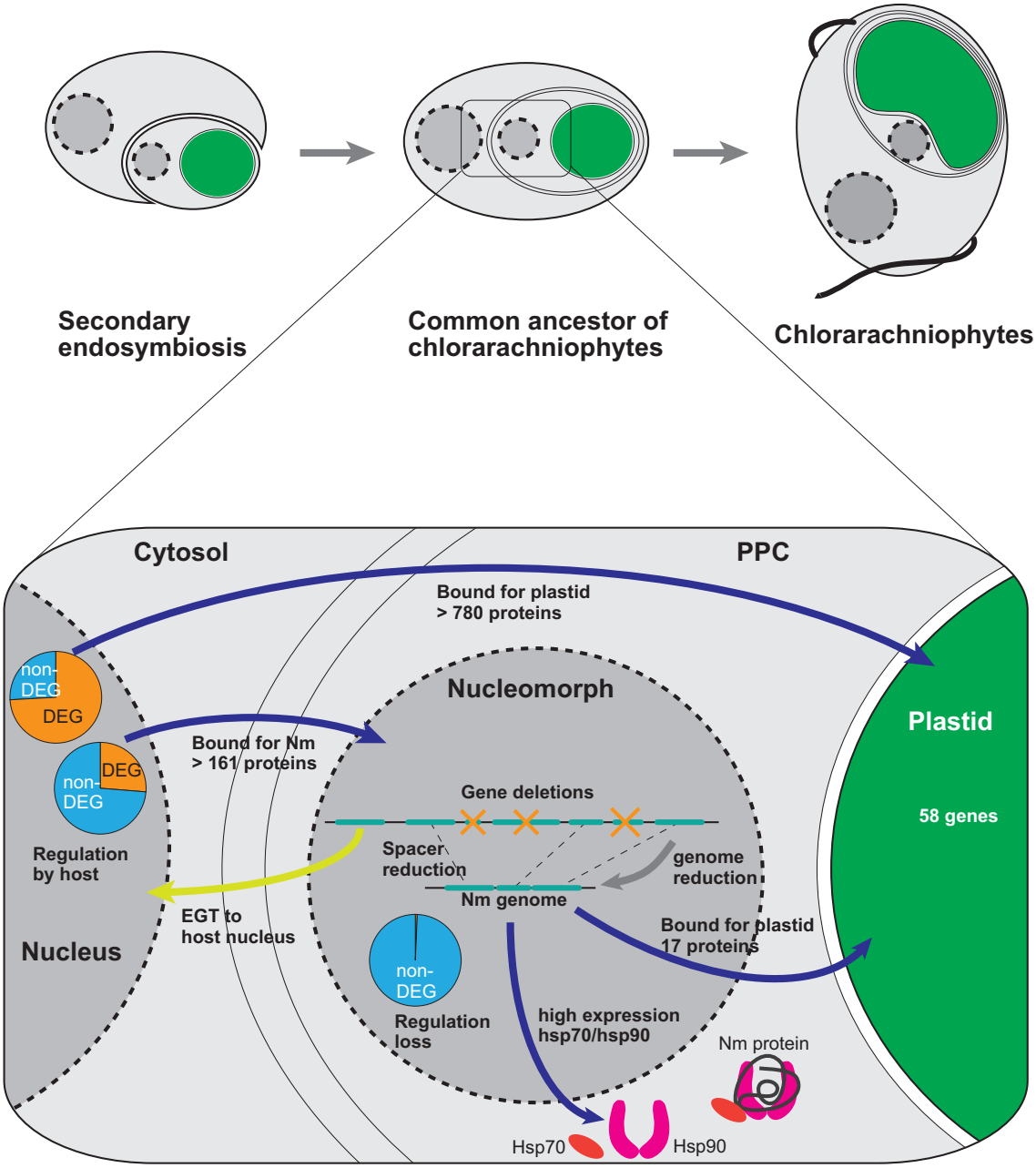

a
C. mesostigmatica
505 genes
b
G. theta
486 genes
c
C. parameciium
466 genes
d
A. amoebiformis
288 genes
e
B. natans
300 genes
f
GC content (%)
Substitution rate
HEG
LEG
Relative gene expression (RPKM)

Figure 6.1. Predicted processes of reductive evolution of endosymbiotically-derived genomes in chlorarachniophytes.

Chlorarachniophyte nucleomorph genomes underwent most of their reductive evolution prior to the radiation of extent members of the group. Following the reductive evolution, nucleomorph-encoded proteins lacked their gene expression patterns along the cell cycle. Nucleomorph- or plastid-targeted nuclear-encoded proteins acquired their gene expression patterns to control functions of the organelles.


Secondary endosymbiosis
Common ancestor of chlorarachniophytes
Chlorarachniophytes
Cytosol
PPC
Bound for plastid > 780 proteins
non-DEG
DEG
Bound for Nm > 161 proteins
Nucleomorph
Plastid
58 genes
Gene deletions
Regulation by host
Spacer reduction
genome reduction
Nm genome
EGT to host nucleus
Nucleus
Bound for plastid 17 proteins
Regulation loss
high expression hsp70/hsp90
Nm protein
Hsp70
Hsp90

8.2. Tables

Table 2.1 List of the plastid genomes that I used in the comparative analyses

Classification	Species	Accession NO.
Ulvophyceae	*Oltmannsiellopsis viridis*	DQ291132.1
	Pseudendoclonium akinetum	AY835431.1
	Ulva sp. UNA00071828	KP720616.1
	Bryopsis hypnoides	GQ892829.1
	Bryopsis plumosa	LN810504.1
	Tydemania expeditiones	LN810505.1
Chlorophyceae	*Acutodesmus obliquus*	DQ396875.1
	Chlamydomonas reinhardtii	BK000554.2
	Dunaliella salina	GQ250046.1
	*Volvox carteri**	GU084820.1
	Floydiella terrestris	GU196268.1
	Oedogonium cardiacum	EU677193.1
	Schizomeris leibleinii	HQ700713.1
	Stigeoclonium helveticum	DQ630521.1
Trebouxiophyceae	*Coccomyxa subellipsoidea*	HQ693844.1
	Leptosira terrestris	EF506945.1
	Oocystis solitaria	FJ968739.1
	Geminella minor	KM462883.1
	Gloeotilopsis sterilis	KM462877.1
	Planctonema lauterbornii	KM462880.1
	Ettlia pseudoalveolaris	KM462869.1
	Fusochloris perforata	KM462882.1
	Microthamnion kuetzingianum	KM462876.1
	Xylochloris irregularis	KM462872.1
	Neocystis brevis	KM462873.1
	Pabia signiensis	KM462866.1
	Koliella longiseta	KM462868.1
	Stichococcus bacillaris	KM462864.1
	Prasiolopsis sp. SAG 84.81	KM462862.1
	"Chlorella" mirabilis	KM462865.1
	Chlorosarcina brevispinosa	KM462875.1
	Myrmecia israelensis	KM462861.1
	Lobosphaera incisa	KM821265.1
	Dictyochloropsis reticulata	KM462860.1
	Watanabea reniformis	KM462863.1
	Botryococcus braunii	KM462884.1
	Choricystis minor	KM462878.1
	Elliptochloris bilobata	KM462887.1
	Trebouxiophyceae sp. MX-AZ01	JX402620.1
	Paradoxia multiseta	KM462879.1
Chlorodendrophyceae	*Tetraselmis subcordiformis**	GANN00000000.1
Pedinophyceae	*Pedinomonas minor*	FJ968740.1
	Pedinomonas tuberculata	KM462867.1
	Marsupiomonas sp. NIES 1824	KM462870.1
Chlorellales	*Chlorella vulgaris*	AB001684.1
	Chlorella variabilis	KJ718922.1
	Chlorella sorokiniana	KJ397925.1
	Chlorella sp. ArM0029B	KF554427.1
	Parachlorella kessleri	FJ968741.1
	Dicloster acuatus	KM462885.1
	Pseudochloris wilhelmii	KM462886.1
	Marvania geminata	KM462888.1
Prasinophytes	*Prasinoderma coloniale*	KJ746598.1
	Prasinophyceae sp. MBIC10622	KJ746602.1
	Prasinococcus sp. CCMP1194	KJ746597.1
	Pyramimonas parkeae	FJ493499.1
	Ostreococcus tauri	CR954199.2
	Monomastix sp. OKE-1	FJ493497.1
	Nephroselmis olivacea	AF137379.1
	Nephroselmis astigmatica	KJ746600.1
	Pycnococcus provasolii	FJ493498.1
	Picocystis salinarum	KJ746599.1

	Prasinophyceae sp. CCMP1205	KJ746601.1
Chlorarachniophytes	*Bigelowiella natans*	DQ851108.1
	Lotharella oceanica	KF438023.1

*used only by phylogenetic analyses

Table 2.2 General features of the plastid genomes of four chlorarachniophytes

Features	*Gymnochlora stellata*	*Lotharella vacuolata*	*Partenskyella glossopodia*	*Bigelowiella natans**	*Lotharella oceanica***
Genome size (bp)					
Total	67,451	71,557	72,620	69,166	70,997
LSC region (number of genes)	50,468 (70)	51,319 (69)	48,607 (68)	46,186 (67)	51,500(70)
SSC region (number of genes)	4,367 (9)	6,697 (9)	4,567 (10)	4,124 (9)	6,801(10)
IR (number of genes)	6,308 (9)	6,770 (8)	9,723 (10)	9,387 (11)	6,348(7)
(A + T)%	70.2	70.2	71.9	70.5	69.4
Number of genes***					
Total	97	94	98	98	94
Protein-coding	60	59	61	61	59
tRNA	31	29	31	31	29
rRNA	6	6	6	6	6
Number of Introns					
Group I	0	1	1	0	1
Group II	2	2	2	0	2
Average intergenic length (bp)	89.6	107.9	107.9	90.7	112.7

*Rogers et al. 2007

**Tanifuji et al. 2014, updated.

***including duplicated genes

Table 2.3 Conserved genes in the plastid genomes of chlorarachniophytes

Function	Conserved genes
ATP synthase	*atpA*, *atpB*, *atpE*, *atpF*, *atpH*, *atpI*
Cytochrome	*petA*, *petB*, *petD*, *petG*
Photosystem I	*psaA*, *psaB*, *psaC*, *psaJ*
Photosystem II	*psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbH*, *psbI*, *psbJ*, *psbK*, *psbL*, *psbM*, *psbN*, *psbT*
RubisCO	*rbcL*
LSU ribosomal proteins	*rpl14*, *rpl16*, *rpl19*, *rpl20*, *rpl2*, *rpl23*, *rpl36, rpl5*
SSU ribosomal proteins	*rps11*, *rps12*, *rps14*, *rps18*, *rps19*, *rps2*, *rps3*, *rps4*, *rps7*, *rps8*, *rps9*
RNA polymearse	*rpoA*, *rpoB*, *rpoC1*, *rpoC2*
Miscellaneous	*chlI*, *clpP*, *tufA*, *ycf1*, *ycf3*, *ycf4*, *maturase-like*[a]
rRNAs	*rrnS*, *rrnL*, *rrn5S*
tRNAs	*trnA* (UGC), *trnC* (GCA), *trnD* (GUC), *trnE* (UUC), *trnF* (GAA), *trnG* (GCC)[b], *trnG* (UCC), *trnH* (GUG)[b], *trnI* (GAU), *trnK* (UUU), *trnL* (CAA)[c], *trnL* (UAA), *trnL* (UAG), *trnM* (CAU), *trnN* (GUU), *trnP* (UGG), *trnQ* (UUG), *trnR* (ACG), *trnR* (UCU), *trnS* (GCU), *trnT* (UGU), *trnV* (UAC), *trnW* (CCA), *trnY* (GUA)

[a]lacking in the plastid genome of *B. natans*.

[b]lacking in the plastid genome of *L. vacuolata*.

[c]lacking in the plastid genome of *L. oceanica*.

Table 3.1 Genome features of nucleomorph genomes in chlorarachniophytes

	A. amoebiformis	*L. vacuolata*	*B. natans**	*L. oceanica***
Genome size (bp)	373,958	431,876	372,879	~611,658
Chr. 1	131,920	166,173	140,598	~210,000
Chr. 2	124,024	141,647	134,144	207,543
Chr. 3	118,014	124,056	98,137	194,115
GC content (%)	30.0	24.7	28.5	33.0
Number of genes	340	359	332	636
Protein-coding (including duplicates)	295 (300)	294 (319)	288 (288)	338(596)
rRNAs	3 (18)	3 (18)	3 (18)	3(18)
tRNAs	21	19	22	19
snRNAs	3	3	4	3
Introns (Introns/genes)	793 (2.6)	1028 (3.2)	865 (3.0)	1021(1.6)
Gene density (genes/kbp)	0.91	0.83	0.89	1.04

*They were updated from the original article (Gilson et al. 2006).

**They were updated from the original article (Tanifuji et al. 2014).

Table 3.2 Number of genes encoded by nucleomorph genomes of chlorarachniophytes

	Chlorarachniophytes			
Gene symbol	***Lo***	***Lv***	***Bn***	***Aa***
Translation				
rla0	1	1	1	1
rpl10	1	1	1	1
rpl10A	2	2	1	1
rpl11	1	1	1	1
rpl13A	1	1	1	1
rpl14A	1		1	1
rpl15	1	1	1	1
rpl17	1	1	1	1
rpl18	1	1	1	1
rpl18a	1	1	1	1
rpl19	1	1	1	1
rpl2	1	1	1	1
rpl21A	1	1	1	1
rpl23	1	1	1	1
rpl23	1	1	1	1
rpl24			1	
rpl27	2	2	1	1
rpl27a	1	1	1	1
rpl3	1	1	1	1
rpl30			1	1
rpl32	1	1	1	1
rpl34	1	1	1	
rpl37a	1	1	1	1
rpl44	1	1	1	1
rpl4A	1	1	1	1
rpl5	1	1	1	1
rpl7	1	1	1	1
rpl7AE-1	1	1	1	1
rpl7AE-2	1	1	1	1
rpl8	1	1	1	1
rpl9	1	1	1	1
rpl9-2		1		
rps0	1	1	1	1
rps10	1	1	1	1
rps11	1	1	1	1
rps12	1	2	1	1
rps13	1	1	1	1
rps14	1	1	1	1
rps15	1	1	1	1
rps15A	1	1	1	1
rps16	1	1	1	1
rps17E	1	1	1	1
rps18	1	1	1	1
rps2	1	1	1	1
rps21e	1	1	1	1
rps23	1	1	1	1
rps24	1	1		1
rps26	1	1	1	1
rps27	1	1	1	1
rps27a	1	1	1	1
rps28	1	1	1	1
rps3	1	1	1	1
rps30	1	1	1	1
rps3a	1	1	1	1
rps4	1	1	1	1
rps5	1	1	1	1
rps6	1	1	1	1
rps7	1	1	1	1
rps8	1	1	1	1
rps8	1	1	1	1
rps9	2	2	1	1
ef2	1	1	1	1
eif-4A2	2	2	1	1
eif2G	1	1	1	1
eif-4C		1	1	1
eif6	1	1	1	1
Transcription				
clf1	1	1	1	1
cwc22	1	1	1	1
cwf24	1	1	1	1

duf572	1	1	1	1
gsp2	3	3	1	1
imp4	1	1	1	1
mago	1	1	1	1
mak16	1	1	1	1
mce	1	1	1	1
mcf1.25	1	1	1	1
cc1-like	1	1	1	1
fcf1	1	1	1	1
nip7	1	1	1	1
nop1	1	1	1	1
nop2	1	1	1	1
nop56	1	2	1	1
DNA metabolism and cell cycle control				
cdc28	1	1	1	1
cdc48	1	1	1	1
cdc5	1	1	1	1
H3	1	1	1	1
H4	1	1	1	1
mcm-like	1	1	1	
mcm2	1	1	1	1
mcm4	1	1	1	1
pcna			1	1
psf2	1	1		
rad25			1	1
rad51	1	1	1	1
rpoB	1	1	1	
rpoF	1	1	1	1
RNA metabolism				
dbp1	1	1	1	1
dib1	1	1	1	1
dip2			1	1
ATP/GTPbp	1	1	1	1
G10	1	1	1	1
hub1	1	1	1	1
mis3	1	1	1	
mrp1	1	1	1	1
myb	1	1	1	1
ppci	1	1	1	1
prl1	1	1	1	1
prp14	1	1	1	1
prp16	5	1	1	1
prp17	1	1	1	1
prp19	1	1	1	1
prp22	1	1	1	1
prp38	1	1	1	1
prp43-1	1	1	1	1
prp43-2	1	1	1	1
prp45	1	1	1	1
prp6	1	1	1	1
prp8	1	1	1	1
rbp11	2	2	1	
rnabp1	1	1	1	1
rnabp2	1	1	1	1
rpa1	1	1	1	1
rpa2	1	1	1	1
rpa5	1	1	1	1
rpabc5	1	1	1	1
rpabc6	1	1	1	1
rpb1	1	1	1	1
rpb10	1	1	1	1
rpb2	1	1	1	1
rpb3	1	1	1	1
rpb8	1	1	1	1
rpc1	1	1	1	1
rpc10	1	1	1	1
rpc2	1	1	1	1
phf5	1	1	1	1
rfc4			1	
rhel1	1	1	1	
ngp1	1	1	1	1
ngp2	1	1	1	1
sap62	1	1	1	1
sbp1	1	1	1	1
sf1	1	1	1	1
sf3a1	1	1	1	1
sf3a3	1	1	1	1

sf3b	1	1	1	1
sf3b4	1	1	1	1
sf3b5	1	1	1	1
sf3bA	1	1	1	1
sfb1	1	1	1	1
snRNP-F	1	1	1	1
snRNPE-1			1	
snRNPE-2	1	1	1	1
snRNPG	1	1	1	1
snRNPSm-D	1	1	1	1
snrnpSmD2	1	1	1	1
snrnpSmD3	1	1	1	1
snrpB	1		1	1
sys1	1	1	1	1
tfIIA-gamma	1	1	1	1
tfIIB	1	1	1	1
tfIId	1	1	1	1
tfIIIB-brf	1	1		1
u2AF	1	1	1	1
U5snRNP_116kDa	1	1	1	1
U5snRNP_200kDa	1	1	1	1
ysh1	1	1	1	1
Protein fate and degradation				
hsp70	8	1	1	1
hsp90	1	1	1	1
pno1	1	1	1	1
prsA6			1	
tcpB	1	1	1	1
tcpD	1	1	1	1
tcpE	1	1	1	1
tcpG	1	1	1	
tcpH	1	1	1	1
tcpT	1	1	1	1
ub2			1	
Ub-like	1	1	1	
Plastid-targeted				
clpC	5	1	1	1
clpP-1	1	1	1	1
clpP-2	1	1	1	1
clpP-3	1	2	1	1
clpP-4	1	1	1	1
clpP-5	1	1	1	1
clpP-6	1	1	1	1
cpn60	1	1	1	1
dnaK	7	6	1	6
murL	1	1	1	1
secY	1	2	1	1
rpoD	1	1	1	1
sufB	1	1	1	1
tatC	1	1	1	1
tic20	1	1	1	1
toc75	1	1	1	1
ycf16	1	1	1	1
Hypothetical (only shared ORFs by all chlorarachniophytes)				
-	1 (orf450)	1 (orf788)	1 (orf489)	1 (orf412)
-	2 (orf328)	2 (orf336)	1 (orf324)	1 (orf317)
-	1 (orf1420)	1 (orf1432)	1 (orf1433)	1 (orf1424)
-	1 (orf221)	1 (orf218)	1 (orf229)	1 (orf250)
-	1 (orf729)	1 (orf719)	1 (orf709)	1 (orf307)
-	1 (orf326)	1 (orf315)	1 (orf248)	1 (orf313)
-	1 (orf175)	1 (orf2318)	1 (orf303)	1 (orf316)
-	1 (orf106)	1 (orf107)	1 (orf82)	1 (orf109)
-	1 (orf86)	1 (orf88)	1 (orf90)	1 (orf99)
-	1 (orf699)	1 (orf746)	1 (orf784)	1 (orf792)
-	1 (orf400)	1 (orf442)	1 (orf403)	1 (orf403)
-	1 (orf200)	1 (orf201)	1 (orf199)	1 (orf209)
-	1 (orf103)	1 (orf103)	1 (orf102)	1 (orf111)
-	1 (orf158)	1 (orf158)	1 (orf107)	1 (orf155)
-	1 (orf333)	1 (orf339)	1 (orf339)	1 (orf337)
-	1 (orf695)	1 (orf693)	1 (orf576)	1 (orf567)
-	1 (orf195)	1 (orf197)	1 (orf199)	1 (orf205)
-	1 (orf833)	1 (orf828)	1 (orf822)	1 (orf818)

Aa: A. amoebiformis, *Bn: B. natans*, *Lo: L. oceanica*, *Lv: L. vacuolata*,

Table 3.3 Genome size variation and its factors

	A. amoebiformis	*L. vacuolata*	*B. natans*	*L. oceanica*
Genome size (kbp)	374.0	431.9	372.9	~611.7
Average size of shared proteins (aa)	346.8	355.7	347.3	351.0
Intergenic length (bp)*	86.6	130.4	112.5	163.0
Size of duplicated region (kbp)	63.6	109.2	52.4	282.7
internal regions	0.0	37.8	0.0	54.5
subtelomeric regions	63.6	71.4	52.4	228.2

*Variations are significantly shown by ANOVA ($p < 0.01$)

Table 3.4 Number of introns of highly conservative genes in chlorarachniophytes

	4 species	3 species	2 species	1 species	Number of positions
Lo-orf400/Lv-orf442/Bn-orf403/Aa-orf403	0	4	1	1	6
cdc28	3	0	0	2	5
clf1	4	1	2	1	8
clpC	0	0	0	0	0
clpP-2	2	1	0	0	3
clpP-3	0	0	0	1	1
cpn60	0	0	0	0	0
dbp1	1	0	1	0	2
dnaK	0	0	0	1	1
ef2	0	0	1	1	2
eif-4A2	5	3	2	0	10
eif2G	1	2	3	0	6
ATP/GTPbp	3	0	1	0	4
nop1	0	0	0	0	0
gsp2	1	3	1	0	5
HSP70	0	0	0	0	0
hsp90	0	0	0	0	0
mrp1	2	1	1	0	4
murL	2	5	1	1	9
myb1	4	3	1	0	8
prp16	7	1	0	0	8
prp22	1	5	1	2	9
prp43-1	7	2	2	3	14
prp43-2	5	3	2	0	10
prp8	0	0	3	0	3
rad51	3	1	1	1	6
rpa2	1	10	0	3	14
rpabc5	2	1	1	1	5
rpb1	0	3	1	0	4
rpb2	4	5	2	0	11
rpc2	4	4	1	4	13
rpl10	2	1	1	0	4
rpl10e	2	2	0	0	4
rpl23	0	1	2	0	3
rpl3	3	0	2	0	5
rpl4A	0	1	1	0	2
rpl8	4	1	0	0	5
rps13	0	3	0	0	3
rps2	2	0	0	0	2
rps23	1	3	0	0	4
rps5	3	2	0	1	6
rps9	2	0	0	1	3
sf3b	1	10	3	6	20
sf3b4	3	1	1	0	5
sfb1	6	2	0	0	8
rpoD	4	1	3	0	8
sufB	0	0	0	1	1
sys1	4	1	0	1	6
tatC	3	1	0	0	4
tfIId	3	2	0	1	6
toc75	0	0	0	3	3
u2AF	1	2	0	0	3
U5snRNP 200kDa	0	0	0	1	1
U5snRNP116kDa	2	4	2	1	9

ycf16	3	0	0	1	4
Total	**111**	**96**	**44**	**39**	**290**
%	**38.3**	**33.1**	**15.2**	**13.4**	

Table 3.5 Average number of homologous genes in syntenic blocks

	B. natans	***A. amoebiformis***	***L. vacuolata***	***L. oceanica***
B. natans	-	6.2(n=17)	5.9(n=13)	5.8(n=14)
A. amoebiformis		-	6.5(n=11)	6.9(n=11)
L. vacuolata			-	11.5(n=21)
L. oceanica				-

Table 4.1 Relative expression of nucleomorph genes in *Amorphochlora amoebiformis*

Chromosome No.	Gene name	Number of reads	Mapped by some reads (bp)	Gene length (bp)	Mapping rate	RPKM	GC content (%)
AaNm1	dnak_1	11906	2366	2366	1.00	3978.8	48.2
AaNm1	AAMCHR101	650	506	506	1.00	1015.7	48.3
AaNm1	AAMCHR102	353	505	508	0.99	549.4	31.2
AaNm1	rpl32	1000	367	392	0.94	2017.1	24.3
AaNm1	mce	1857	1165	1193	0.98	1230.8	23.4
AaNm1	rps12	2435	510	510	1.00	3775.1	24.9
AaNm1	rnabp1	3748	384	384	1.00	7717.4	28.7
AaNm1	dbp1	3730	1539	1539	1.00	1916.3	28.2
AaNm1	eif-4C	813	392	392	1.00	1639.9	25.6
AaNm1	rpl9	3038	563	563	1.00	4266.6	28.7
AaNm1	fibrillarin	4374	719	719	1.00	4810.1	28.3
AaNm1	trf	2420	752	752	1.00	2544.5	23.8
AaNm1	AAMCHR103	378	403	422	0.95	708.2	20.2
AaNm1	BNATCHR1110	1865	1269	1277	0.99	1154.8	25.2
AaNm1	BNATCHR1111	574	600	696	0.86	652.1	22.3
AaNm1	rpb1	11805	4486	4510	0.99	2069.6	28.3
AaNm1	BNATCHR1112	3761	526	526	1.00	5653.5	25.6
AaNm1	eif-4A2	4010	1265	1265	1.00	2506.4	28.0
AaNm1	pcna	2692	880	894	0.98	2380.9	26.7
AaNm1	AAMCHR104	280	602	638	0.94	347.0	21.1
AaNm1	aamchr105	1027	344	344	1.00	2360.6	22.9
AaNm1	rpabc6	2785	375	375	1.00	5872.2	26.3
AaNm1	rds3	1605	409	409	1.00	3102.8	33.3
AaNm1	rpl10e	3574	717	717	1.00	3941.3	27.9
AaNm1	aamchr131_tcpH?	1394	798	798	1.00	1381.2	24.3
AaNm1	aamchr132	654	200	200	1.00	2585.5	25.9
AaNm1	aamchr106	283	342	345	0.99	648.6	22.0
AaNm1	rpl27a	3373	524	524	1.00	5089.7	28.5
AaNm1	aamchr107	1262	519	520	1.00	1918.9	22.6
AaNm1	aamchr108	426	439	440	1.00	765.5	20.3
AaNm1	rpb3	868	643	687	0.94	999.0	25.4
AaNm1	rpa5	2276	1117	1134	0.99	1586.9	27.2
AaNm1	rpl44	691	217	237	0.92	2305.3	28.4
AaNm1	sf3b	8162	3638	3677	0.99	1755.1	26.6
AaNm1	IMP4	1259	618	618	1.00	1610.8	24.4
AaNm1	rps8_1	2442	500	500	1.00	3861.7	24.7
AaNm1	rps16	2234	499	509	0.98	3470.3	28.3
AaNm1	rpl10	2004	818	818	1.00	1937.1	23.5
AaNm1	rpl3	6751	1182	1205	0.98	4429.8	30.8
AaNm1	BNATCHR282	3656	2560	2567	1.00	1126.1	23.1
AaNm1	rps14	4677	473	473	1.00	7818.2	29.3
AaNm1	tfIIB	3563	1048	1048	1.00	2688.2	26.1
AaNm1	aamchr109	348	672	687	0.98	400.5	21.7
AaNm1	fet5	3861	817	817	1.00	3736.6	24.9
AaNm1	BNATCHR284	8744	1211	1211	1.00	5709.1	27.1
AaNm1	aamchr110	764	203	203	1.00	2975.8	25.0
AaNm1	aamchr111	356	185	185	1.00	1521.5	22.6
AaNm1	prp45	1247	669	669	1.00	1473.8	25.6
AaNm1	aamchr133	4384	1227	1227	1.00	2825.1	27.3
AaNm1	aamchr112	787	299	299	1.00	2081.2	23.9
AaNm1	u2AF	4466	630	630	1.00	5605.1	31.8
AaNm1	rpl27	1300	465	465	1.00	2210.5	25.9
AaNm1	aamchr113	149	485	512	0.95	230.1	21.2
AaNm1	rps18	1348	342	342	1.00	3116.5	29.3
AaNm1	BNATCHR2119_rpl15?	3825	705	706	1.00	4283.8	25.6
AaNm1	aamchr114	1729	215	215	1.00	6358.6	26.4
AaNm1	aamchr115	4119	1468	1512	0.97	2154.0	23.8
AaNm1	aamchr116	717	887	943	0.94	601.2	22.0
AaNm1	U5snRNP_116kDa	8729	2782	2802	0.99	2463.2	26.7
AaNm1	aamchr117	3440	2517	2518	1.00	1080.2	22.2

AaNm1	rpl21	2399	401	401	1.00	4730.3	22.7
AaNm1	rpl18A	2118	520	520	1.00	3220.5	21.9
AaNm1	rps15	2785	401	401	1.00	5491.4	28.1
AaNm1	bnatchr194	708	997	998	1.00	560.9	22.5
AaNm1	aamchr234	2463	1412	1414	1.00	1377.3	25.3
AaNm1	snrpB	783	292	292	1.00	2120.2	24.4
AaNm1	aamchr118	349	338	359	0.94	768.7	20.2
AaNm1	aamchr235_conserved	1431	464	464	1.00	2438.5	22.8
AaNm1	aamchr119	1525	386	386	1.00	3123.8	21.4
AaNm1	aamchr120	500	275	277	0.99	1427.2	17.8
AaNm1	aamchr121	199	399	404	0.99	389.5	23.0
AaNm1	aamchr122	951	293	293	1.00	2566.4	24.5
AaNm1	p120	2820	1365	1380	0.99	1615.7	20.3
AaNm1	snrnpSmD2	1176	345	345	1.00	2695.2	28.5
AaNm1	ppci	2747	977	977	1.00	2223.1	23.7
AaNm1	rps26	1531	332	332	1.00	3646.2	29.2
AaNm1	rps2	3443	853	853	1.00	3191.5	27.3
AaNm1	eef2	25951	2534	2534	1.00	8097.5	33.9
AaNm1	rpl23	5809	462	462	1.00	9941.7	34.9
AaNm1	clpP-4	2197	619	619	1.00	2806.4	26.7
AaNm1	rps10	889	288	288	1.00	2440.7	21.8
AaNm1	sap62	1791	740	740	1.00	1913.7	27.9
AaNm1	clpP-3	4895	725	725	1.00	5338.5	29.6
AaNm1	aamchr236	1545	716	716	1.00	1706.2	21.6
AaNm1	duf572	986	454	455	1.00	1713.4	25.0
AaNm1	aamchr123	1430	3384	3615	0.94	312.8	21.6
AaNm1	sfb1	8915	2823	2846	0.99	2476.8	28.6
AaNm1	aamchr124	678	455	455	1.00	1178.2	23.5
AaNm1	aamchr137	1997	512	512	1.00	3084.0	22.6
AaNm1	aamchr125	2647	416	416	1.00	5031.1	23.3
AaNm1	rpl4A	3669	890	908	0.98	3195.0	28.4
AaNm1	snRNPE-2	1304	260	290	0.90	3555.4	23.4
AaNm1	cdc5	4290	1457	1457	1.00	2328.1	25.2
AaNm1	prp19	978	457	493	0.93	1568.5	21.4
AaNm1	aamchr238_conserved	799	419	444	0.94	1422.9	22.8
AaNm1	mcm2	3633	1904	1904	1.00	1508.7	24.6
AaNm1	rps11	1719	439	439	1.00	3096.1	23.9
AaNm1	BNATCHR290	2133	649	649	1.00	2598.7	27.8
AaNm1	rpc10	976	176	176	1.00	4384.7	27.1
AaNm1	tcpE	2030	1570	1594	0.98	1007.0	25.2
AaNm1	rpl7A-2	1159	490	519	0.94	1765.7	22.4
AaNm1	rpl37ae	2605	303	303	1.00	6797.8	29.6
AaNm1	aamchr126	1321	433	458	0.95	2280.6	21.1
AaNm1	rad25	1981	1574	1677	0.94	934.0	24.3
AaNm1	aamchr127	1376	1820	1841	0.99	591.0	23.9
AaNm1	eif6	1511	816	819	1.00	1458.8	23.9
AaNm1	prp43-1	7918	2070	2070	1.00	3024.5	28.8
AaNm1	clf1	2114	1720	1756	0.98	951.9	22.4
AaNm1	aamchr238_conserved	1597	1014	1014	1.00	1245.3	24.3
AaNm1	cdc28	9468	2352	2352	1.00	3182.9	27.9
AaNm1	rpl5	3502	830	852	0.97	3250.0	23.8
AaNm1	eif2G	2287	1323	1355	0.98	1334.5	28.3
AaNm1	aamchr128	905	1143	1218	0.94	587.5	20.4
AaNm1	aamchr129	546	197	197	1.00	2191.4	23.9
AaNm1	cwc22	1821	1288	1357	0.95	1061.0	23.1
AaNm1	rps28	966	309	315	0.98	2424.8	29.5
AaNm1	rpl15	3023	501	501	1.00	4770.9	27.3
AaNm1	aamchr130	1396	1421	1443	0.98	764.9	20.5
AaNm1	dnaK_2	10736	2366	2366	1.00	3587.8	48.1
AaNm2	dnaK_4	11103	2366	2366	1.00	3710.5	48.2
AaNm2	aamchr201	404	501	506	0.99	631.3	48.5
AaNm2	HSP70	138365	1939	2012	0.96	54375.3	33.7
AaNm2	rpl11	3615	568	573	0.99	4988.3	26.5
AaNm2	G10	3168	482	483	1.00	5186.1	27.7
AaNm2	bnatchr362	3581	1146	1166	0.98	2428.3	22.4

AaNm2	rpl10a	2289	704	704	1.00	2570.8	25.0
AaNm2	prp38	2230	509	509	1.00	3464.1	26.5
AaNm2	aamchr202	1279	389	389	1.00	2599.7	26.0
AaNm2	sf3b4	1909	719	719	1.00	2099.3	28.2
AaNm2	aamchr203	648	347	347	1.00	1476.6	24.7
AaNm2	nucleolar_GTPase	1747	777	777	1.00	1777.8	25.0
AaNm2	rps7	1546	643	643	1.00	1901.1	25.6
AaNm2	aamchr204	3342	1629	1629	1.00	1622.1	23.0
AaNm2	aamchr205	2241	1415	1426	0.99	1242.6	23.7
AaNm2	rps9	2302	560	564	0.99	3227.2	25.0
AaNm2	aamchr206	603	658	674	0.98	707.4	23.6
AaNm2	aamchr207	777	757	888	0.85	691.8	23.1
AaNm2	bnatchr184	21806	4274	4274	1.00	4034.1	26.6
AaNm2	tcpB	1172	1579	1579	1.00	586.9	22.7
AaNm2	nop56	2259	1152	1152	1.00	1550.5	24.0
AaNm2	bnatchr2118	344	437	437	1.00	622.4	19.4
AaNm2	aamchr208	144	410	466	0.88	244.3	21.0
AaNm2	rhelx	1227	989	993	1.00	977.0	22.4
AaNm2	pno1	400	528	528	1.00	599.0	20.2
AaNm2	aamchr209	251	602	602	1.00	329.7	19.1
AaNm2	H3	2285	409	409	1.00	4417.4	23.7
AaNm2	aamchr210	2675	923	945	0.98	2238.2	23.7
AaNm2	aamchr211	447	584	722	0.81	489.5	20.7
AaNm2	rps5	3215	801	801	1.00	3173.6	27.7
AaNm2	rpl17	3915	501	501	1.00	6178.7	24.9
AaNm2	aamchr212	1485	257	257	1.00	4568.7	23.3
AaNm2	aamchr213	575	344	344	1.00	1321.6	25.2
AaNm2	mcf1.25	1885	652	652	1.00	2286.0	24.3
AaNm2	rps13	2269	452	452	1.00	3969.2	31.3
AaNm2	prp6_homolog	2286	2054	2056	1.00	879.1	23.7
AaNm2	rps4	4191	795	803	0.99	4126.7	27.3
AaNm2	snRNPG	778	251	251	1.00	2450.8	27.3
AaNm2	clpP-5	1901	602	602	1.00	2496.8	23.6
AaNm2	rps21e_bnatchr2101	4276	264	264	1.00	12806.7	26.8
AaNm2	aamchr214	527	297	297	1.00	1403.0	22.2
AaNm2	nip7	852	536	536	1.00	1256.8	22.0
AaNm2	sf3a3	1588	1192	1394	0.86	900.7	20.7
AaNm2	aamchr215	203	715	772	0.93	207.9	20.1
AaNm2	aamchr216	447	509	537	0.95	658.2	21.0
AaNm2	H4	885	300	300	1.00	2332.5	24.8
AaNm2	aamchr217	2922	2013	2044	0.98	1130.3	22.7
AaNm2	rnabp2	1477	667	667	1.00	1750.9	23.1
AaNm2	bnatchr297	742	354	354	1.00	1657.3	22.6
AaNm2	rps3a	1902	705	716	0.98	2100.4	22.7
AaNm2	sf3bA	2403	814	814	1.00	2334.2	27.5
AaNm2	secY	10166	1379	1379	1.00	5828.9	26.4
AaNm2	murL	8868	1906	1922	0.99	3648.2	26.0
AaNm2	rps23	2392	497	501	0.99	3775.1	29.8
AaNm2	aamchr218	1092	458	458	1.00	1885.2	24.5
AaNm2	aamchr219	459	898	993	0.90	365.5	19.6
AaNm2	rpl7Ae	1081	398	422	0.94	2025.4	30.0
AaNm2	aamchr220	527	478	504	0.95	826.8	21.8
AaNm2	rpl19	1574	474	474	1.00	2625.6	22.4
AaNm2	aamchr221	7749	2309	2309	1.00	2653.5	22.9
AaNm2	aamchr222	586	367	375	0.98	1235.6	23.0
AaNm2	ysh1	1811	1673	1717	0.97	834.0	22.8
AaNm2	rpc1	7594	4255	4283	0.99	1401.9	24.2
AaNm2	prp43-2	3884	2078	2081	1.00	1475.7	24.3
AaNm2	aamchr223	686	589	589	1.00	920.9	21.6
AaNm2	aamchr224	6172	3352	3394	0.99	1437.9	23.8
AaNm2	aamchr225	2566	1727	1758	0.98	1154.1	22.0
AaNm2	aamchr238	3118	1266	1266	1.00	1947.4	24.1
AaNm2	rps8_2	3775	608	608	1.00	4909.3	30.1
AaNm2	bnatchr371	2365	1739	1739	1.00	1075.3	23.9
AaNm2	aamchr226	12	124	219	0.57	43.3	22.4

AaNm2	ycf16	17455	846	846	1.00	16313.7	28.8
AaNm2	aamchr227	10521	528	528	1.00	15755.3	25.7
AaNm2	aamchr228	1263	562	630	0.89	1585.1	22.2
AaNm2	snRNPSm-D	845	256	266	0.96	2511.8	24.1
AaNm2	rps17E	1160	376	376	1.00	2439.3	21.2
AaNm2	aamchr229	1465	1620	1648	0.98	702.9	22.6
AaNm2	bnatchr2105	11349	1055	1055	1.00	8505.7	25.8
AaNm2	aamchr230	946	323	355	0.91	2107.0	18.7
AaNm2	prp17	3861	1173	1198	0.98	2548.3	26.8
AaNm2	prl1	4202	1160	1190	0.97	2792.0	27.6
AaNm2	aamchr231	576	549	566	0.97	804.7	22.2
AaNm2	mcm4	3812	1845	1850	1.00	1629.2	23.1
AaNm2	toc75	11916	2375	2393	0.99	3937.2	27.7
AaNm2	rps30	853	158	158	1.00	4268.7	25.2
AaNm2	bantchr2108	3386	1024	1024	1.00	2614.5	25.4
AaNm2	sys1	2465	1279	1301	0.98	1498.1	27.0
AaNm2	aamchr239	517	697	698	1.00	585.7	21.4
AaNm2	rpabc5	1333	679	679	1.00	1552.3	24.8
AaNm2	aamchr232	7732	2657	2657	1.00	2300.9	24.2
AaNm2	rps3	4129	916	916	1.00	3564.1	23.8
AaNm2	dib1	5335	458	458	1.00	9210.3	30.7
AaNm2	bnatchr1116	3335	990	990	1.00	2663.6	24.1
AaNm2	aamchr234	480	380	380	1.00	998.8	23.1
AaNm2	aamchr235	1450	758	758	1.00	1512.5	22.6
AaNm2	aamchr236	1377	545	545	1.00	1997.7	23.9
AaNm2	aamchr237	475	444	444	1.00	845.9	24.9
AaNm2	clpP-1	4510	681	681	1.00	5236.4	27.1
AaNm2	mago	1160	488	488	1.00	1879.5	25.7
AaNm2	snrnpSmD3	1407	427	427	1.00	2605.4	28.2
AaNm2	rbp8	1491	478	478	1.00	2466.3	29.0
AaNm2	dnaK_3	11810	2366	2366	1.00	3946.7	48.2
AaNm3	dnak_6	10973	2366	2366	1.00	3667.0	48.2
AaNm3	aamchr301	449	506	506	1.00	701.6	48.3
AaNm3	aamchr302	1205	567	611	0.93	1559.4	23.4
AaNm3	dip2	1582	1102	1110	0.99	1126.9	25.3
AaNm3	rpl7A-1	3071	981	981	1.00	2475.2	25.6
AaNm3	rbp2	9258	3665	3677	1.00	1990.8	29.5
AaNm3	aamchr304	801	299	299	1.00	2118.2	22.7
AaNm3	cwf24	1723	251	251	1.00	5427.7	26.2
AaNm3	rad51	2411	1019	1019	1.00	1870.8	28.5
AaNm3	bnatchr2116	1443	385	385	1.00	2963.5	29.1
AaNm3	aamchr306	723	407	407	1.00	1404.6	23.5
AaNm3	aamchr307	3558	2474	2546	0.97	1105.0	23.1
AaNm3	u5snRNP_200kDa	9439	5445	5445	1.00	1370.7	25.5
AaNm3	sig2	4179	1283	1287	1.00	2567.4	27.0
AaNm3	rps27	3012	355	355	1.00	6708.6	28.9
AaNm3	myb1	8102	2068	2078	1.00	3082.8	29.5
AaNm3	cdc48-like	2992	1768	1773	1.00	1334.3	26.0
AaNm3	rpc2	8871	3775	3784	1.00	1853.6	26.3
AaNm3	rpl2	2412	464	464	1.00	4110.2	28.2
AaNm3	aamchr308	2199	1516	1516	1.00	1146.9	24.2
AaNm3	aamchr309	898	611	611	1.00	1162.1	23.1
AaNm3	mak16	719	637	670	0.95	848.5	21.2
AaNm3	snRNP-F	868	297	297	1.00	2310.8	25.7
AaNm3	gsp2	11180	669	669	1.00	13213.5	31.3
AaNm3	aamchr310	2676	381	381	1.00	5553.5	24.8
AaNm3	tic20	3415	696	696	1.00	3879.6	26.0
AaNm3	hub1	241	214	254	0.84	750.2	18.7
AaNm3	aamchr311	59	136	164	0.83	284.5	19.4
AaNm3	rpa2	13178	3545	3558	1.00	2928.5	27.5
AaNm3	aamchr312	593	973	985	0.99	476.0	23.6
AaNm3	tfIId	1603	695	704	0.99	1800.4	28.1
AaNm3	rps0	3043	677	677	1.00	3554.0	29.6
AaNm3	aamchr313	512	311	311	1.00	1301.7	23.7
AaNm3	tcpD	2829	1557	1630	0.96	1372.3	24.0

AaNm3	rpl13A	2013	628	630	1.00	2526.4	23.0
AaNm3	rpoF	1305	1395	1424	0.98	724.6	21.7
AaNm3	BNATCHR186	3623	806	806	1.00	3554.2	23.9
AaNm3	sf3b5	1703	212	212	1.00	6351.6	24.4
AaNm3	aamchr314	1259	932	959	0.97	1038.0	22.2
AaNm3	tcpT	632	1500	1646	0.91	303.6	21.1
AaNm3	aamchr315	535	710	740	0.96	571.6	21.5
AaNm3	BNATCHR374	1439	678	690	0.98	1649.0	23.9
AaNm3	aamchr316	2353	385	385	1.00	4832.4	24.4
AaNm3	sufB	25007	1538	1566	0.98	12626.2	31.8
AaNm3	clpP-2	5338	857	857	1.00	4924.9	28.3
AaNm3	tatC	5355	905	905	1.00	4678.6	27.3
AaNm3	aamchr317	243	211	211	1.00	910.6	25.0
AaNm3	mrp1	9600	978	978	1.00	7761.3	30.4
AaNm3	nbp1	347	346	408	0.85	672.5	19.2
AaNm3	rpl8	4067	847	857	0.99	3752.3	30.8
AaNm3	prp8	30550	6824	6824	1.00	3539.8	28.3
AaNm3	rpa1	8001	4503	4506	1.00	1404.0	26.6
AaNm3	prp14	761	333	333	1.00	1806.9	25.9
AaNm3	aamchr318	1590	507	507	1.00	2479.7	23.3
AaNm3	aamchr319	814	798	798	1.00	806.5	22.0
AaNm3	BNATCHR359	1984	524	524	1.00	2993.7	25.4
AaNm3	bnatchr360	1252	294	294	1.00	3367.1	22.9
AaNm3	sf1_bnatchr361	4336	1215	1215	1.00	2821.7	23.9
AaNm3	aamchr321	1102	560	577	0.97	1510.1	23.4
AaNm3	rbp10	576	234	234	1.00	1946.3	30.8
AaNm3	conserved_ATPase	3641	2603	2622	0.99	1098.0	23.9
AaNm3	msl1	3117	760	760	1.00	3242.8	26.1
AaNm3	clpP-6	3733	702	702	1.00	4204.6	28.4
AaNm3	rps6	3051	643	672	0.96	3589.8	27.1
AaNm3	ClpC	75917	2729	2729	1.00	21995.7	32.7
AaNm3	prp16	7045	2421	2421	1.00	2300.9	27.0
AaNm3	aamchr322	3791	1931	1931	1.00	1552.3	24.2
AaNm3	sbp1	598	794	799	0.99	591.8	24.1
AaNm3	aamchr323	640	308	308	1.00	1643.0	23.7
AaNm3	rpl30	592	325	347	0.94	1348.9	22.7
AaNm3	aamchr324	387	323	323	1.00	947.4	23.5
AaNm3	cpn60	7406	1685	1685	1.00	3475.3	30.7
AaNm3	hsp90	122871	2123	2123	1.00	45761.7	32.2
AaNm3	prp22	3050	1911	1911	1.00	1262.0	29.0
AaNm3	rps27A	1985	322	322	1.00	4874.3	26.1
AaNm3	tfIIA-gamma	516	434	434	1.00	940.1	21.7
AaNm3	RNApolymeraseII_TF	1051	782	812	0.96	1023.4	26.0
AaNm3	aamchr325	1575	1737	1838	0.95	677.5	29.4
AaNm3	dnaK_5	11394	2364	2366	1.00	3807.7	48.2
	Sum	1264727			**Average**	3032.2	25.8

Mapping rates were calculated by gene lengths and mapped regions.
RPKM (Reads Per Kilobase of exon per Million mapped reads) was calculated by the following formula;
=(number of reads for each gene)*1000000/(total number of reads)*1000/(gene length)

Table 4.2 Relative expression of nucleomorph genes in *Bigelowiella natans*

Chromosome No.	Gene name	Number of reads	Mapped by some reads (bp)	Gene length (bp)	Mapping rate	RPKM	GC content (%)
BnNm1	sf3b4	18320	722	722	1.00	3878.5	29.5
BnNm1	BNATCHR182	11644	353	353	1.00	5042.0	23.2
BnNm1	BNATCHR183_nucleolar_GTPase	7578	768	768	1.00	1508.2	23.1
BnNm1	BNATCHR184	109938	4301	4301	1.00	3907.1	26.8
BnNm1	BNATCHR185	1009	332	332	1.00	464.6	19.8
BnNm1	tcpD	38292	1693	1693	1.00	3457.2	26.6
BnNm1	rpl13A	20139	620	620	1.00	4965.1	25.8
BnNm1	rpoF	10297	1289	1289	1.00	1221.1	22.9
BnNm1	BNATCHR186	35277	747	747	1.00	7218.6	24.3
BnNm1	sf3b5	10330	218	218	1.00	7243.1	27.4
BnNm1	snRNPE-1	6565	923	923	1.00	1087.2	22.9
BnNm1	snRNPE-2	6829	280	280	1.00	3728.0	25.5
BnNm1	rpl7A-1	25477	1149	1149	1.00	3389.3	22.9
BnNm1	dip2	16377	1155	1155	1.00	2167.4	23.8
BnNm1	BNATCHR187	3196	338	338	1.00	1445.3	23.0
BnNm1	tcpE	23768	1656	1656	1.00	2193.9	26.8
BnNm1	rpl7A-2	16715	512	512	1.00	4990.2	24.9
BnNm1	rpl37ae	11637	269	269	1.00	6612.5	27.7
BnNm1	BNATCHR188	994	392	394	0.99	385.6	18.8
BnNm1	rad25	22037	1711	1738	0.98	1938.1	26.0
BnNm1	bnatchr113	11301	1874	1878	1.00	919.8	22.7
BnNm1	eif6	13787	760	760	1.00	2772.9	23.1
BnNm1	prp43-1	34668	2086	2086	1.00	2540.4	29.0
BnNm1	clpP-1	23717	751	751	1.00	4827.2	26.5
BnNm1	BNATCHR189	3331	435	455	0.96	1119.0	21.1
BnNm1	prp14	6380	322	322	1.00	3028.6	25.3
BnNm1	prp22	25531	2063	2063	1.00	1891.7	26.4
BnNm1	rps27A	8869	294	294	1.00	4611.1	26.1
BnNm1	tfIIA-gamma	7682	430	430	1.00	2730.8	21.6
BnNm1	snRNP-F	14508	255	255	1.00	8696.5	26.6
BnNm1	BNATCHR190	16747	1681	1681	1.00	1522.8	21.9
BnNm1	rbp10	6049	216	216	1.00	4280.6	27.8
BnNm1	BNATCHR191	45363	2853	2860	1.00	2424.5	25.2
BnNm1	BNATCHR192	5029	1942	1951	1.00	394.0	21.1
BnNm1	BNATCHR193	2571	315	315	1.00	1247.6	21.2
BnNm1	bnatchr1121	13244	213	213	1.00	9504.3	24.1
BnNm1	rpl18A	11851	555	555	1.00	3263.9	24.1
BnNm1	rps15	15563	391	392	1.00	6068.6	28.5
BnNm1	BNATCHR194_ATPase	29563	2280	2280	1.00	1982.0	23.2
BnNm1	BNATCHR195	1109	266	266	1.00	637.3	19.1
BnNm1	rps12	22426	390	390	1.00	8789.6	24.3
BnNm1	mce	2571	1259	1259	1.00	312.1	19.7
BnNm1	rpl32	9983	358	358	1.00	4262.4	24.0
BnNm1	BNATCHR196	6101	845	845	1.00	1103.6	22.7
BnNm1	BNATCHR197	25674	1133	1139	0.99	3445.5	21.5
BnNm1	sf3a3	18847	1387	1410	0.98	2043.2	22.3
BnNm1	rps7	13343	627	627	1.00	3252.9	23.2
BnNm1	BNATCHR198	2849	644	644	1.00	676.2	19.7
BnNm1	BNATCHR199	5695	868	868	1.00	1002.9	21.3
BnNm1	ubiquitin-like	1956	177	177	1.00	1689.2	22.6
BnNm1	BNATCHR1100	6781	1110	1123	0.99	923.0	22.4
BnNm1	BNATCHR1101	6747	951	951	1.00	1084.5	21.7
BnNm1	BNATCHR1102	7992	2023	2036	0.99	600.0	20.5
BnNm1	snRNPSm-D	23826	252	252	1.00	14452.1	24.8
BnNm1	rps17E	5470	328	329	1.00	2541.4	21.3
BnNm1	BNATCHR1103	2866	254	254	1.00	1724.7	27.4
BnNm1	BNATCHR1104	414	404	404	1.00	156.6	20.2
BnNm1	bnatchr1122	3161	191	191	1.00	2529.7	20.8
BnNm1	rpa1	34966	4079	4079	1.00	1310.3	23.8
BnNm1	prp8	121668	6797	6797	1.00	2736.1	28.2

BnNm1	rpl8	25265	934	941	0.99	4104.0	30.3
BnNm1	nbp1	1789	424	424	1.00	644.9	23.3
BnNm1	mrp1	41381	943	943	1.00	6707.6	30.1
BnNm1	tatC	43569	965	965	1.00	6901.3	25.9
BnNm1	clpP-2	49435	853	853	1.00	8858.6	28.3
BnNm1	sufB	141944	1469	1469	1.00	14769.8	32.2
BnNm1	BNATCHR1105	17586	499	499	1.00	5387.0	23.6
BnNm1	rpl30	7515	347	347	1.00	3310.4	22.7
BnNm1	BNATCHR1106	1816	329	329	1.00	843.7	20.2
BnNm1	cpn60	33674	1649	1649	1.00	3121.4	29.4
BnNm1	BNATCHR1107	5814	437	437	1.00	2033.6	22.3
BnNm1	prp19	11351	496	496	1.00	3498.1	26.4
BnNm1	hsp90	518834	2087	2087	1.00	38000.2	30.3
BnNm1	BNATCHR1108	477	491	491	1.00	148.5	19.9
BnNm1	G10	21495	501	501	1.00	6558.1	25.9
BnNm1	rpl11	17743	476	476	1.00	5697.7	27.4
BnNm1	ysh1	27435	1767	1767	1.00	2373.3	24.9
BnNm1	rpoB	8974	2301	2339	0.98	586.5	21.8
BnNm1	rpl15	15296	512	512	1.00	4566.5	23.7
BnNm1	rnabp1	32777	501	501	1.00	10000.3	25.7
BnNm1	dbp1	18623	1542	1542	1.00	1846.1	25.1
BnNm1	eif-4C	10681	432	432	1.00	3779.3	27.2
BnNm1	rpl9	7882	568	572	0.99	2106.3	23.6
BnNm1	fibrellarin	20100	761	761	1.00	4037.3	30.3
BnNm1	trf	2167	801	801	1.00	413.5	20.9
BnNm1	BNATCHR1110	23460	1546	1546	1.00	2319.5	24.8
BnNm1	BNATCHR1111_conserved	4170	801	801	1.00	795.8	23.6
BnNm1	rpb1	41876	4532	4532	1.00	1412.4	28.7
BnNm1	BNATCHR1112	26616	622	622	1.00	6540.8	28.9
BnNm1	eif-4A2	32108	1212	1212	1.00	4049.4	29.3
BnNm1	pcna	23206	853	853	1.00	4158.4	26.9
BnNm1	BNATCHR1113	3921	616	616	1.00	973.0	20.6
BnNm1	rpl14A	5849	393	393	1.00	2274.9	27.5
BnNm1	BNATCHR1114	19121	1597	1621	0.99	1803.0	21.6
BnNm1	BNATCHR1115	2885	362	362	1.00	1218.2	20.7
BnNm1	prp43-2	19644	2072	2072	1.00	1449.2	24.8
BnNm1	rpc1	57332	4313	4313	1.00	2031.9	26.0
BnNm1	rps3	18135	826	826	1.00	3356.0	22.9
BnNm1	dib1	12214	475	475	1.00	3930.5	28.1
BnNm1	rps9	24104	581	581	1.00	6341.5	26.7
BnNm1	BNATCHR1116	29937	1049	1049	1.00	4362.3	25.7
BnNm1	rbp11	3163	332	332	1.00	1456.3	24.1
BnNm1	BNATCHR1117	26561	347	347	1.00	11700.2	27.3
BnNm1	rpl4A	26493	836	836	1.00	4844.0	29.7
BnNm1	BNATCHR1118	5083	500	500	1.00	1553.9	21.5
BnNm1	BNATCHR1119	36447	1634	1634	1.00	3409.5	25.1
BnNm1	BNATCHR1120	398	266	266	1.00	228.7	17.7
BnNm1	sbp1	4493	918	918	1.00	748.1	27.2
BnNm2	dnaK	75240	1952	1961	1.00	5864.8	37.2
BnNm2	rpl3	47273	1188	1188	1.00	6082.4	28.8
BnNm2	BNATCHR282	22531	2560	2564	1.00	1343.2	21.8
BnNm2	rps14	13483	494	494	1.00	4171.9	32.4
BnNm2	tfIIB	22069	639	639	1.00	5279.1	26.3
BnNm2	BNATCHR283	2336	682	682	1.00	523.6	22.5
BnNm2	fet5	12213	723	723	1.00	2582.0	24.5
BnNm2	BNATCHR284_conserved	46286	1307	1307	1.00	5413.2	29.1
BnNm2	secY	42534	1581	1581	1.00	4112.3	25.1
BnNm2	murL	42145	1748	1748	1.00	3685.4	26.2
BnNm2	rps23	16524	503	503	1.00	5021.4	33.1
BnNm2	BNATCHR285	4103	532	547	0.97	1146.6	22.0
BnNm2	BNATCHR286	6961	781	781	1.00	1362.4	22.4
BnNm2	rpl7Ae	7028	422	422	1.00	2545.7	30.3
BnNm2	BNATCHR287_conserved	6473	518	518	1.00	1910.1	25.1
BnNm2	rpl10	29817	832	832	1.00	5478.0	26.3
BnNm2	rps16	19962	513	513	1.00	5947.9	26.4

BnNm2	rps8	35932	458	458	1.00	11992.1	25.5
BnNm2	imp4	23463	716	716	1.00	5009.0	24.5
BnNm2	sf3b	60070	3514	3514	1.00	2613.0	26.6
BnNm2	rpl44	11527	356	356	1.00	4949.3	27.3
BnNm2	rpa5	16133	1012	1012	1.00	2436.8	26.0
BnNm2	rpb3	9288	894	894	1.00	1588.1	25.1
BnNm2	BNATCHR288	1403	417	422	0.99	508.2	21.0
BnNm2	BNATCHR289	5023	435	435	1.00	1765.0	22.8
BnNm2	rpl27a	22243	515	515	1.00	6601.8	25.9
BnNm2	tcpH	10049	1604	1604	1.00	957.6	23.3
BnNm2	rpl10e	25175	638	638	1.00	6031.5	29.0
BnNm2	rds3	13985	373	373	1.00	5731.0	32.1
BnNm2	rpabc6	12027	404	404	1.00	4550.5	30.7
BnNm2	hub1	5229	272	272	1.00	2938.5	26.2
BnNm2	bnatchr2120	4286	224	224	1.00	2924.7	35.1
BnNm2	rpl10a	29852	707	707	1.00	6454.1	28.0
BnNm2	rpc10	9887	173	173	1.00	8735.7	27.6
BnNm2	BNATCHR290	13509	618	618	1.00	3341.3	25.2
BnNm2	rps11	11640	416	416	1.00	4277.0	24.7
BnNm2	mcm2	19652	2268	2308	0.98	1301.5	23.0
BnNm2	BNATCHR291	3991	694	694	1.00	879.0	20.1
BnNm2	BNATCHR292	1064	344	344	1.00	472.8	19.7
BnNm2	tcpB	14475	1566	1579	0.99	1401.3	24.4
BnNm2	mak16	24253	618	618	1.00	5998.7	26.0
BnNm2	BNATCHR293	2858	699	699	1.00	625.0	20.7
BnNm2	tcpG	9190	1514	1514	1.00	927.8	23.0
BnNm2	rpl2	28734	499	499	1.00	8801.9	28.6
BnNm2	rpc2	39644	3364	3377	1.00	1794.4	26.9
BnNm2	cdc48-like	18727	1677	1677	1.00	1706.9	24.9
BnNm2	BNATCHR294	3964	786	786	1.00	770.9	20.6
BnNm2	BNATCHR295	5394	592	592	1.00	1392.7	22.5
BnNm2	H4	10173	305	305	1.00	5098.3	27.4
BnNm2	BNATCHR296	14053	1984	1984	1.00	1082.7	21.3
BnNm2	rnabp2	6159	615	615	1.00	1530.8	22.8
BnNm2	BNATCHR297	8672	340	345	0.99	3842.2	24.6
BnNm2	rps3a	32090	722	722	1.00	6793.8	23.0
BnNm2	sf3bA	24230	749	749	1.00	4944.8	27.7
BnNm2	BNATCHR298	17019	500	500	1.00	5202.9	22.5
BnNm2	BNATCHR299	28949	3953	3992	0.99	1108.5	21.2
BnNm2	duf572	6914	344	344	1.00	3072.2	23.5
BnNm2	BNATCHR2100	7048	709	709	1.00	1519.5	21.0
BnNm2	clpP-3	36031	623	623	1.00	8840.3	30.1
BnNm2	myb1	54204	2043	2043	1.00	4055.5	29.1
BnNm2	rps21e	11510	344	344	1.00	5114.4	25.4
BnNm2	rps27	18820	319	319	1.00	9018.0	28.4
BnNm2	sig2	20277	1293	1302	0.99	2380.5	25.2
BnNm2	u5snRNP	58224	5671	5678	1.00	1567.4	24.4
BnNm2	BNATCHR2102	2930	358	358	1.00	1251.0	20.8
BnNm2	BNATCHR2119_conserved	13417	422	424	1.00	4836.9	23.8
BnNm2	BNATCHR2103	2116	445	445	1.00	726.8	22.3
BnNm2	clpP-4	14985	627	627	1.00	3653.2	26.3
BnNm2	rpl23	33343	467	467	1.00	10913.6	31.9
BnNm2	rps18	18436	516	516	1.00	5461.3	27.7
BnNm2	BNATCHR2104	7386	602	610	0.99	1850.8	23.7
BnNm2	rpl27	14260	451	451	1.00	4833.1	21.5
BnNm2	u2AF	34085	638	638	1.00	8166.2	30.8
BnNm2	BNATCHR2105	41078	805	805	1.00	7800.0	25.4
BnNm2	rpl23	4504	268	269	1.00	2559.3	19.8
BnNm2	prp17	19717	1287	1295	0.99	2327.3	28.7
BnNm2	prl1	15383	1232	1237	1.00	1900.9	26.3
BnNm2	BNATCHR2107	1547	504	536	0.94	441.2	20.1
BnNm2	mcm4	19119	1984	1984	1.00	1473.0	21.8
BnNm2	toc75	77106	2456	2456	1.00	4798.9	27.1
BnNm2	rps30	16027	187	187	1.00	13100.6	25.0
BnNm2	BNATCHR2108	16870	1027	1027	1.00	2510.9	22.4

BnNm2	sys1	29199	1341	1341	1.00	3328.3	28.0
BnNm2	BNATCHR2109	4455	875	875	1.00	778.3	21.9
BnNm2	rpabc5	8573	676	710	0.95	1845.7	23.3
BnNm2	BNATCHR2110	31437	2621	2621	1.00	1833.4	23.0
BnNm2	ClpC	286997	2691	2693	1.00	16290.0	32.5
BnNm2	prp16	27621	2357	2357	1.00	1791.3	25.8
BnNm2	nip7	5136	580	580	1.00	1353.6	21.5
BnNm2	BNATCHR2112	4355	333	333	1.00	1999.0	26.3
BnNm2	BNATCHR2113	1438	503	512	0.98	429.3	22.1
BnNm2	BNATCHR2114	11733	1906	1920	0.99	934.1	22.0
BnNm2	BNATCHR2115	3307	431	431	1.00	1172.8	23.6
BnNm2	BNATCHR2116	15297	267	267	1.00	8757.4	30.5
BnNm2	rad51	26581	1091	1091	1.00	3724.1	31.0
BnNm2	cwf24	2554	290	290	1.00	1346.2	20.3
BnNm2	bnatchr2121_conserved	6903	309	309	1.00	3414.7	22.0
BnNm2	rbp2	57932	3652	3652	1.00	2424.8	30.0
BnNm2	cdc5	27681	1687	1687	1.00	2508.1	26.4
BnNm2	rpl34	8986	330	330	1.00	4162.3	24.5
BnNm2	prp45	12964	586	586	1.00	3381.6	26.2
BnNm2	BNATCHR2117	9475	370	370	1.00	3914.3	21.9
BnNm2	BNATCHR2118	16867	443	443	1.00	5819.9	23.0
BnNm2	nop5	11479	1156	1156	1.00	1517.8	22.4
BnNm2	gsp2	59021	736	736	1.00	12257.7	31.2
BnNm3	rps13	20635	608	608	1.00	5187.8	30.3
BnNm3	prp6	26933	2135	2135	1.00	1928.3	23.9
BnNm3	rps4	24603	805	805	1.00	4671.7	27.0
BnNm3	snRNPG	6716	253	253	1.00	4057.6	23.6
BnNm3	clpP-5	26666	736	736	1.00	5538.1	24.7
BnNm3	rbp8	7877	500	500	1.00	2408.1	26.6
BnNm3	rhel1	9502	1483	1483	1.00	979.4	24.7
BnNm3	sfb1	57625	2885	2885	1.00	3053.1	29.6
BnNm3	BNATCHR358	3079	778	778	1.00	604.9	22.4
BnNm3	BNATCHR359	13930	362	362	1.00	5882.0	25.9
BnNm3	BNATCHR360	1729	227	247	0.92	1070.0	16.7
BnNm3	BNATCHR361	17806	1123	1144	0.98	2379.1	26.0
BnNm3	BNATCHR362	30291	1135	1135	1.00	4079.4	24.1
BnNm3	mago	8337	505	505	1.00	2523.5	26.6
BnNm3	snrnpSmD3	16419	464	464	1.00	5408.9	26.5
BnNm3	ycf16	89146	884	884	1.00	15414.5	29.2
BnNm3	ub2	14617	1348	1348	1.00	1657.5	25.7
BnNm3	tcpT	15232	1695	1695	1.00	1373.6	23.4
BnNm3	snrpB	5464	287	287	1.00	2910.1	26.1
BnNm3	rpl24	8669	467	467	1.00	2837.5	27.5
BnNm3	BNATCHR363	1828	402	402	1.00	695.1	21.1
BnNm3	BNATCHR364	5043	357	357	1.00	2159.2	23.7
BnNm3	BNATCHR365	1729	392	392	1.00	674.2	18.8
BnNm3	BNATCHR366	6194	185	185	1.00	5117.7	27.4
BnNm3	p120	10268	1333	1333	1.00	1177.4	21.4
BnNm3	snrnpSmD2	11736	324	324	1.00	5536.7	26.8
BnNm3	ppci	13925	998	998	1.00	2132.8	21.9
BnNm3	rps26	10341	334	334	1.00	4732.6	27.6
BnNm3	rps2	20557	824	838	0.98	3749.7	25.7
BnNm3	eef2	85969	2539	2539	1.00	5175.6	32.3
BnNm3	rps10	7732	293	293	1.00	4033.7	22.4
BnNm3	sap62	15560	720	720	1.00	3303.4	26.1
BnNm3	rps0	41433	635	635	1.00	9973.6	25.5
BnNm3	tfIId	23915	708	708	1.00	5163.2	30.2
BnNm3	BNATCHR367	3376	1011	1011	1.00	510.4	22.2
BnNm3	BNATCHR368	3113	332	332	1.00	1433.2	21.3
BnNm3	BNATCHR369	27033	1707	1707	1.00	2420.7	23.6
BnNm3	BNATCHR370	4208	1112	1112	1.00	578.4	20.7
BnNm3	prp38	5404	476	489	0.97	1689.2	23.6
BnNm3	BNATCHR371	19821	1901	1901	1.00	1593.8	22.4
BnNm3	rps8	30551	588	588	1.00	7942.0	26.7
BnNm3	BNATCHR372	24704	3179	3191	1.00	1183.4	22.5

BnNm3	BNATCHR385	1003	275	311	0.88	493.0	23.4
BnNm3	BNATCHR373	24416	2678	2678	1.00	1393.6	22.8
BnNm3	prsA6	4441	697	701	0.99	968.4	21.9
BnNm3	BNATCHR374_conserved	6874	645	676	0.95	1554.3	23.8
BnNm3	rpa2	47811	3432	3432	1.00	2129.4	27.0
BnNm3	mcf1.25	9346	645	645	1.00	2214.9	24.6
BnNm3	BNATCHR375	4849	323	323	1.00	2294.7	25.0
BnNm3	BNATCHR376	3425	276	276	1.00	1896.8	20.9
BnNm3	rpl17	23011	477	477	1.00	7373.9	23.3
BnNm3	rps5	22822	719	719	1.00	4851.8	30.6
BnNm3	BNATCHR377	5093	676	676	1.00	1151.6	24.3
BnNm3	BNATCHR378	14461	963	963	1.00	2295.4	24.3
BnNm3	H3	8462	427	427	1.00	3029.2	25.9
BnNm3	BNATCHR379	9537	661	661	1.00	2205.4	21.9
BnNm3	pno1	10866	650	650	1.00	2555.3	20.7
BnNm3	rhelx	9669	1049	1049	1.00	1408.9	22.8
BnNm3	BNATCHR381	6522	964	964	1.00	1034.1	20.9
BnNm3	rps6	34821	796	796	1.00	6686.6	26.8
BnNm3	clf1	25981	1867	1867	1.00	2127.1	24.2
BnNm3	rfc4	10555	1037	1037	1.00	1555.8	25.9
BnNm3	cdc28	30311	2396	2396	1.00	1933.7	27.8
BnNm3	rpl5	16294	796	796	1.00	3128.9	24.5
BnNm3	eif2G	16426	1376	1376	1.00	1824.7	27.0
BnNm3	BNATCHR382	9683	1045	1045	1.00	1416.4	23.1
BnNm3	BNATCHR383	9873	349	349	1.00	4324.2	24.2
BnNm3	cwc22	21515	1590	1590	1.00	2068.3	22.9
BnNm3	rps28	5977	234	234	1.00	3904.3	29.3
BnNm3	rpl19	12827	490	490	1.00	4001.4	21.2
BnNm3	tic20	21319	700	700	1.00	4655.3	28.2
BnNm3	clpP-6	10708	671	671	1.00	2439.3	22.8
BnNm3	msl1	19559	746	746	1.00	4007.6	27.4
BnNm3	BNATCHR384_conserved_atpase	29147	2662	2662	1.00	1673.7	25.2
BnNm3	hsp70	428335	1832	1832	1.00	35738.6	35.2
	Sum	6529430			**Average**	3691.6	24.9

Mapping rates were calculated by gene lengths and mapped regions.

RPKM (Reads Per Kilobase of exon per Million mapped reads) was calculated by the following formula;

=(number of reads for each gene)*1000000/(total number of reads)*1000/(gene length)

Table 4.3 Relative expression of nucleomorph genes in *Chroomonas mesostigmatica*

Chromosome No.	Gene name	Number of reads	Mapped by some reads (bp)	Gene length (bp)	Mapping rate	RPKM	GC content (%)
CmNm1	CMESO_5	2805	416	620	0.67	1079.1	27.5
CmNm1	tfIIB-brf	1111	1283	1283	1.00	206.5	37.7
CmNm1	CMESO_7	303	497	497	1.00	145.4	30.1
CmNm1	CMESO_9	1158	344	344	1.00	802.9	22.6
CmNm1	dbp4	7525	1490	1490	1.00	1204.5	33.9
CmNm1	snrpF	1964	299	299	1.00	1566.7	36.3
CmNm1	CMESO_13	2803	659	659	1.00	1014.5	31.4
CmNm1	CMESO_14	4023	650	650	1.00	1476.2	32.0
CmNm1	Ha-orf253	1466	638	638	1.00	548.0	27.4
CmNm1	CMESO_16	7707	2180	2180	1.00	843.2	29.3
CmNm1	kin(aaB)	9385	1322	1322	1.00	1693.2	30.0
CmNm1	cycB	1934	1064	1064	1.00	433.5	25.2
CmNm1	tfIIIC	5263	2276	2276	1.00	551.5	23.1
CmNm1	Gt-orf261	27043	1916	1916	1.00	3366.4	28.9
CmNm1	rps6	9178	647	647	1.00	3383.4	28.9
CmNm1	hira	1891	1823	1823	1.00	247.4	24.9
CmNm1	Gt-orf176	4736	614	614	1.00	1839.7	28.5
CmNm1	fet5	6503	761	761	1.00	2038.1	30.6
CmNm1	gyrA	12486	2606	2606	1.00	1142.8	33.0
CmNm1	tcpD	17010	1565	1565	1.00	2592.3	32.0
CmNm1	crm	5074	3074	3074	1.00	393.7	23.3
CmNm1	Gt-orf216	1925	809	809	1.00	567.5	30.4
CmNm1	nat10	8509	2639	2639	1.00	769.0	28.2
CmNm1	CMESO_30	1074	377	377	1.00	679.5	19.6
CmNm1	CMESO_31	9557	3445	3446	1.00	661.5	19.9
CmNm1	BMS1-like	12398	1859	1859	1.00	1590.7	26.3
CmNm1	CMESO_33	669	1469	1514	0.97	105.4	18.3
CmNm1	Gt-orf272	1338	1105	1112	0.99	287.0	20.8
CmNm1	tic22	11609	983	983	1.00	2816.7	29.0
CmNm1	clpP1	72427	794	794	1.00	21756.2	36.7
CmNm1	rps5	5516	572	575	0.99	2288.0	35.1
CmNm1	CMESO_38	1369	1200	1256	0.96	260.0	21.0
CmNm1	CMESO_39	355	376	404	0.93	209.6	19.0
CmNm1	mcm4	14104	1969	1982	0.99	1697.2	27.1
CmNm1	Ha-orf310	1072	964	971	0.99	263.3	19.4
CmNm1	bysl	761	719	719	1.00	252.4	25.0
CmNm1	U5_snRNP	7443	5243	5300	0.99	334.9	25.1
CmNm1	Ha-orf889	5878	2647	2669	0.99	525.3	22.1
CmNm1	CMESO_45	1351	365	365	1.00	882.8	22.4
CmNm1	Ha-orf90	3476	347	347	1.00	2389.2	29.0
CmNm1	cdc28	5230	2556	2633	0.97	473.8	25.2
CmNm1	sui1	2321	344	344	1.00	1609.2	27.8
CmNm1	tcpT	16376	1595	1595	1.00	2448.8	31.0
CmNm1	taf30	10097	416	416	1.00	5789.0	28.3
CmNm1	Gt-orf102	8529	308	308	1.00	6604.7	28.5
CmNm1	sen2	3229	806	806	1.00	955.5	27.1
CmNm1	snrpD3	1817	251	251	1.00	1726.6	31.0
CmNm1	Ha-orf206	3583	626	626	1.00	1365.1	23.0
CmNm1	cbbX	191990	1202	1202	1.00	38095.8	35.4
CmNm1	rps27	3329	296	296	1.00	2682.4	25.9
CmNm1	rpa1	11591	4538	4538	1.00	609.2	26.7
CmNm1	snu13	2050	410	410	1.00	1192.5	29.0
CmNm1	rps26	3903	320	320	1.00	2909.1	29.3
CmNm1	prp8	11734	6623	6629	1.00	422.2	26.3
CmNm1	tcpZ	8803	1565	1565	1.00	1341.6	30.2
CmNm1	CMESO_64	1253	2539	2567	0.99	116.4	18.3
CmNm1	eif6	4178	698	698	1.00	1427.6	29.3
CmNm1	rps15	4688	491	491	1.00	2277.2	28.9
CmNm1	CMESO_68	1608	506	506	1.00	757.9	22.7
CmNm1	Gt-orf938	6874	3605	3605	1.00	454.8	22.2
CmNm1	cdc28	4154	1889	1889	1.00	524.5	28.6

CmNm1	rpl13	5201	389	389	1.00	3188.9	28.2
CmNm1	sen34	1659	772	776	0.99	509.9	23.8
CmNm1	Ha-orf868	3531	2738	2738	1.00	307.6	21.4
CmNm1	tubA	16597	1346	1346	1.00	2941.0	38.0
CmNm1	rps13	6048	455	455	1.00	3170.3	28.3
CmNm1	rps28	3797	197	197	1.00	4597.0	33.8
CmNm1	Ha-orf293	745	871	887	0.98	200.3	21.4
CmNm1	ubiquitin	14939	236	236	1.00	15097.8	35.9
CmNm1	CMESO_80	70	582	626	0.93	26.7	18.5
CmNm1	rpb4	1306	413	413	1.00	754.2	25.8
CmNm1	CMESO_82	109	281	281	1.00	92.5	20.9
CmNm1	Gt-orf266	3765	842	842	1.00	1066.5	30.8
CmNm1	kin(cdc)	6048	902	902	1.00	1599.2	32.6
CmNm1	eif2G	12114	1304	1304	1.00	2215.7	32.6
CmNm1	rrp3	7812	1196	1196	1.00	1557.9	30.7
CmNm1	nmt1	1823	1061	1061	1.00	409.8	24.2
CmNm1	cpn60	125904	1841	1841	1.00	16311.3	35.5
CmNm1	rpl27A	5908	431	431	1.00	3269.4	31.0
CmNm1	sys1	9248	1304	1304	1.00	1691.5	28.5
CmNm1	gidA	17798	1952	1952	1.00	2174.7	32.4
CmNm1	rps3A	6264	692	692	1.00	2159.0	26.8
CmNm1	rub	9703	593	593	1.00	3902.6	30.0
CmNm1	Gt-orf477	4835	1466	1466	1.00	786.6	25.8
CmNm1	Ha-orf821	36506	4856	4856	1.00	1793.0	22.3
CmNm1	rpl10	6673	653	653	1.00	2437.3	33.9
CmNm1	CMESO_97	454	775	788	0.98	137.4	19.1
CmNm1	Gt-cdc28	10416	1859	1859	1.00	1336.4	28.3
CmNm1	mrs2	3227	1172	1172	1.00	656.7	32.5
CmNm1	CMESO_100	1148	1820	1883	0.97	145.4	19.2
CmNm1	rla0	12421	938	938	1.00	3158.3	28.8
CmNm1	Gt-orf419a	1170	1490	1490	1.00	187.3	22.6
CmNm1	rps23	7105	437	437	1.00	3877.8	37.4
CmNm1	rpl15	12180	614	614	1.00	4731.3	34.0
CmNm1	snrpE	1480	248	248	1.00	1423.4	27.3
CmNm1	ubiquitin	12161	236	236	1.00	12290.2	35.9
CmNm1	CMESO_108	407	250	251	1.00	386.7	29.0
CmNm1	snrpB	1086	257	257	1.00	1007.9	32.2
CmNm1	hsp90	391787	2138	2138	1.00	43706.4	31.0
CmNm1	kin(mps1)	8766	1800	1814	0.99	1152.6	22.8
CmNm1	dib1	1020	407	407	1.00	597.7	24.0
CmNm1	Ha-orf1234	3100	3610	3902	0.93	189.5	20.7
CmNm1	Cp-3gp362	5646	1217	1217	1.00	1106.5	23.1
CmNm1	CMESO_117	672	185	185	1.00	866.4	22.0
CmNm1	CMESO_118	128	407	407	1.00	75.0	17.4
CmNm1	rpc9	1171	302	302	1.00	924.8	25.1
CmNm1	rpl27	4463	467	467	1.00	2279.4	28.0
CmNm1	cenp-A	8872	401	401	1.00	5276.9	31.6
CmNm1	rpb11	2056	338	338	1.00	1450.8	28.3
CmNm1	CMESO_128	755	329	329	1.00	547.3	21.5
CmNm1	Gt-orf143	417	459	476	0.96	208.9	21.6
CmNm1	sut	19297	2378	2378	1.00	1935.4	30.1
CmNm1	eif4A	9223	1211	1211	1.00	1816.5	30.2
CmNm1	ufd	5566	881	881	1.00	1506.9	32.3
CmNm1	Gt-orf160	26368	566	566	1.00	11111.3	33.7
CmNm1	rcl1	1675	1043	1043	1.00	383.0	26.1
CmNm1	dhm	11982	1649	1649	1.00	1733.1	29.6
CmNm1	ATP/GTPbp	1560	1061	1061	1.00	350.7	29.2
CmNm1	eif4E	6538	557	557	1.00	2799.6	29.2
CmNm1	rpl7A	14988	701	701	1.00	5099.5	26.6
CmNm1	rpl18	4326	446	446	1.00	2313.4	28.4
CmNm1	clpP2	48272	704	704	1.00	16354.1	36.5
CmNm1	CMESO_143	2962	1583	1598	0.99	442.1	20.5
CmNm1	cdc48	6398	1967	1967	1.00	775.8	30.3
CmNm1	rpl19	7688	569	569	1.00	3222.6	28.6
CmNm1	snrpG	3618	227	227	1.00	3801.4	29.8

CmNm1	taf13	3704	542	542	1.00	1630.0	25.0
CmNm1	Ha-orf244	2503	722	722	1.00	826.9	23.5
CmNm1	rfc2	4075	956	956	1.00	1016.7	26.5
CmNm1	Ha-orf732	2227	2200	2213	0.99	240.0	20.5
CmNm1	rpl12	5285	470	470	1.00	2681.9	34.0
CmNm1	mcm6	6545	2372	2372	1.00	658.1	26.0
CmNm1	rpl34	3442	347	347	1.00	2365.8	22.4
CmNm1	Ha-orf512	1417	1493	1508	0.99	224.1	20.1
CmNm1	Gt-orf180	2315	1392	1394	1.00	396.1	22.7
CmNm1	CMESO_158	913	779	848	0.92	256.8	19.4
CmNm1	U5_snRNP	7239	2693	2693	1.00	641.1	29.3
CmNm1	CMESO_160	10888	1298	1298	1.00	2000.7	28.4
CmNm1	smc2	7425	3214	3371	0.95	525.3	23.3
CmNm1	rpl7	4972	743	743	1.00	1596.0	27.8
CmNm1	CMESO_164	795	653	653	1.00	290.4	22.0
CmNm1	CMESO_165	1389	515	515	1.00	643.3	21.9
CmNm1	ebi	11609	1376	1376	1.00	2012.2	27.3
CmNm1	rps3	4149	653	653	1.00	1515.4	33.9
CmNm1	rpl1	12350	875	875	1.00	3366.4	32.8
CmNm1	tfIIB	1625	1031	1031	1.00	375.9	32.4
CmNm1	CMESO_170	31	242	269	0.90	27.5	13.7
CmNm1	brx1	1057	718	746	0.96	337.9	22.1
CmNm1	rpl35A	3358	335	335	1.00	2390.8	27.4
CmNm1	CMESO_173	5009	719	719	1.00	1661.6	30.0
CmNm1	Ha-orf1040	3348	2998	3104	0.97	257.3	20.5
CmNm1	Ha-orf62	3653	191	191	1.00	4561.6	30.2
CmNm1	rps30	1694	170	170	1.00	2376.7	32.7
CmNm1	CMESO_177	6618	1352	1352	1.00	1167.5	25.9
CmNm1	CMESO_178	2955	1292	1292	1.00	545.5	32.8
CmNm1	Ha-orf253	1523	794	794	1.00	457.5	27.2
CmNm1	CMESO_180	4080	650	650	1.00	1497.1	32.0
CmNm1	CMESO_181	2831	659	659	1.00	1024.6	31.4
CmNm1	snrpF	2117	299	299	1.00	1688.7	36.3
CmNm1	dbp4	7311	1490	1490	1.00	1170.3	33.9
CmNm1	CMESO_185	950	344	344	1.00	658.7	22.6
CmNm1	CMESO_187	344	497	497	1.00	165.1	30.1
CmNm1	tfIIB-brf	1184	1283	1283	1.00	220.1	37.7
CmNm1	CMESO_189	2794	443	620	0.71	1074.8	27.5
CmNm2	CMESO_198	1356	376	620	0.61	521.6	27.9
CmNm2	tfIIB-brf	1126	1283	1283	1.00	209.3	37.7
CmNm2	CMESO_200	319	494	497	0.99	153.1	30.1
CmNm2	CMESO_202	1019	344	344	1.00	706.5	22.6
CmNm2	dbp4	7552	1490	1490	1.00	1208.9	33.9
CmNm2	snrpF	2011	299	299	1.00	1604.1	36.3
CmNm2	CMESO_206	1836	659	659	1.00	664.5	31.5
CmNm2	CMESO_207	4134	650	650	1.00	1516.9	32.1
CmNm2	Ha-orf253	1495	794	794	1.00	449.1	27.2
CmNm2	CMESO_209	3090	1292	1292	1.00	570.4	32.8
CmNm2	CMESO_210	2807	449	461	0.97	1452.3	31.4
CmNm2	uceE2	4760	440	440	1.00	2580.2	32.7
CmNm2	rpl8	10093	779	779	1.00	3090.2	39.0
CmNm2	Ha-orf102	1349	317	317	1.00	1015.0	25.2
CmNm2	CMESO_215	5932	764	764	1.00	1851.9	25.1
CmNm2	gblp1	5837	2416	2444	0.99	569.6	24.5
CmNm2	CMESO_217	1746	284	284	1.00	1466.3	18.9
CmNm2	rpl31	3315	230	230	1.00	3437.6	29.0
CmNm2	rpl11B	5561	524	524	1.00	2531.2	33.0
CmNm2	Ha-orf269	1657	812	812	1.00	486.7	23.7
CmNm2	Ha-orf893	2748	2437	2564	0.95	255.6	19.9
CmNm2	mcm8	8077	1898	1898	1.00	1015.0	24.7
CmNm2	CMESO_223	381	179	179	1.00	507.7	25.6
CmNm2	mcm9	2107	2030	2030	1.00	247.6	24.5
CmNm2	CMESO_225	773	935	935	1.00	197.2	20.5
CmNm2	CMESO_226	761	486	491	0.99	369.7	19.1
CmNm2	CMESO_227	6291	653	653	1.00	2297.8	31.0

CmNm2	imp4	1452	635	635	1.00	545.4	28.5
CmNm2	Ha-orf421	3002	1974	1979	1.00	361.8	21.9
CmNm2	tbl3	6824	2372	2372	1.00	686.2	24.7
CmNm2	rpl18A	9935	530	530	1.00	4470.9	32.4
CmNm2	rps11	4707	560	560	1.00	2004.7	31.7
CmNm2	Gt-orf112	2051	374	374	1.00	1308.0	20.8
CmNm2	CMESO_234	235	1013	1097	0.92	51.1	18.2
CmNm2	CMESO_235	98	173	173	1.00	135.1	23.6
CmNm2	Gt-orf228	2767	758	758	1.00	870.6	23.5
CmNm2	Ha-orf134	313	410	410	1.00	182.1	18.2
CmNm2	Ha-orf194-436	3279	2052	2054	1.00	380.8	22.7
CmNm2	Gt-orf197	40826	617	617	1.00	15781.7	32.0
CmNm2	CMESO_240	1203	1247	1247	1.00	230.1	18.6
CmNm2	Gt-orf115	1253	374	374	1.00	799.1	24.8
CmNm2	mcm5	5125	1850	1850	1.00	660.7	25.3
CmNm2	H4	4592	392	392	1.00	2794.0	38.4
CmNm2	H3	5282	464	464	1.00	2715.1	35.1
CmNm2	rpb2	9760	3668	3668	1.00	634.6	33.9
CmNm2	CMESO_246	3187	2566	2585	0.99	294.1	19.6
CmNm2	rpl23	9537	422	422	1.00	5390.2	39.2
CmNm2	CMESO_248	87	578	647	0.89	32.1	16.4
CmNm2	Gt-orf187	7780	707	707	1.00	2624.6	33.9
CmNm2	CMESO_250	2198	224	224	1.00	2340.4	14.2
CmNm2	hfc136	41313	1217	1217	1.00	8096.5	34.6
CmNm2	ATPbp	1614	791	791	1.00	486.7	24.6
CmNm2	rps24	5402	395	395	1.00	3261.8	28.3
CmNm2	Gt-orf301	15839	971	971	1.00	3890.6	26.0
CmNm2	rpb8	8330	503	503	1.00	3949.8	27.2
CmNm2	CMESO_257	14402	1343	1343	1.00	2557.7	26.8
CmNm2	eif5A	7388	473	473	1.00	3725.4	32.5
CmNm2	Cp-2gp261	13383	3470	3470	1.00	919.9	24.4
CmNm2	der1	6798	647	647	1.00	2506.0	30.1
CmNm2	Ha-orf390	4263	1178	1178	1.00	863.1	24.4
CmNm2	CMESO_262	2762	290	290	1.00	2271.6	21.6
CmNm2	kea1	29336	2213	2213	1.00	3161.7	34.7
CmNm2	CMESO_264	191	179	179	1.00	254.5	18.3
CmNm2	rps10B	1935	287	287	1.00	1608.1	22.9
CmNm2	rpl26	3600	374	374	1.00	2295.8	29.3
CmNm2	hsp70	471679	1955	1955	1.00	57544.4	34.6
CmNm2	nip7	2295	536	536	1.00	1021.2	27.6
CmNm2	rpc1	8656	3986	3986	1.00	517.9	28.5
CmNm2	CMESO_270	823	653	653	1.00	300.6	24.2
CmNm2	Ha-orf155	119	503	515	0.98	55.1	20.7
CmNm2	fcf1	3106	437	437	1.00	1695.2	25.8
CmNm2	mcm7	23820	2174	2174	1.00	2613.3	25.0
CmNm2	CMESO_275	465	1091	1091	1.00	101.7	19.9
CmNm2	dph1	5262	1271	1271	1.00	987.4	25.9
CmNm2	asf	1165	473	473	1.00	587.4	22.2
CmNm2	CMESO_278	1759	2214	2237	0.99	187.5	19.5
CmNm2	Ha-orf530	1259	1556	1577	0.99	190.4	21.2
CmNm2	impA	7667	1559	1559	1.00	1173.0	31.3
CmNm2	erf1	13701	1277	1277	1.00	2559.0	29.6
CmNm2	tcpE	6224	1601	1601	1.00	927.2	34.5
CmNm2	CMESO_284	115	227	227	1.00	120.8	22.8
CmNm2	Gt-orf187	8087	800	800	1.00	2411.0	33.1
CmNm2	tfIIA-S	4012	359	359	1.00	2665.4	30.0
CmNm2	Gt-orf446	2910	1838	1910	0.96	363.4	19.2
CmNm2	Gt-orf365	4210	1253	1253	1.00	801.4	24.2
CmNm2	tha4	18822	572	572	1.00	7848.3	31.6
CmNm2	rpabc6	2682	401	401	1.00	1595.2	27.6
CmNm2	rpabc5	2561	650	650	1.00	939.7	31.2
CmNm2	met	7424	1394	1394	1.00	1270.2	27.0
CmNm2	smc3	8300	3219	3278	0.98	603.9	20.7
CmNm2	sufD	12264	1460	1460	1.00	2003.5	29.6
CmNm2	tubG	5761	1310	1310	1.00	1048.9	30.3

CmNm2	Ha-orf108	2153	329	329	1.00	1560.8	36.7
CmNm2	CMESO_297	473	904	923	0.98	122.2	18.3
CmNm2	CMESO_298	14	194	278	0.70	12.0	19.0
CmNm2	nog1	3570	1379	1379	1.00	617.5	25.1
CmNm2	Gt-orf82	783	395	395	1.00	472.8	27.8
CmNm2	Gt-orf163	8238	533	533	1.00	3686.4	30.1
CmNm2	rpl9	6262	557	557	1.00	2681.4	29.7
CmNm2	Gt-orf227	2456	731	731	1.00	801.3	23.9
CmNm2	taf90	3164	1553	1553	1.00	485.9	26.0
CmNm2	Ha-orf373	4631	1184	1184	1.00	932.9	23.0
CmNm2	CMESO_308	3141	443	443	1.00	1691.1	24.3
CmNm2	Gt-orf193	14893	653	653	1.00	5439.7	28.6
CmNm2	tcpA	11272	1604	1604	1.00	1676.1	31.8
CmNm2	Ha-orf132	1085	410	410	1.00	631.2	23.4
CmNm2	U5snRNP	5517	905	905	1.00	1454.0	29.6
CmNm2	rps15A	5362	392	392	1.00	3262.5	31.6
CmNm2	snrpD2	4033	257	257	1.00	3742.8	27.1
CmNm2	tfIIE	2626	665	665	1.00	941.8	24.0
CmNm2	CMESO_316	1005	671	671	1.00	357.2	20.8
CmNm2	CMESO_317	619	599	599	1.00	246.5	19.8
CmNm2	CMESO_318	450	1437	1505	0.95	71.3	19.9
CmNm2	engA	18007	1589	1589	1.00	2702.8	31.9
CmNm2	Cp-2gp291	4017	2542	2546	1.00	376.3	21.8
CmNm2	CMESO_321	811	1006	1103	0.91	175.4	18.8
CmNm2	secE	6136	401	401	1.00	3649.6	23.4
CmNm2	rip1	4931	848	848	1.00	1386.9	26.4
CmNm2	rpl24	1236	197	197	1.00	1496.4	20.7
CmNm2	has1	7184	1424	1424	1.00	1203.3	27.9
CmNm2	nop5	3084	1172	1172	1.00	627.6	28.0
CmNm2	CMESO_328	6514	653	653	1.00	2379.2	31.0
CmNm2	Gt-orf532	39249	1949	1949	1.00	4803.1	30.4
CmNm2	CMESO_330	354	519	641	0.81	131.7	17.6
CmNm2	Ha-orf154	1970	536	536	1.00	876.6	25.3
CmNm2	kin(snf1)	7480	1457	1457	1.00	1224.5	32.4
CmNm2	pi4K	1909	1827	1850	0.99	246.1	22.4
CmNm2	Gt-orf534	31109	866	866	1.00	8567.8	33.1
CmNm2	rps15	47305	389	389	1.00	29004.2	29.7
CmNm2	smc1	7443	3437	3461	0.99	512.9	22.9
CmNm2	Gt-orf365	2014	998	998	1.00	481.3	21.8
CmNm2	Ha-orf831	7971	2474	2480	1.00	766.6	23.5
CmNm2	CMESO_339	1956	1343	1343	1.00	347.4	19.8
CmNm2	Ha-orf106	3650	362	362	1.00	2404.8	31.1
CmNm2	Gt-orf139	19259	473	473	1.00	9711.3	34.2
CmNm2	CMESO_342	3372	2162	2162	1.00	372.0	19.1
CmNm2	CMESO_343	978	335	335	1.00	696.3	19.0
CmNm2	Gt-orf125	3028	458	458	1.00	1576.9	25.7
CmNm2	dnaG	3805	2207	2207	1.00	411.2	24.4
CmNm2	Gt-orf331	1085	1256	1313	0.96	197.1	21.4
CmNm2	rpl36	4283	245	245	1.00	4169.5	24.8
CmNm2	rps17	6104	338	338	1.00	4307.3	30.4
CmNm2	rpl21	9799	479	479	1.00	4879.2	30.8
CmNm2	rps19	2720	425	425	1.00	1526.5	30.8
CmNm2	CMESO_351	451	941	941	1.00	114.3	18.8
CmNm2	iap100	62701	3368	3368	1.00	4440.2	27.5
CmNm2	imb1	9414	2576	2576	1.00	871.6	28.1
CmNm2	taf	5100	1310	1310	1.00	928.5	24.6
CmNm2	Ha-orf121	971	371	371	1.00	624.2	24.5
CmNm2	rsp4	5660	629	629	1.00	2146.2	34.4
CmNm2	Ha-orf249	1427	767	767	1.00	443.7	22.8
CmNm2	rpb3	6652	914	914	1.00	1735.8	27.4
CmNm2	CMESO_359	1085	890	890	1.00	290.8	18.5
CmNm2	CMESO_360	877	1304	1331	0.98	157.2	20.9
CmNm2	CMESO_361	471	529	533	0.99	210.8	21.3
CmNm2	rps8	22190	1127	1127	1.00	4696.1	31.7
CmNm2	snrpD	280	239	239	1.00	279.4	23.3

CmNm2	ggt	2421	1175	1178	1.00	490.2	23.3
CmNm2	smc4	8231	3527	3557	0.99	551.9	22.9
CmNm2	hda	2562	1130	1130	1.00	540.8	28.5
CmNm2	CMESO_367	1164	716	716	1.00	387.7	22.2
CmNm2	hlip	9394	380	380	1.00	5896.2	27.0
CmNm2	CMESO_370	267	266	266	1.00	239.4	19.9
CmNm2	rpl6B	4397	599	599	1.00	1750.8	24.3
CmNm2	CMESO_373	1608	2909	2933	0.99	130.8	19.4
CmNm2	CMESO_374	2115	659	659	1.00	765.5	31.5
CmNm2	snrpF	2099	299	299	1.00	1674.3	36.3
CmNm2	dbp4	7730	1490	1490	1.00	1237.4	33.9
CmNm2	CMESO_378	1272	344	344	1.00	881.9	22.6
CmNm2	CMESO_380	315	497	497	1.00	151.2	30.1
CmNm2	tfIIB-brf	1192	1283	1283	1.00	221.6	37.7
CmNm2	CMESO_382	1340	377	620	0.61	515.5	27.9
CmNm3	CMESO_391	1372	379	620	0.61	527.8	27.9
CmNm3	tfIIB-brf	1085	1283	1283	1.00	201.7	37.7
CmNm3	CMESO_393	311	497	497	1.00	149.2	30.1
CmNm3	CMESO_395	1199	344	344	1.00	831.3	22.6
CmNm3	dbp4	7517	1490	1490	1.00	1203.3	33.9
CmNm3	snrpF	2058	299	299	1.00	1641.6	36.3
CmNm3	CMESO_399	2182	659	659	1.00	789.7	31.5
CmNm3	CMESO_400	3987	650	650	1.00	1463.0	32.1
CmNm3	CMESO_401	1466	794	794	1.00	440.4	27.2
CmNm3	CMESO_402	3098	1292	1292	1.00	571.9	32.8
CmNm3	CMESO_403	6542	1352	1352	1.00	1154.1	25.9
CmNm3	rps30	1707	170	170	1.00	2394.9	32.7
CmNm3	pcna	4923	779	779	1.00	1507.3	31.2
CmNm3	rad25	5633	1913	1913	1.00	702.3	29.3
CmNm3	Ha-orf239	319	676	722	0.94	105.4	21.0
CmNm3	rad51	4481	995	995	1.00	1074.1	36.8
CmNm3	CMESO_409	671	1521	1547	0.98	103.5	19.8
CmNm3	Gt-orf357	4310	1109	1109	1.00	926.9	29.1
CmNm3	Gt-orf249	36857	806	806	1.00	10906.6	28.6
CmNm3	eif1A	4717	443	443	1.00	2539.6	30.9
CmNm3	CMESO_413	359	828	860	0.96	99.6	18.5
CmNm3	CMESO_414	175	844	851	0.99	49.0	19.1
CmNm3	Gt-orf168	8983	473	473	1.00	4529.6	30.8
CmNm3	CMESO_416	842	629	644	0.98	311.8	18.1
CmNm3	pop2	1217	773	773	1.00	375.5	26.6
CmNm3	gblp2	10865	944	944	1.00	2745.1	34.4
CmNm3	rpl24	1318	368	368	1.00	854.2	24.9
CmNm3	CMESO_421	864	230	230	1.00	896.0	24.2
CmNm3	rpl14	4736	410	410	1.00	2755.1	27.0
CmNm3	CMESO_423	134	179	179	1.00	178.5	18.3
CmNm3	Ha-orf154	3035	458	458	1.00	1580.5	22.2
CmNm3	nol10	5974	1037	1037	1.00	1374.0	29.0
CmNm3	pab2	3051	1629	1637	1.00	444.5	23.6
CmNm3	trf	2220	1150	1160	0.99	456.5	24.2
CmNm3	rpl23A	1734	317	317	1.00	1304.6	23.3
CmNm3	CMESO_429	1853	1434	1478	0.97	299.0	18.5
CmNm3	Ha-orf590	2238	1775	1775	1.00	300.7	21.1
CmNm3	tcpG	6324	1535	1535	1.00	982.6	29.4
CmNm3	tif211	5039	800	800	1.00	1502.3	27.8
CmNm3	ftsZ	8513	1232	1232	1.00	1648.1	38.5
CmNm3	rps4	5116	752	752	1.00	1622.6	29.3
CmNm3	snrpD2	915	248	248	1.00	880.0	24.1
CmNm3	mcm3	13359	2177	2177	1.00	1463.6	24.7
CmNm3	tfIID	7344	788	788	1.00	2222.8	28.9
CmNm3	Ha-orf221	4711	707	707	1.00	1589.3	22.6
CmNm3	GTP-bp	6855	929	929	1.00	1759.9	27.3
CmNm3	rpb1	15738	4421	4421	1.00	849.0	32.2
CmNm3	Ha-orf146	1683	398	398	1.00	1008.6	25.3
CmNm3	rps25	1015	269	269	1.00	899.9	20.0
CmNm3	Ha-orf522	4590	1598	1598	1.00	685.1	22.3

CmNm3	Gt-orf228	10689	716	716	1.00	3560.6	33.9
CmNm3	eif2B	8156	506	506	1.00	3844.4	29.6
CmNm3	gyrB	12101	2064	2072	1.00	1392.9	31.6
CmNm3	CMESO_447	849	549	575	0.95	352.2	20.7
CmNm3	Gt-orf755	6651	2453	2453	1.00	646.7	26.0
CmNm3	Ha-orf308	3218	926	926	1.00	828.9	26.9
CmNm3	CMESO_451	3132	848	848	1.00	880.9	21.9
CmNm3	Ha-orf479	1997	1445	1466	0.99	324.9	22.3
CmNm3	snrpD1	1408	272	272	1.00	1234.6	21.6
CmNm3	rps2	11313	731	731	1.00	3691.2	35.0
CmNm3	CMESO_457	2745	794	794	1.00	824.6	22.1
CmNm3	CMESO_458	58	292	302	0.97	45.8	20.1
CmNm3	rpl5	8494	722	722	1.00	2805.9	30.6
CmNm3	CMESO_460	539	446	446	1.00	288.2	19.9
CmNm3	mce	1103	1061	1061	1.00	247.9	21.8
CmNm3	rpl32	8495	371	371	1.00	5461.3	28.5
CmNm3	U3snoRNP	16295	1478	1478	1.00	2629.6	24.8
CmNm3	Ha-orf178	3707	539	539	1.00	1640.4	26.3
CmNm3	tubB	16676	1325	1325	1.00	3001.8	37.5
CmNm3	nop56	3961	1195	1196	1.00	789.9	27.3
CmNm3	tcpB	14012	1517	1517	1.00	2203.0	32.9
CmNm3	rps14	6396	428	428	1.00	3564.2	39.2
CmNm3	rps20	4495	377	377	1.00	2843.8	27.0
CmNm3	Gt-orf387	16869	1277	1277	1.00	3150.7	31.7
CmNm3	rps16	9911	428	428	1.00	5523.0	32.4
CmNm3	mak16	13501	656	656	1.00	4908.7	26.5
CmNm3	rpl3	12383	1166	1166	1.00	2533.0	32.4
CmNm3	Ha-orf255	1100	776	776	1.00	338.1	20.6
CmNm3	rps9	8503	551	551	1.00	3680.6	31.7
CmNm3	CMESO_480	842	728	728	1.00	275.9	22.1
CmNm3	mcm2	3531	2451	2459	1.00	342.5	25.7
CmNm3	CMESO_482	765	1500	1931	0.78	94.5	17.1
CmNm3	rpl0A	6127	653	653	1.00	2237.9	27.5
CmNm3	rpl17	11256	485	485	1.00	5535.4	34.6
CmNm3	Ha-orf336	1655	1004	1004	1.00	393.2	22.1
CmNm3	nop2	2431	1103	1103	1.00	525.7	28.5
CmNm3	sbp1	1797	929	929	1.00	461.4	26.5
CmNm3	rpl37A	6423	278	278	1.00	5510.6	36.2
CmNm3	Gt-orf272	2976	918	926	0.99	766.5	25.8
CmNm3	G10	1611	758	758	1.00	506.9	21.2
CmNm3	cdc48b	95245	2291	2291	1.00	9915.6	35.0
CmNm3	Gt-orf456	3170	1373	1373	1.00	550.7	24.8
CmNm3	CMESO_494	841	1288	1325	0.97	151.4	18.2
CmNm3	Gt-orf187	8276	800	800	1.00	2467.4	33.0
CmNm3	Ha-orf65	328	197	197	1.00	397.1	23.7
CmNm3	Gt-orf497	6577	1121	1121	1.00	1399.3	31.4
CmNm3	sen1	9074	2261	2261	1.00	957.2	27.1
CmNm3	CMESO_499	4911	554	554	1.00	2114.3	30.6
CmNm3	CMESO_500	1632	749	749	1.00	519.7	18.7
CmNm3	BRSK	8053	1973	1973	1.00	973.5	24.3
CmNm3	CMESO_502	223	428	428	1.00	124.3	18.4
CmNm3	rfc3	1533	842	842	1.00	434.2	24.3
CmNm3	Ha-orf448	5975	1258	1292	0.97	1103.0	23.9
CmNm3	rps21	4152	260	260	1.00	3808.8	29.1
CmNm3	rpl40	3713	179	179	1.00	4947.4	35.0
CmNm3	CMESO_507	942	1249	1250	1.00	179.7	20.0
CmNm3	Gt-orf323	11686	1010	1010	1.00	2759.6	28.5
CmNm3	CMESO_509	208	563	605	0.93	82.0	18.8
CmNm3	prl1	2445	989	989	1.00	589.6	31.1
CmNm3	CMESO_511	1747	947	947	1.00	440.0	20.0
CmNm3	CMESO_512	651	422	422	1.00	367.9	23.4
CmNm3	Ha-orf338	3033	1010	1010	1.00	716.2	20.0
CmNm3	sof1	10980	1262	1262	1.00	2075.1	29.9
CmNm3	CMESO_515	101	218	218	1.00	110.5	20.5
CmNm3	cdc2	5374	1022	1022	1.00	1254.2	27.7

CmNm3	kin(ABC1)	11755	2618	2618	1.00	1070.9	28.8
CmNm3	rpa5	2925	1052	1052	1.00	663.2	27.8
CmNm3	ubc2	1434	458	458	1.00	746.8	31.4
CmNm3	Gt-orf258	4591	944	944	1.00	1159.9	24.4
CmNm3	fkbp	14544	875	875	1.00	3964.4	29.2
CmNm3	CMESO_522	1098	884	884	1.00	296.2	17.5
CmNm3	rpa2	9998	3383	3383	1.00	704.9	30.8
CmNm3	Ha-orf803	3435	2420	2420	1.00	338.5	23.3
CmNm3	Gt-orf446	2647	1821	1904	0.96	331.6	19.2
CmNm3	tfIIA-S	2537	353	353	1.00	1714.2	30.5
CmNm3	Gt-orf187	7987	800	800	1.00	2381.2	33.1
CmNm3	rpoD	18044	1457	1457	1.00	2953.8	35.5
CmNm3	Ha-orf193	1121	599	599	1.00	446.4	23.2
CmNm3	Ha-orf445	1860	1277	1340	0.95	331.1	19.5
CmNm3	gsp2	22363	641	641	1.00	8321.0	34.6
CmNm3	rpl30	8085	329	329	1.00	5861.2	30.9
CmNm3	rpc10	758	134	134	1.00	1349.2	31.9
CmNm3	rpf1	1570	566	566	1.00	661.6	21.5
CmNm3	rps27A	4759	431	431	1.00	2633.6	27.3
CmNm3	reb1	2962	932	932	1.00	758.0	32.7
CmNm3	rad3	4065	2279	2279	1.00	425.4	23.1
CmNm3	rpb7	2932	524	524	1.00	1334.6	28.6
CmNm3	rpc2	23802	3290	3290	1.00	1725.5	30.4
CmNm3	Ha-orf203	461	614	614	1.00	179.1	20.0
CmNm3	tcpH	11089	1592	1592	1.00	1661.3	32.7
CmNm3	rpl13A	4330	590	590	1.00	1750.4	31.1
CmNm3	Ha-orf483	1710	1487	1487	1.00	274.3	22.1
CmNm3	rrp3	1353	1232	1232	1.00	261.9	25.0
CmNm3	pab1	8375	1472	1472	1.00	1357.0	25.5
CmNm3	CMESO_550	954	965	965	1.00	235.8	18.2
CmNm3	Ha-orf88	953	296	296	1.00	767.9	19.5
CmNm3	CMESO_552	2873	683	683	1.00	1003.3	24.1
CmNm3	h2B	4096	374	374	1.00	2612.1	30.4
CmNm3	CMESO_554	2126	1412	1412	1.00	359.1	21.9
CmNm3	ef2	35495	2546	2546	1.00	3325.2	37.0
CmNm3	Ha-orf154	2907	452	452	1.00	1533.9	25.6
CmNm3	CMESO_557	1832	1460	1460	1.00	299.3	19.4
CmNm3	rpb10	1440	230	230	1.00	1493.3	29.4
CmNm3	CMESO_559	8892	6255	6296	0.99	336.9	19.3
CmNm3	nop1	7840	824	824	1.00	2269.3	31.0
CmNm3	cbf5	3520	1127	1127	1.00	744.9	25.9
CmNm3	rla1	10816	410	410	1.00	6292.0	30.2
CmNm3	rli1	14886	1790	1790	1.00	1983.5	28.0
CmNm3	gidB	2637	758	758	1.00	829.7	28.3
CmNm3	CMESO_565	7948	2180	2180	1.00	869.6	29.3
CmNm3	Ha-orf253	1738	638	638	1.00	649.7	27.4
CmNm3	CMESO_567	3874	650	650	1.00	1421.5	32.1
CmNm3	CMESO_568	2207	659	659	1.00	798.8	31.5
CmNm3	snrpF	1978	299	299	1.00	1577.8	36.3
CmNm3	dbp4	7555	1490	1490	1.00	1209.3	33.9
CmNm3	CMESO_572	1147	344	344	1.00	795.3	22.6
CmNm3	CMESO_574	298	497	497	1.00	143.0	30.1
CmNm3	tfIIB-brf	1082	1283	1283	1.00	201.1	37.7
CmNm3	CMESO_576	1311	375	620	0.60	504.3	27.9
	Sum	4192729			**Average**	2021.2	27.0

Mapping rates were calculated by gene lengths and mapped regions.
RPKM (Reads Per Kilobase of exon per Million mapped reads) was calculated by the following formula;
=(number of reads for each gene)*1000000/(total number of reads)*1000/(gene length)

Table 4.4 Relative expression of nucleomorph genes in *Guillardia theta*

Chromosome No.	Gene name	Number of reads	Mapped by some reads (bp)	Gene length (bp)	Mapping rate	RPKM	GC content (%)
GtNm1	orf110	32	293	332	0.88	47.2	52.9
GtNm1	orf210	115	632	632	1.00	89.1	53.7
GtNm1	orf151	65	448	455	0.98	69.9	41.9
GtNm1	orf65	80	197	197	1.00	198.8	44.4
GtNm1	orf69	104	209	209	1.00	243.6	55.7
GtNm1	ubc4	620	443	443	1.00	685.3	52.9
GtNm1	hsp82	132027	2054	2054	1.00	31472.9	29.8
GtNm1	orf119	316	354	359	0.99	431.0	17.5
GtNm1	orf470	512	1291	1412	0.91	177.5	15.3
GtNm1	tubG	5195	1274	1274	1.00	1996.6	25.3
GtNm1	orf113	204	341	341	1.00	292.9	19.6
GtNm1	ski2	837	2103	2159	0.97	189.8	20.6
GtNm1	tcpE	6555	1535	1535	1.00	2090.9	28.4
GtNm1	rpl13A	2331	584	584	1.00	1954.4	26.3
GtNm1	orf433	604	1183	1301	0.91	227.3	17.4
GtNm1	pab1	2400	1163	1169	0.99	1005.2	22.6
GtNm1	orf291	151	767	875	0.88	84.5	16.3
GtNm1	orf419	67	907	1259	0.72	26.1	17.6
GtNm1	orf177	10	224	533	0.42	9.2	14.2
GtNm1	rpoD	4649	1382	1382	1.00	1647.1	31.5
GtNm1	kea1	31080	2180	2180	1.00	6980.7	32.7
GtNm1	orf341	1403	995	1025	0.97	670.2	17.1
GtNm1	der1	3071	605	605	1.00	2485.4	26.6
GtNm1	sf3b_3	792	2617	3185	0.82	121.8	18.1
GtNm1	eif5A	1822	476	476	1.00	1874.2	27.9
GtNm1	rpl23	4030	422	422	1.00	4675.9	35.2
GtNm1	orf519	653	1270	1559	0.81	205.1	16.6
GtNm1	rpb2	4166	3620	3620	1.00	563.5	29.2
GtNm1	H3	5576	410	410	1.00	6659.1	33.1
GtNm1	H4	4805	314	314	1.00	7492.7	33.7
GtNm1	mcm?	1436	1586	1598	0.99	440.0	19.9
GtNm1	orf225	676	677	677	1.00	488.9	21.1
GtNm1	orf115	381	347	347	1.00	537.6	22.1
GtNm1	orf236	2115	674	710	0.95	1458.6	19.7
GtNm1	orf197	29077	593	593	1.00	24008.7	29.0
GtNm1	gyrA	12294	2633	2633	1.00	2286.2	27.1
GtNm1	tcpD	3601	1559	1559	1.00	1131.0	26.5
GtNm1	CRM	375	1988	2849	0.70	64.4	18.6
GtNm1	orf216	481	650	650	1.00	362.3	27.8
GtNm1	orf68	533	206	206	1.00	1266.9	20.8
GtNm1	PP1	4400	914	914	1.00	2357.1	28.0
GtNm1	orf176	622	530	530	1.00	574.6	21.7
GtNm1	fet5	4643	761	761	1.00	2987.4	24.8
GtNm1	rpl26	2146	326	326	1.00	3223.2	22.9
GtNm1	rps10B	694	281	281	1.00	1209.3	25.5
GtNm1	clpP1	60576	764	764	1.00	38822.3	34.6
GtNm1	tic22	13522	923	923	1.00	7173.2	26.1
GtNm1	orf272	383	787	818	0.96	229.3	19.0
GtNm1	BMS1-like	1426	1656	1670	0.99	418.1	20.1
GtNm1	prsA2	2434	701	701	1.00	1700.1	26.4
GtNm1	nat10	2112	2576	2579	1.00	401.0	22.7
GtNm1	mcm4	3636	1625	1625	1.00	1095.6	19.4
GtNm1	orf171	813	515	515	1.00	773.0	18.2
GtNm1	prsA5	4088	698	698	1.00	2867.7	27.9
GtNm1	rps5	5900	575	575	1.00	5024.1	28.6
GtNm1	tfIIIC	1015	1745	1997	0.87	248.9	18.4
GtNm1	cycB	715	984	1043	0.94	335.7	21.5
GtNm1	rpl23A	848	353	353	1.00	1176.2	20.9
GtNm1	mcm6	4726	2234	2234	1.00	1035.8	20.4
GtNm1	rpl12	3448	434	434	1.00	3890.0	34.0
GtNm1	orf590	257	1453	1772	0.82	71.0	18.6

GtNm1	rfC	1080	872	872	1.00	606.4	20.6
GtNm1	orf224	9024	674	674	1.00	6555.6	22.5
GtNm1	orf532	15885	1598	1598	1.00	4867.3	22.5
GtNm1	rcl1	659	1001	1001	1.00	322.3	20.1
GtNm1	dhm	3392	1628	1628	1.00	1020.2	28.3
GtNm1	ATP/GTP_bp	536	992	992	1.00	264.6	24.2
GtNm1	eif4E	2508	527	527	1.00	2330.2	27.7
GtNm1	rpl7A	6885	620	620	1.00	5437.3	25.1
GtNm1	rpl18	2663	503	503	1.00	2592.3	22.8
GtNm1	clpP2	50588	701	701	1.00	35334.9	31.8
GtNm1	orf461	1773	1308	1385	0.94	626.8	18.3
GtNm1	cdc48	1678	1785	1820	0.98	451.4	25.3
GtNm1	rpl19	2936	568	572	0.99	2513.2	23.4
GtNm1	snrpG	341	224	224	1.00	745.4	24.4
GtNm1	taf13	380	506	506	1.00	367.7	21.3
GtNm1	asf	236	441	461	0.96	250.7	19.0
GtNm1	dph1	1840	1244	1244	1.00	724.2	23.5
GtNm1	mcm7	1133	1935	2012	0.96	275.7	20.2
GtNm1	fcf1	314	383	383	1.00	401.4	23.4
GtNm1	orf382	490	1140	1148	0.99	209.0	17.1
GtNm1	orf183	768	551	551	1.00	682.5	24.3
GtNm1	eif2G	4335	1475	1475	1.00	1439.0	29.8
GtNm1	orf665	362	1610	1997	0.81	88.8	16.6
GtNm1	orf128	118	379	386	0.98	149.7	19.4
GtNm1	orf248	944	746	746	1.00	619.6	18.3
GtNm1	U5_snRNP	1627	2411	2483	0.97	320.8	21.6
GtNm1	ruvb-like	182	1017	1175	0.87	75.8	17.5
GtNm1	orf308	157	753	926	0.81	83.0	17.8
GtNm1	orf313	63	610	941	0.65	32.8	15.6
GtNm1	prsB6	629	719	719	1.00	428.3	25.6
GtNm1	hsp70	232304	1952	1952	1.00	58270.8	32.4
GtNm1	orf100	4068	302	302	1.00	6595.5	21.8
GtNm1	imp4	677	623	623	1.00	532.1	22.6
GtNm1	orf304	136	726	914	0.79	72.9	16.0
GtNm1	gblp	2792	2042	2042	1.00	669.5	19.5
GtNm1	rpl31	1019	293	293	1.00	1702.9	28.2
GtNm1	rpl11B	2423	524	524	1.00	2264.1	29.7
GtNm1	orf147	88	443	443	1.00	97.3	17.8
GtNm1	mcm?	633	1347	1415	0.95	219.0	18.6
GtNm1	prsB5	1154	617	617	1.00	915.8	31.9
GtNm1	orf636	130	1595	1910	0.84	33.3	18.8
GtNm1	orf244	38	521	734	0.71	25.3	16.1
GtNm1	orf446	201	991	1340	0.74	73.4	17.4
GtNm1	rps25	119	263	263	1.00	221.5	18.9
GtNm1	orf472	921	1340	1418	0.94	318.0	19.2
GtNm1	orf228	6159	686	686	1.00	4396.0	30.9
GtNm1	eif2B	5799	512	512	1.00	5545.7	26.7
GtNm1	gyrB	10376	2054	2054	1.00	2473.5	28.8
GtNm1	orf168	121	493	506	0.97	117.1	18.1
GtNm1	orf755	1695	2267	2267	1.00	366.1	21.9
GtNm1	orf304	577	914	914	1.00	309.1	19.5
GtNm1	hsf	961	557	557	1.00	844.8	24.4
GtNm1	cpeT	3137	668	668	1.00	2299.4	26.3
GtNm1	orf122	508	368	368	1.00	675.9	17.6
GtNm1	orf370	5519	1112	1112	1.00	2430.1	20.9
GtNm1	orf101	4553	305	305	1.00	7309.2	20.6
GtNm1	hlip	7866	383	383	1.00	10056.1	23.2
GtNm1	orf160	6703	482	482	1.00	6809.2	36.9
GtNm1	erf1	2911	1229	1229	1.00	1159.7	28.0
GtNm1	orf313	558	900	941	0.96	290.3	19.6
GtNm1	impA	1942	1538	1544	1.00	615.9	27.8
GtNm1	orf85	49	212	257	0.82	93.4	14.7
GtNm1	sufD	11269	1403	1403	1.00	3932.8	24.4
GtNm1	orf1019	6930	2970	3059	0.97	1109.2	19.1
GtNm1	met	3716	1382	1382	1.00	1316.6	24.3

GtNm1	rpabc5	1864	644	644	1.00	1417.2	27.3
GtNm1	rpabc6	4996	380	380	1.00	6437.4	29.9
GtNm1	tha4	4241	227	227	1.00	9147.8	27.2
GtNm1	orf443	373	1272	1331	0.96	137.2	18.8
GtNm1	orf365	983	1097	1097	1.00	438.8	22.1
GtNm1	prsA6	690	707	707	1.00	477.9	20.6
GtNm1	rpc1	3162	3869	3869	1.00	400.2	25.6
GtNm1	nip7	489	602	602	1.00	397.7	21.6
GtNm1	rps15	1664	419	419	1.00	1944.5	28.3
GtNm1	eif6	3107	686	686	1.00	2217.6	26.8
GtNm1	orf222	4413	668	668	1.00	3234.7	32.7
GtNm1	orf262	144	682	788	0.87	89.5	17.1
GtNm1	orf124	34	343	374	0.92	44.5	17.6
GtNm1	orf773	259	1633	2321	0.70	54.6	17.0
GtNm1	orf365	144	827	1097	0.75	64.3	19.1
GtNm1	smc1	1751	2619	3017	0.87	284.2	20.0
GtNm1	rps15	9054	380	380	1.00	11666.2	24.1
GtNm1	orf534	40336	1604	1604	1.00	12313.0	26.8
GtNm1	PI4K	6465	1455	1565	0.93	2022.7	20.3
GtNm1	kin(snf1)	1463	1418	1418	1.00	505.2	27.8
GtNm1	orf714	212	1292	2144	0.60	48.4	15.5
GtNm1	rpl6B	3374	509	509	1.00	3245.6	22.7
GtNm1	snrpD	175	182	182	1.00	470.8	21.9
GtNm1	ggt	677	1061	1082	0.98	306.4	20.3
GtNm1	hda	1011	1124	1124	1.00	440.4	27.2
GtNm1	trf	270	948	992	0.96	133.3	21.0
GtNm1	pab2	841	1381	1439	0.96	286.2	21.0
GtNm1	nol10	1613	938	938	1.00	842.0	22.9
GtNm1	orf147	1198	443	443	1.00	1324.1	22.1
GtNm1	orf69b	103	209	209	1.00	241.3	55.7
GtNm1	orf65b	100	197	197	1.00	248.5	44.4
GtNm1	orf151b	93	452	455	0.99	100.1	41.9
GtNm1	orf210b	128	632	632	1.00	99.2	53.7
GtNm1	orf110b	48	307	332	0.92	70.8	52.9
GtNm2	ORF110	32	320	332	0.96	47.2	52.9
GtNm2	ORF210	103	623	632	0.99	79.8	53.7
GtNm2	ORF151	87	455	455	1.00	93.6	41.9
GtNm2	ORF65	130	197	197	1.00	323.1	44.4
GtNm2	ORF69	120	209	209	1.00	281.1	55.7
GtNm2	ubc4	736	443	443	1.00	813.5	52.9
GtNm2	tfIId	3682	749	749	1.00	2407.0	39.7
GtNm2	mcm3	1701	2021	2021	1.00	412.1	22.1
GtNm2	rps4	3958	773	773	1.00	2507.1	30.1
GtNm2	ftsZ	7277	1199	1199	1.00	2971.7	35.0
GtNm2	tif211	1213	794	794	1.00	748.0	21.3
GtNm2	prsA1	1249	713	713	1.00	857.7	22.5
GtNm2	tcpG	6684	1508	1508	1.00	2170.2	25.0
GtNm2	tfIIA-S	1254	386	386	1.00	1590.7	23.0
GtNm2	ORF546	1336	1608	1640	0.98	398.9	18.2
GtNm2	ORF419	1694	1259	1259	1.00	658.8	19.5
GtNm2	ranbpm	499	842	842	1.00	290.2	26.1
GtNm2	rps3a	5652	659	659	1.00	4199.4	22.1
GtNm2	rub	14181	479	479	1.00	14495.9	28.8
GtNm2	h2B	22010	308	308	1.00	34989.9	32.0
GtNm2	gsp2	34766	641	641	1.00	26556.5	32.1
GtNm2	rpl30	7047	323	323	1.00	10682.6	28.4
GtNm2	rpc10	1841	221	221	1.00	4078.8	33.8
GtNm2	rps28	1282	197	197	1.00	3186.4	29.8
GtNm2	rps13	4298	443	443	1.00	4750.5	27.9
GtNm2	tubA	6605	1346	1346	1.00	2402.7	35.1
GtNm2	ORF626	84	1249	1880	0.66	21.9	16.8
GtNm2	sen34	150	570	728	0.78	100.9	21.5
GtNm2	rpl13	1244	383	383	1.00	1590.4	22.7
GtNm2	cdc28	1214	1847	1847	1.00	321.8	22.3
GtNm2	ORF938	442	2354	2816	0.84	76.9	18.5

GtNm2	ORF129	242	389	389	1.00	304.6	19.0
GtNm2	rpl14	662	356	356	1.00	910.5	21.3
GtNm2	nop5	1457	1150	1184	0.97	602.5	22.1
GtNm2	has1	4619	1409	1409	1.00	1605.1	25.1
GtNm2	rpl24	1184	194	194	1.00	2988.3	21.0
GtNm2	rip1	783	788	788	1.00	486.5	22.7
GtNm2	secE	1182	440	440	1.00	1315.3	23.8
GtNm2	ORF300	196	542	902	0.60	106.4	16.1
GtNm2	ORF731	432	1858	2195	0.85	96.4	16.2
GtNm2	engA	9973	1490	1490	1.00	3277.3	27.0
GtNm2	ORF430	173	914	1292	0.71	65.6	17.3
GtNm2	ORF170	221	388	512	0.76	211.3	15.4
GtNm2	rsp4	2557	599	599	1.00	2090.2	28.3
GtNm2	ORF80	42	223	242	0.92	85.0	19.3
GtNm2	ORF339	106	659	1019	0.65	50.9	15.7
GtNm2	imb1	1013	2511	2567	0.98	193.2	21.5
GtNm2	iap100	72301	3212	3212	1.00	11021.5	23.5
GtNm2	ORF82	6132	248	248	1.00	12106.7	27.3
GtNm2	ORF163	9409	491	491	1.00	9382.9	25.2
GtNm2	rpl9	5421	566	566	1.00	4689.6	27.0
GtNm2	nog1	1987	1190	1190	1.00	817.6	21.4
GtNm2	pop2	2049	785	785	1.00	1278.0	22.8
GtNm2	ORF168	3275	506	506	1.00	3169.1	24.3
GtNm2	ORF145	123	436	437	1.00	137.8	18.3
GtNm2	ORF142	65	369	428	0.86	74.4	14.0
GtNm2	eif1A	4135	431	431	1.00	4697.6	29.6
GtNm2	ORF249	18787	749	749	1.00	12281.4	28.4
GtNm2	cbp	369	515	515	1.00	350.8	21.7
GtNm2	ORF99	289	299	299	1.00	473.3	20.0
GtNm2	rpl8	4441	764	764	1.00	2846.2	32.9
GtNm2	uce-E2	4057	434	434	1.00	4577.1	28.7
GtNm2	ORF187	15267	563	563	1.00	13277.6	26.6
GtNm2	rps24	1184	389	389	1.00	1490.3	23.1
GtNm2	ncbP2	466	293	293	1.00	778.7	20.7
GtNm2	ATPbp	949	710	710	1.00	654.5	21.7
GtNm2	hcf136	53440	1187	1187	1.00	22044.0	30.8
GtNm2	tcpT	10746	1547	1547	1.00	3401.2	24.6
GtNm2	taf30	9740	413	413	1.00	11547.4	27.1
GtNm2	ORF102	4848	308	308	1.00	7707.0	26.9
GtNm2	sen2	1297	515	515	1.00	1233.1	20.7
GtNm2	prsS10b	2186	1187	1187	1.00	901.7	26.8
GtNm2	snrpD3	531	239	239	1.00	1087.9	24.6
GtNm2	ORF177	11157	533	533	1.00	10249.3	21.0
GtNm2	cbbX	141718	1115	1115	1.00	62233.5	33.3
GtNm2	rps27	2334	245	245	1.00	4664.5	23.2
GtNm2	rpa1	4479	4252	4304	0.99	509.5	23.6
GtNm2	snu13	646	377	377	1.00	839.0	25.4
GtNm2	rps26	1649	296	296	1.00	2727.7	29.0
GtNm2	prp8	4426	5556	6173	0.90	351.1	21.4
GtNm2	tcpZ	4123	1574	1574	1.00	1282.6	25.5
GtNm2	ORF227	1328	683	683	1.00	952.0	18.0
GtNm2	prsS1	1703	1530	2081	0.74	400.7	18.9
GtNm2	taf90	786	1390	1409	0.99	273.1	20.1
GtNm2	ORF80a	279	242	242	1.00	564.5	19.3
GtNm2	ORF231	128	512	695	0.74	90.2	15.9
GtNm2	rps19	1826	431	431	1.00	2074.4	27.3
GtNm2	rpl21	2923	470	470	1.00	3045.1	24.4
GtNm2	rps17	2476	356	356	1.00	3405.5	19.9
GtNm2	prsS12	371	776	776	1.00	234.1	20.3
GtNm2	prsB4	521	581	581	1.00	439.1	21.6
GtNm2	prsS4	3939	1175	1175	1.00	1641.4	28.9
GtNm2	rpb3	1219	899	899	1.00	663.9	23.8
GtNm2	ORF206	132	450	620	0.73	104.2	16.3
GtNm2	ORF605	1378	1599	1817	0.88	371.3	16.8
GtNm2	rps8	2857	536	536	1.00	2609.9	34.1

GtNm2	ORF139	5346	419	419	1.00	6247.2	35.2
GtNm2	ORF111	2691	335	335	1.00	3933.2	21.4
GtNm2	ORF125	477	377	377	1.00	619.5	21.7
GtNm2	dnaG	1142	1955	1955	1.00	286.0	21.1
GtNm2	ORF331	318	995	995	1.00	156.5	22.3
GtNm2	rpl36	465	227	227	1.00	1003.0	21.5
GtNm2	prsB1	1664	596	596	1.00	1367.0	26.6
GtNm2	ORF419a	611	1108	1259	0.88	237.6	20.0
GtNm2	rps23	5794	437	437	1.00	6491.9	34.5
GtNm2	rpl15	5352	614	614	1.00	4268.0	28.6
GtNm2	rpl10	2657	566	566	1.00	2298.5	30.2
GtNm2	ORF729	8197	2189	2189	1.00	1833.5	21.6
GtNm2	ORF477	2671	1433	1433	1.00	912.6	21.4
GtNm2	prsS13	2685	845	845	1.00	1555.8	27.9
GtNm2	rrp3	1612	1160	1160	1.00	680.4	24.4
GtNm2	nmt1	2657	1025	1025	1.00	1269.2	21.1
GtNm2	cpn60	27817	1769	1769	1.00	7699.4	32.5
GtNm2	rpl27A	2029	434	434	1.00	2289.1	30.6
GtNm2	sys1	4184	1277	1277	1.00	1604.3	25.7
GtNm2	gidA	23639	1949	1949	1.00	5938.7	30.4
GtNm2	ORF861	3366	2550	2585	0.99	637.6	19.2
GtNm2	ORF270	1274	812	812	1.00	768.2	33.1
GtNm2	sen1	1065	1917	2078	0.92	250.9	24.1
GtNm2	cdc2	662	887	887	1.00	365.4	23.1
GtNm2	rbp1	401	299	299	1.00	656.7	26.3
GtNm2	ORF64	161	193	194	0.99	406.3	13.8
GtNm2	sof1	939	1097	1097	1.00	419.1	23.5
GtNm2	ORF338	24	437	1016	0.43	11.6	17.4
GtNm2	ORF125	43	314	377	0.83	55.8	17.5
GtNm2	ORF307	95	731	923	0.79	50.4	17.0
GtNm2	prl1	154	885	920	0.96	82.0	21.8
GtNm2	ORF86	240	260	260	1.00	452.0	15.7
GtNm2	ORF323	6189	971	971	1.00	3120.9	26.5
GtNm2	ORF325	112	865	977	0.89	56.1	15.6
GtNm2	rpl40	1346	194	194	1.00	3397.2	30.3
GtNm2	ORF201	2121	524	605	0.87	1716.6	19.5
GtNm2	prsS6	1497	1193	1193	1.00	614.4	27.9
GtNm2	tfIIB-brf	1167	1184	1184	1.00	482.6	24.3
GtNm2	snrpE	127	236	236	1.00	263.5	21.1
GtNm2	ORF245	836	737	737	1.00	555.4	21.3
GtNm2	u5snRNP	1126	3517	4310	0.82	127.9	18.6
GtNm2	ORF861	1679	2166	2585	0.84	318.0	20.0
GtNm2	ORF148	42	293	446	0.66	46.1	15.4
GtNm2	tfIIE	230	595	623	0.96	180.8	19.7
GtNm2	snrpD2	633	248	248	1.00	1249.8	21.7
GtNm2	rps15A	1768	392	392	1.00	2208.4	26.5
GtNm2	prsS7	2123	1184	1184	1.00	878.0	30.0
GtNm2	ste4	609	905	905	1.00	329.5	22.5
GtNm2	ORF95	281	287	287	1.00	479.4	18.8
GtNm2	ORF196	321	588	590	1.00	266.4	17.6
GtNm2	gtp-bp	548	701	701	1.00	382.8	20.9
GtNm2	rpb1	3508	4421	4421	1.00	388.5	28.7
GtNm2	ORF116	163	350	350	1.00	228.0	17.1
GtNm2	rpf1	474	534	560	0.95	414.4	19.6
GtNm2	ORF203	80	433	611	0.71	64.1	18.3
GtNm2	rpl37A	1170	275	275	1.00	2083.2	29.0
GtNm2	rpl5	5906	698	698	1.00	4143.0	22.3
GtNm2	ORF160	111	408	482	0.85	112.8	14.3
GtNm2	mce	1501	1073	1073	1.00	684.9	20.7
GtNm2	gblp	7700	935	935	1.00	4032.3	30.1
GtNm2	rpl24	352	320	320	1.00	538.6	17.8
GtNm2	ubc4	760	443	443	1.00	840.0	52.9
GtNm2	ORF69b	104	209	209	1.00	243.6	55.7
GtNm2	ORF65b	89	197	197	1.00	221.2	44.4
GtNm2	ORF151b	75	455	455	1.00	80.7	41.9

GtNm2	ORF210b	86	623	632	0.99	66.6	53.7
GtNm2	ORF110b	15	207	332	0.62	22.1	52.9
GtNm3	orf110	33	321	332	0.97	48.7	52.9
GtNm3	orf210	116	632	632	1.00	89.9	53.7
GtNm3	orf151	66	455	455	1.00	71.0	41.9
GtNm3	orf65	85	197	197	1.00	211.3	44.4
GtNm3	orf69	87	209	209	1.00	203.8	55.7
GtNm3	ubc4	797	443	443	1.00	880.9	52.9
GtNm3	tfIID	3746	749	749	1.00	2448.8	39.7
GtNm3	snrpG	341	191	191	1.00	874.2	23.4
GtNm3	orf247	294	728	743	0.98	193.7	18.8
GtNm3	orf274	117	788	824	0.96	69.5	17.8
GtNm3	orf380	218	1092	1142	0.96	93.5	18.9
GtNm3	orf272	1211	818	818	1.00	724.9	22.0
GtNm3	rps21	1101	248	248	1.00	2173.7	24.1
GtNm3	orf199	120	480	599	0.80	98.1	16.0
GtNm3	rpl35A	2242	296	296	1.00	3708.7	23.2
GtNm3	orf323	69	791	971	0.81	34.8	15.8
GtNm3	rfc3	255	721	791	0.91	157.8	19.1
GtNm3	kin(snf2)	2188	1205	1205	1.00	889.1	28.5
GtNm3	rps29A	631	170	170	1.00	1817.4	30.4
GtNm3	cenp-A	1197	311	311	1.00	1884.5	33.3
GtNm3	rpl27	2446	434	434	1.00	2759.6	23.2
GtNm3	rpc9	345	323	323	1.00	523.0	19.8
GtNm3	orf176	243	530	530	1.00	224.5	24.5
GtNm3	orf357	1369	1073	1073	1.00	624.7	23.0
GtNm3	orf406	320	880	1220	0.72	128.4	17.0
GtNm3	pcna	2330	776	776	1.00	1470.2	29.7
GtNm3	rad25	635	1837	1853	0.99	167.8	26.9
GtNm3	orf224	158	671	674	1.00	114.8	17.5
GtNm3	rad51	850	995	995	1.00	418.3	34.3
GtNm3	kin	2083	1064	1064	1.00	958.6	31.5
GtNm3	orf81	479	245	245	1.00	957.3	19.9
GtNm3	gidB	676	350	350	1.00	945.7	21.1
GtNm3	rli1	3221	1796	1796	1.00	878.1	24.3
GtNm3	rla1	3887	314	314	1.00	6061.2	23.8
GtNm3	cbf5	618	830	932	0.89	324.7	19.0
GtNm3	dbp4	1253	1328	1328	1.00	462.0	22.5
GtNm3	orf301	3506	905	905	1.00	1896.9	23.8
GtNm3	rpb8	2108	530	530	1.00	1947.5	24.3
GtNm3	orf370	2053	1112	1112	1.00	904.0	19.8
GtNm3	rpbY	1111	326	326	1.00	1668.7	27.5
GtNm3	rpl34	1498	311	311	1.00	2358.4	28.8
GtNm3	hira	3037	1635	1670	0.98	890.4	23.1
GtNm3	rps6	2322	629	629	1.00	1807.5	27.0
GtNm3	orf261	12647	785	785	1.00	7888.5	28.1
GtNm3	orf283	3775	850	851	1.00	2172.0	19.5
GtNm3	orf85	75	257	257	1.00	142.9	20.5
GtNm3	rps27A	1915	347	347	1.00	2702.2	22.4
GtNm3	orf142	1206	428	428	1.00	1379.7	19.6
GtNm3	orf187	1116	563	563	1.00	970.6	18.3
GtNm3	rpc2	5803	3149	3149	1.00	902.3	26.4
GtNm3	rpb7	1125	590	590	1.00	933.6	28.3
GtNm3	rad3	216	1872	2120	0.88	49.9	18.0
GtNm3	reb1	614	821	821	1.00	366.2	24.5
GtNm3	tcpH	3962	1538	1538	1.00	1261.3	27.2
GtNm3	orf744	794	1792	2234	0.80	174.0	17.7
GtNm3	orf753	690	1777	2261	0.79	149.4	17.5
GtNm3	rpl18A	2733	479	479	1.00	2793.7	25.2
GtNm3	rps11	1309	416	416	1.00	1540.7	27.8
GtNm3	orf112	704	338	338	1.00	1019.8	23.0
GtNm3	orf144	70	423	434	0.97	79.0	14.9
GtNm3	orf214	165	611	644	0.95	125.5	18.8
GtNm3	orf134	212	404	404	1.00	256.9	20.2
GtNm3	orf193	8921	581	581	1.00	7518.2	27.0

GtNm3	tcpA	3341	1595	1595	1.00	1025.6	24.7
GtNm3	orf187	708	563	563	1.00	615.7	19.0
GtNm3	orf177	807	533	533	1.00	741.3	21.3
GtNm3	rpa5	2125	1019	1019	1.00	1021.1	23.2
GtNm3	ubc2	228	443	443	1.00	252.0	26.6
GtNm3	orf258	897	763	776	0.98	566.0	20.8
GtNm3	fkbp	7621	734	734	1.00	5083.8	25.3
GtNm3	orf223	843	671	671	1.00	615.1	17.7
GtNm3	orf379	999	1119	1139	0.98	429.5	20.6
GtNm3	orf625	52	985	1877	0.52	13.6	16.7
GtNm3	rpa2	4894	3329	3329	1.00	719.8	25.6
GtNm3	rps2	7448	665	665	1.00	5483.9	32.4
GtNm3	snrpD1	1622	233	233	1.00	3408.5	19.2
GtNm3	orf247	179	695	743	0.94	118.0	18.8
GtNm3	orf158	14	189	476	0.40	14.4	15.7
GtNm3	prp2-like	167	1004	1772	0.57	46.1	17.6
GtNm3	mrs2	2333	1148	1148	1.00	995.1	28.4
GtNm3	rla0	6034	893	893	1.00	3308.5	24.6
GtNm3	orf222	4448	668	668	1.00	3260.3	32.7
GtNm3	orf479	154	941	1439	0.65	52.4	17.3
GtNm3	orf160	4641	482	482	1.00	4714.5	28.4
GtNm3	orf341	805	896	1025	0.87	384.5	18.2
GtNm3	orf416	1431	1186	1250	0.95	560.5	18.5
GtNm3	EF2	19130	2546	2546	1.00	3679.0	32.9
GtNm3	orf209	806	629	629	1.00	627.4	20.3
GtNm3	orf470	354	1275	1412	0.90	122.8	17.7
GtNm3	prsA3	1313	698	698	1.00	921.0	25.6
GtNm3	rpb10	349	215	215	1.00	794.8	26.4
GtNm3	orf1613	254	2093	4841	0.43	25.7	16.0
GtNm3	nop1	1154	776	776	1.00	728.1	31.3
GtNm3	orf338	163	702	1016	0.69	78.6	18.7
GtNm3	orf118	313	356	356	1.00	430.5	16.8
GtNm3	orf160	6829	482	482	1.00	6937.2	36.9
GtNm3	orf141	93	407	425	0.96	107.1	16.4
GtNm3	orf180	283	542	542	1.00	255.7	19.0
GtNm3	snrpF	176	221	221	1.00	389.9	19.4
GtNm3	orf845	652	2029	2537	0.80	125.8	18.3
GtNm3	orf395	1483	1077	1187	0.91	611.7	18.9
GtNm3	orf228	1202	686	686	1.00	857.9	21.7
GtNm3	tfIIB	1327	977	977	1.00	665.0	27.4
GtNm3	rpl1	3648	767	767	1.00	2328.8	32.2
GtNm3	kin(aaB)	830	745	770	0.97	527.8	23.0
GtNm3	orf153	7	169	461	0.37	7.4	14.9
GtNm3	orf132	273	398	398	1.00	335.9	21.3
GtNm3	orf266	1599	800	800	1.00	978.7	22.0
GtNm3	kin(cdc2)	3484	911	911	1.00	1872.6	30.3
GtNm3	prsS8	2087	1202	1202	1.00	850.1	27.0
GtNm3	orf62	383	188	188	1.00	997.5	16.9
GtNm3	Ebi	2413	1304	1304	1.00	906.1	21.6
GtNm3	rps3	5677	656	656	1.00	4237.3	29.4
GtNm3	orf150	1110	452	452	1.00	1202.4	23.0
GtNm3	prsA7	3218	725	725	1.00	2173.3	24.9
GtNm3	orf72	210	218	218	1.00	471.7	20.5
GtNm3	prsB7	1467	710	710	1.00	1011.7	26.0
GtNm3	kin(mps1)	1211	1210	1610	0.75	368.3	19.6
GtNm3	orf127	116	334	383	0.87	148.3	16.9
GtNm3	orf203	381	611	611	1.00	305.3	19.9
GtNm3	sui1	812	449	449	1.00	885.5	25.3
GtNm3	orf160	6497	482	482	1.00	6599.9	36.9
GtNm3	rpl7	12812	743	743	1.00	8443.1	22.7
GtNm3	prsB3	2795	602	602	1.00	2273.3	27.4
GtNm3	nop56	2403	1256	1256	1.00	936.8	24.3
GtNm3	tcpB	2993	1502	1502	1.00	975.7	27.7
GtNm3	rps14	1552	479	479	1.00	1586.5	32.3
GtNm3	rps20	1591	350	350	1.00	2225.7	23.9

GtNm3	orf387	4684	1163	1163	1.00	1972.0	26.2
GtNm3	rps16	4411	428	428	1.00	5046.2	30.8
GtNm3	mak16	3435	647	647	1.00	2599.5	23.9
GtNm3	rpl3	14080	1127	1127	1.00	6117.2	29.7
GtNm3	orf232	1856	698	698	1.00	1302.0	17.0
GtNm3	rps9	5400	545	545	1.00	4851.4	29.9
GtNm3	orf201	356	438	605	0.72	288.1	19.5
GtNm3	mcm2	920	1671	1865	0.90	241.5	19.5
GtNm3	sbp1	440	866	866	1.00	248.8	23.2
GtNm3	nop2	1026	1076	1076	1.00	466.9	25.6
GtNm3	orf277	287	763	833	0.92	168.7	19.2
GtNm3	rpl17	6562	461	461	1.00	6969.6	28.4
GtNm3	rpl10A	8453	650	650	1.00	6367.5	26.4
GtNm3	orf319	1556	959	959	1.00	794.4	16.5
GtNm3	G10	4067	557	557	1.00	3575.1	19.4
GtNm3	cdc48	41829	2258	2258	1.00	9070.4	32.8
GtNm3	orf456	588	1368	1370	1.00	210.2	21.6
GtNm3	orf305	170	607	917	0.66	90.8	18.1
GtNm3	prsS6	2823	1151	1151	1.00	1200.9	26.2
GtNm3	orf272	494	818	818	1.00	295.7	21.2
GtNm3	tubB	7555	1325	1325	1.00	2791.9	35.4
GtNm3	rpl32	2777	359	359	1.00	3787.5	25.3
GtNm3	orf177	476	533	533	1.00	437.3	18.5
GtNm3	ste13	1159	1145	1145	1.00	495.6	24.5
GtNm3	orf143	479	429	431	1.00	544.2	21.5
GtNm3	sut	9593	2252	2252	1.00	2085.7	27.5
GtNm3	eif4a	3732	1157	1157	1.00	1579.4	29.8
GtNm3	ufd	2348	527	527	1.00	2181.5	27.3
GtNm3	tfIID	3735	749	749	1.00	2441.6	39.7
GtNm3	ubc4b	668	443	443	1.00	738.3	52.9
GtNm3	orf69b	70	209	209	1.00	164.0	55.7
GtNm3	orf65b	84	197	197	1.00	208.8	44.4
GtNm3	orf151b	71	438	455	0.96	76.4	41.9
GtNm3	orf210b	81	627	632	0.99	62.8	53.7
GtNm3	orf110b	27	320	332	0.96	39.8	52.9
	Sum	2042331			**Average**	2333.2	25.3

Mapping rates were calculated by gene lengths and mapped regions.
RPKM (Reads Per Kilobase of exon per Million mapped reads) was calculated by the following formula;
=(number of reads for each gene)*1000000/(total number of reads)*1000/(gene length)

Table 4.5 Relative expression of nucleomorph genes in *Cryptomonas paramecium*

Chromosome No.	Gene name	Number of reads	Mapped by some reads (bp)	Gene length (bp)	Mapping rate	RPKM	GC content (%)
CpNm1	ubc4	2706	443	443	1.00	2563.6	38.1
CpNm1	BRSK	10200	1148	1148	1.00	3728.9	31.2
CpNm1	rpl40	2171	152	152	1.00	5994.3	28.8
CpNm1	hsp90	74211	2042	2042	1.00	15252.2	31.9
CpNm1	CPARA_1gp005	8981	371	371	1.00	10159.5	23.7
CpNm1	CPARA_1gp006	497	671	782	0.86	266.7	19.8
CpNm1	CPARA_1gp007	1094	1125	1136	0.99	404.2	19.3
CpNm1	nat10	10364	2573	2573	1.00	1690.5	25.7
CpNm1	prsA2	6676	689	689	1.00	4066.5	29.3
CpNm1	bms1-like	6767	1829	1829	1.00	1552.8	26.4
CpNm1	CPARA_1gp011	828	413	413	1.00	841.4	19.3
CpNm1	CPARA_1gp012	160	727	737	0.99	91.1	17.8
CpNm1	tic22	8274	905	905	1.00	3837.0	29.6
CpNm1	clpP1	24783	731	731	1.00	14228.4	36.5
CpNm1	mcm4	10851	1880	1880	1.00	2422.3	21.7
CpNm1	CPARA_1gp016	599	791	791	1.00	317.8	19.1
CpNm1	CPARA_1gp017	76	300	335	0.90	95.2	19.0
CpNm1	prsA5	3849	692	692	1.00	2334.3	29.1
CpNm1	rps5	7646	575	575	1.00	5580.7	32.1
CpNm1	tfIIIC	6696	1760	1760	1.00	1596.7	24.2
CpNm1	CPARA_1gp021	467	437	437	1.00	448.5	19.2
CpNm1	cycB	3173	956	956	1.00	1392.9	27.0
CpNm1	rpl23A	4688	374	374	1.00	5260.6	26.4
CpNm1	Ha-orf146	5761	323	323	1.00	7485.4	25.3
CpNm1	rpb1	20154	4358	4358	1.00	1940.9	30.3
CpNm1	GTP-bp	1751	942	965	0.98	761.5	24.4
CpNm1	CPARA_1gp027	758	272	272	1.00	1169.6	26.7
CpNm1	CPARA_1gp028	351	152	152	1.00	969.1	19.6
CpNm1	CPARA_1gp029	744	803	803	1.00	388.8	19.4
CpNm1	tfIID	7079	695	695	1.00	4274.7	33.3
CpNm1	mcm3	3517	2075	2075	1.00	711.3	22.0
CpNm1	rps4	9347	728	728	1.00	5388.4	28.7
CpNm1	ftsZ	4226	1052	1052	1.00	1685.9	35.8
CpNm1	tif211	3348	803	803	1.00	1749.8	26.5
CpNm1	prsA1	2477	722	722	1.00	1439.8	26.7
CpNm1	tcpG	16860	1547	1547	1.00	4573.9	29.8
CpNm1	tfIIA-S	4285	371	371	1.00	4847.3	22.6
CpNm1	Ha-orf590	5694	1709	1709	1.00	1398.3	23.7
CpNm1	CPARA_1gp039	3752	1091	1091	1.00	1443.3	20.4
CpNm1	CPARA_1gp040	1042	239	239	1.00	1829.7	22.5
CpNm1	CPARA_1gp041	461	612	626	0.98	309.1	19.1
CpNm1	CPARA_1gp042	27	272	359	0.76	31.6	14.7
CpNm1	hda	3575	1112	1112	1.00	1349.2	28.6
CpNm1	ggt	5602	872	872	1.00	2696.2	24.7
CpNm1	CPARA_1gp045	523	170	170	1.00	1291.1	22.2
CpNm1	snrpD	619	233	233	1.00	1115.0	24.4
CpNm1	rpl6B	8841	593	593	1.00	6257.0	28.5
CpNm1	CPARA_1gp048	1124	200	200	1.00	2358.6	21.9
CpNm1	CPARA_1gp049	449	769	797	0.96	236.4	17.8
CpNm1	CPARA_1gp050	604	866	893	0.97	283.9	19.4
CpNm1	snrpF	1022	266	266	1.00	1612.5	21.7
CpNm1	Gt-orf180	3403	1403	1403	1.00	1017.9	22.4
CpNm1	CPARA_1gp053	2788	704	704	1.00	1662.0	20.3
CpNm1	U5_snRNP_115kDa	8393	2555	2555	1.00	1378.6	26.4
CpNm1	CPARA_1gp055	3775	1253	1253	1.00	1264.4	23.6
CpNm1	CPARA_1gp056	669	650	650	1.00	432.0	19.2
CpNm1	Ha-orf106	1362	308	308	1.00	1855.9	26.5
CpNm1	Ha-orf108	2787	353	353	1.00	3313.5	30.2
CpNm1	prsB6	1881	737	737	1.00	1071.1	26.7
CpNm1	CPARA_1gp060	640	461	461	1.00	582.6	21.4

CpNm1	Ha-orf831	3618	2327	2327	1.00	652.5	22.9
CpNm1	smc1	3912	3065	3179	0.96	516.5	19.6
CpNm1	rps15	7621	419	419	1.00	7633.4	25.7
CpNm1	Gt-orf534	26821	839	839	1.00	13416.3	35.1
CpNm1	pi4K	2922	1040	1040	1.00	1179.1	22.7
CpNm1	CPARA_1gp066	549	479	479	1.00	481.0	21.3
CpNm1	kin(snf1)	5445	1340	1340	1.00	1705.4	28.0
CpNm1	CPARA_1gp068	721	410	410	1.00	738.0	20.2
CpNm1	CPARA_1gp069	2953	434	434	1.00	2855.6	24.4
CpNm1	Ha-orf530	3650	1331	1331	1.00	1150.9	20.3
CpNm1	CPARA_1gp071	1464	371	371	1.00	1656.1	21.0
CpNm1	CPARA_1gp072	1586	614	614	1.00	1084.1	21.6
CpNm1	rla0	9005	773	773	1.00	4889.1	26.1
CpNm1	CPARA_1gp074	3006	1763	1763	1.00	715.6	21.5
CpNm1	mrs2	3323	1121	1121	1.00	1244.1	29.0
CpNm1	CPARA_1gp076	1359	593	593	1.00	961.8	20.5
CpNm1	CPARA_1gp077	152	173	173	1.00	368.7	14.9
CpNm1	CPARA_1gp078	5345	698	698	1.00	3213.8	25.0
CpNm1	snrpD1	3902	266	266	1.00	6156.4	25.1
CpNm1	rps2	23255	692	692	1.00	14103.6	34.2
CpNm1	rpa2	10393	3308	3308	1.00	1318.5	28.0
CpNm1	Ha-orf803	2647	2381	2384	1.00	466.0	21.5
CpNm1	CPARA_1gp083	3603	1136	1136	1.00	1331.1	22.1
CpNm1	CPARA_1gp084	4806	575	575	1.00	3507.8	22.6
CpNm1	CPARA_1gp085	1589	876	905	0.97	736.9	20.6
CpNm1	rpa5	5315	1289	1289	1.00	1730.5	26.6
CpNm1	CPARA_1gp087	1161	242	242	1.00	2013.4	27.2
CpNm1	Gt-orf193	7940	560	560	1.00	5950.5	29.2
CpNm1	kin(cdc)	1263	221	221	1.00	2398.5	34.7
CpNm1	tcpA	12117	1604	1604	1.00	3170.4	29.5
CpNm1	Ha-orf132	576	398	398	1.00	607.4	22.6
CpNm1	U5_snRNP_40kDa	1762	881	881	1.00	839.4	27.2
CpNm1	prsS7	5705	1214	1214	1.00	1972.2	32.8
CpNm1	rps15A	5012	404	404	1.00	5206.6	25.4
CpNm1	snrpD2	2516	257	257	1.00	4108.6	32.2
CpNm1	tfIIE	1375	635	635	1.00	908.8	25.5
CpNm1	CPARA_1gp097	147	215	215	1.00	286.9	16.2
CpNm1	Gt-orf228	5338	734	734	1.00	3052.1	23.9
CpNm1	CPARA_1gp099	3463	443	443	1.00	3280.7	23.4
CpNm1	Ha-orf194-430	5541	1945	1973	0.99	1178.6	20.9
CpNm1	CPARA_1gp101	507	810	854	0.95	249.2	18.9
CpNm1	Gt-orf115	1312	377	377	1.00	1460.5	23.3
CpNm1	CPARA_1gp103	2952	389	389	1.00	3184.8	24.1
CpNm1	CPARA_1gp104	893	617	617	1.00	607.4	23.3
CpNm1	CPARA_1gp105	1354	1196	1196	1.00	475.1	20.6
CpNm1	bysl	1643	746	746	1.00	924.3	22.8
CpNm1	snrpE	1359	149	149	1.00	3827.8	30.7
CpNm1	tfIIB-brf	4463	1157	1157	1.00	1618.9	27.8
CpNm1	prsS6A	5120	1187	1187	1.00	1810.3	31.6
CpNm1	CPARA_1gp110	641	606	617	0.98	436.0	22.2
CpNm1	cbbX	125122	1031	1031	1.00	50932.6	37.4
CpNm1	snrpD3	3788	254	254	1.00	6258.9	27.1
CpNm1	sen2	4913	737	737	1.00	2797.7	25.9
CpNm1	prsS10B	3306	1151	1151	1.00	1205.4	27.9
CpNm1	Gt-orf102	3468	311	311	1.00	4679.9	26.0
CpNm1	taf30	6401	407	407	1.00	6600.5	25.0
CpNm1	tcpT	10712	1586	1586	1.00	2834.6	30.6
CpNm1	sui1	1565	290	290	1.00	2264.8	24.1
CpNm1	prsS13	6386	935	935	1.00	2866.4	28.1
CpNm1	tcpD	13174	1595	1595	1.00	3466.4	29.8
CpNm1	Gt-orf216	1159	698	698	1.00	696.9	28.5
CpNm1	pp1	7643	929	929	1.00	3452.8	29.0
CpNm1	CPARA_1gp123	535	518	518	1.00	433.5	21.2
CpNm1	CPARA_1gp124	359	278	278	1.00	542.0	20.1
CpNm1	CPARA_1gp125	5166	554	554	1.00	3913.5	25.6

CpNm1	Gt-orf261	16428	821	821	1.00	8397.7	30.3
CpNm1	rps6	7120	596	596	1.00	5013.7	27.3
CpNm1	CPARA_1gp128	1407	188	188	1.00	3140.9	25.9
CpNm1	Gt-orf176	2877	491	491	1.00	2459.1	25.0
CpNm1	fet5	7796	758	758	1.00	4316.4	26.7
CpNm1	rpl26	4334	263	263	1.00	6916.0	27.3
CpNm1	rps10B	3554	293	293	1.00	5090.6	25.5
CpNm1	Gt-orf187	2809	590	590	1.00	1998.1	21.5
CpNm1	rpc2	14261	3263	3263	1.00	1834.2	28.2
CpNm1	rpb7	2699	527	527	1.00	2149.4	29.7
CpNm1	rad3	4736	2171	2171	1.00	915.5	22.5
CpNm1	reb1	5916	848	848	1.00	2927.9	29.1
CpNm1	rps27A	7991	326	326	1.00	10287.4	30.0
CpNm1	rpf1	3596	545	545	1.00	2769.1	27.5
CpNm1	rpc10	1975	134	134	1.00	6185.6	35.6
CpNm1	rpl30	9380	326	326	1.00	12075.5	28.4
CpNm1	gsp2	35292	656	656	1.00	22578.4	36.1
CpNm1	h2B	15169	341	341	1.00	18669.1	33.9
CpNm1	CPARA_1gp144	3028	1109	1109	1.00	1145.9	23.1
CpNm1	ef2	56041	2546	2546	1.00	9237.8	35.9
CpNm1	CPARA_1gp146	874	299	299	1.00	1226.8	20.7
CpNm1	CPARA_1gp147	973	464	464	1.00	880.1	21.9
CpNm1	CPARA_1gp148	198	374	383	0.98	217.0	19.8
CpNm1	prsA3	3461	728	728	1.00	1995.2	28.3
CpNm1	rpb10	1664	224	224	1.00	3117.6	31.1
CpNm1	CPARA_1gp151	2113	1241	1241	1.00	714.6	20.2
CpNm1	CPARA_1gp152	15	170	170	1.00	37.0	17.5
CpNm1	CPARA_1gp153	876	1388	1388	1.00	264.9	20.0
CpNm1	CPARA_1gp154	222	510	599	0.85	155.5	19.2
CpNm1	CPARA_1gp155	360	727	959	0.76	157.5	17.0
CpNm1	CPARA_1gp156	535	1028	1028	1.00	218.4	20.5
CpNm1	tcpZ	13342	1556	1556	1.00	3598.6	29.9
CpNm1	prp8	8630	6347	6347	1.00	570.6	22.8
CpNm1	rps26	5903	311	311	1.00	7965.9	29.8
CpNm1	snu13	2271	422	422	1.00	2258.5	29.3
CpNm1	rpa1	14955	4376	4376	1.00	1434.3	26.3
CpNm1	rps27	7122	284	284	1.00	10524.6	31.6
CpNm1	CPARA_1gp163	668	185	185	1.00	1515.4	26.9
CpNm1	rbp1	1463	308	308	1.00	1993.5	26.5
CpNm1	CPARA_1gp165	707	194	194	1.00	1529.5	19.5
CpNm1	sof1	11980	1181	1181	1.00	4257.2	32.5
CpNm1	CPARA_1gp167	868	182	182	1.00	2001.6	24.6
CpNm1	CPARA_1gp168	311	275	275	1.00	474.6	18.1
CpNm1	CPARA_1gp169	683	329	329	1.00	871.3	21.8
CpNm1	CPARA_1gp170	1273	719	719	1.00	743.1	20.3
CpNm1	CPARA_1gp171	3169	953	953	1.00	1395.6	24.9
CpNm1	CPARA_1gp172	5736	383	383	1.00	6285.4	29.2
CpNm2	ubc4	2916	443	443	1.00	2762.5	38.1
CpNm2	BRSK	11560	1283	1283	1.00	3781.4	30.8
CpNm2	rpl40	2763	158	158	1.00	7339.1	28.3
CpNm2	trf	877	1098	1115	0.98	330.1	22.0
CpNm2	pab2	1392	1224	1232	0.99	474.2	21.4
CpNm2	nol10	3811	1049	1049	1.00	1524.7	26.0
CpNm2	Gt-orf160	4331	482	482	1.00	3771.0	27.5
CpNm2	ufd	3694	581	581	1.00	2668.3	27.0
CpNm2	eif4A	9884	1172	1172	1.00	3539.4	30.8
CpNm2	Gt-orf143	657	434	434	1.00	635.3	26.0
CpNm2	ste13	3148	1277	1277	1.00	1034.6	29.3
CpNm2	U3_snRNP	3438	1400	1400	1.00	1030.6	25.2
CpNm2	rpl32	6863	365	365	1.00	7891.2	26.8
CpNm2	CPARA_2gp186	436	332	332	1.00	551.1	21.3
CpNm2	CPARA_2gp187	197	186	224	0.83	369.1	18.7
CpNm2	Gt-orf456	1870	1292	1292	1.00	607.4	22.7
CpNm2	CPARA_2gp189	158	170	170	1.00	390.1	18.7
CpNm2	cdc48b	47324	2261	2261	1.00	8784.2	36.1

CpNm2	G10	2417	761	761	1.00	1332.9	19.0
CpNm2	Gt-orf272	1067	686	686	1.00	652.8	24.2
CpNm2	tubB	8552	1328	1328	1.00	2702.7	35.7
CpNm2	prsS6B	4446	1139	1139	1.00	1638.2	30.9
CpNm2	CPARA_2gp195	173	218	218	1.00	333.1	20.1
CpNm2	sbp1	971	893	893	1.00	456.3	23.9
CpNm2	nop2	3670	1046	1046	1.00	1472.5	26.6
CpNm2	Ha-orf336	3393	941	941	1.00	1513.3	23.9
CpNm2	rpl17	5567	473	473	1.00	4939.5	32.1
CpNm2	rpl10A	6660	653	653	1.00	4280.4	26.9
CpNm2	CPARA_2gp201	96	399	455	0.88	88.5	17.3
CpNm2	CPARA_2gp202	513	254	254	1.00	847.6	20.8
CpNm2	mcm2	2804	2123	2123	1.00	554.3	22.9
CpNm2	rps9	8526	557	557	1.00	6424.1	32.8
CpNm2	CPARA_2gp205	1363	395	395	1.00	1448.2	20.7
CpNm2	CPARA_2gp206	3551	212	212	1.00	7029.7	31.5
CpNm2	rpl3	19261	995	995	1.00	8124.1	31.8
CpNm2	mak16	2319	623	623	1.00	1562.2	24.7
CpNm2	rps16	6152	428	428	1.00	6032.4	35.2
CpNm2	CPARA_2gp210	524	185	185	1.00	1188.7	24.7
CpNm2	rps20	2766	338	338	1.00	3434.4	28.3
CpNm2	rps14	6685	419	419	1.00	6695.9	38.3
CpNm2	tcpB	7963	1508	1508	1.00	2216.1	29.7
CpNm2	nop56	4591	1238	1238	1.00	1556.4	24.5
CpNm2	prsB3	5774	605	605	1.00	4005.4	31.8
CpNm2	rpl7	11470	740	740	1.00	6505.1	28.9
CpNm2	CPARA_2gp217	421	431	431	1.00	409.9	19.4
CpNm2	CPARA_2gp218	418	200	200	1.00	877.1	23.4
CpNm2	CPARA_2gp219	1858	752	752	1.00	1036.9	23.9
CpNm2	CPARA_2gp220	681	320	320	1.00	893.1	21.2
CpNm2	hsf	5330	647	647	1.00	3457.4	27.9
CpNm2	Ha-orf132	4572	911	911	1.00	2106.2	25.7
CpNm2	Gt-orf755	6326	2309	2309	1.00	1149.8	25.2
CpNm2	CPARA_2gp224	552	458	458	1.00	505.8	22.7
CpNm2	eif2B	568	497	497	1.00	479.6	25.9
CpNm2	CPARA_2gp226	4522	1460	1460	1.00	1299.9	22.7
CpNm2	rps25	2985	305	305	1.00	4107.4	24.8
CpNm2	CPARA_2gp228	304	421	434	0.97	294.0	19.3
CpNm2	prsB5	2012	704	704	1.00	1199.4	31.3
CpNm2	rpl11B	7036	545	545	1.00	5418.1	32.1
CpNm2	rpl31	3090	290	290	1.00	4471.8	26.8
CpNm2	gblp1	7977	2315	2315	1.00	1446.1	24.9
CpNm2	Ha-orf102	675	278	278	1.00	1019.0	23.7
CpNm2	rpl8	15342	755	755	1.00	8528.2	35.7
CpNm2	uceE2	2938	440	440	1.00	2802.3	32.4
CpNm2	Gt-orf187	7333	662	662	1.00	4648.8	29.6
CpNm2	rps24	3107	383	383	1.00	3404.6	26.8
CpNm2	CPARA_2gp238	499	170	170	1.00	1231.9	18.7
CpNm2	ATPbp	1857	764	764	1.00	1020.1	26.1
CpNm2	CPARA_2gp240	1833	164	164	1.00	4690.7	22.4
CpNm2	CPARA_2gp241	916	191	191	1.00	2012.7	24.5
CpNm2	CPARA_2gp242	811	617	617	1.00	551.6	20.6
CpNm2	CPARA_2gp243	229	179	179	1.00	536.9	18.9
CpNm2	Ha-orf390	1477	1205	1205	1.00	514.4	20.4
CpNm2	CPARA_2gp245	510	497	497	1.00	430.7	19.7
CpNm2	fcf1	1112	449	449	1.00	1039.4	26.9
CpNm2	mcm7	6102	2066	2066	1.00	1239.5	22.7
CpNm2	dph1	2223	1337	1337	1.00	697.8	24.8
CpNm2	asf	1524	482	482	1.00	1327.0	25.7
CpNm2	tcpH	7264	1532	1532	1.00	1989.9	29.7
CpNm2	CPARA_2gp251	558	620	620	1.00	377.7	21.4
CpNm2	mcm5	7269	1817	1817	1.00	1679.0	22.7
CpNm2	h4	5779	314	314	1.00	7724.0	33.3
CpNm2	h3	3662	413	413	1.00	3721.3	32.4
CpNm2	rpb2	16149	3629	3629	1.00	1867.6	32.0

CpNm2	CPARA_2gp256	1493	854	854	1.00	733.7	21.3
CpNm2	CPARA_2gp257	459	719	719	1.00	267.9	18.3
CpNm2	rpl23	9701	422	422	1.00	9647.7	36.6
CpNm2	eif5A	8339	464	464	1.00	7542.5	30.5
CpNm2	CPARA_2gp260	657	815	815	1.00	338.3	22.8
CpNm2	CPARA_2gp261	1456	1352	1352	1.00	452.0	23.4
CpNm2	CPARA_2gp262	1743	1171	1205	0.97	607.1	22.5
CpNm2	der1	5479	629	629	1.00	3655.7	29.4
CpNm2	kea1	19286	2144	2144	1.00	3775.2	33.3
CpNm2	rpoD	6851	1244	1244	1.00	2311.3	33.7
CpNm2	CPARA_2gp266	56	332	332	1.00	70.8	20.1
CpNm2	CPARA_2gp267	1159	1297	1328	0.98	366.3	21.4
CpNm2	CPARA_2gp268	426	221	221	1.00	809.0	20.7
CpNm2	CPARA_2gp269	1572	812	812	1.00	812.5	21.5
CpNm2	pab1	633	962	1103	0.87	240.9	19.3
CpNm2	rrp3	4891	1211	1211	1.00	1695.0	22.4
CpNm2	rpl13A	6204	563	563	1.00	4624.7	28.2
CpNm2	CPARA_2gp273	100	152	152	1.00	276.1	22.9
CpNm2	CPARA_2gp274	820	723	797	0.91	431.8	17.4
CpNm2	rps11	4192	413	413	1.00	4259.8	28.0
CpNm2	rpl18A	4497	458	458	1.00	4120.8	26.8
CpNm2	CPARA_2gp277	565	1246	1268	0.98	187.0	18.8
CpNm2	CPARA_2gp278	1759	641	641	1.00	1151.7	23.2
CpNm2	tbl3	3927	2273	2273	1.00	725.1	23.5
CpNm2	CPARA_2gp280	466	167	167	1.00	1171.1	19.6
CpNm2	CPARA_2gp281	1160	883	884	1.00	550.7	21.2
CpNm2	imp4	1477	647	647	1.00	958.1	25.3
CpNm2	Ha-orf154	1219	425	425	1.00	1203.7	21.4
CpNm2	rpl14	3300	377	377	1.00	3673.6	25.1
CpNm2	nop5	5619	1181	1181	1.00	1996.8	26.8
CpNm2	has1	11429	1436	1436	1.00	3340.2	28.4
CpNm2	rip1	1034	725	725	1.00	598.6	22.3
CpNm2	secE	699	341	341	1.00	860.3	23.7
CpNm2	CPARA_2gp289	2477	995	995	1.00	1044.8	22.4
CpNm2	CPARA_2gp290	380	685	893	0.77	178.6	17.9
CpNm2	CPARA_2gp291	1592	854	854	1.00	782.4	21.9
CpNm2	engA	5263	1469	1469	1.00	1503.6	26.2
CpNm2	CPARA_2gp293	631	254	254	1.00	1042.6	23.9
CpNm2	CPARA_2gp294	470	944	1004	0.94	196.5	19.7
CpNm2	CPARA_2gp295	132	526	599	0.88	92.5	17.3
CpNm2	CPARA_2gp296	424	305	305	1.00	583.4	21.6
CpNm2	CPARA_2gp297	718	155	155	1.00	1944.1	21.2
CpNm2	rsp4	6357	626	626	1.00	4261.9	32.5
CpNm2	taf	1305	1178	1178	1.00	464.9	21.1
CpNm2	imb1	3998	2579	2579	1.00	650.6	25.0
CpNm2	iap100	25910	3229	3248	0.99	3347.9	26.1
CpNm2	Gt-orf82	5948	392	392	1.00	6368.0	27.2
CpNm2	rpl9	7018	563	563	1.00	5231.5	27.8
CpNm2	nog1	3227	1136	1136	1.00	1192.2	24.1
CpNm2	CPARA_2gp305	738	263	263	1.00	1177.7	27.7
CpNm2	pop2	2528	815	815	1.00	1301.8	26.8
CpNm2	Gt-orf168	2040	443	443	1.00	1932.6	27.3
CpNm2	CPARA_2gp308	996	758	758	1.00	551.5	19.6
CpNm2	eif1A	4887	356	356	1.00	5761.2	30.8
CpNm2	CPARA_2gp310	1687	167	167	1.00	4239.5	28.6
CpNm2	Gt-orf357	3061	1073	1073	1.00	1197.2	27.0
CpNm2	tubG	6602	1277	1277	1.00	2169.7	28.8
CpNm2	tcpE	7928	1598	1598	1.00	2082.1	30.5
CpNm2	erf1	8293	1268	1268	1.00	2744.8	28.6
CpNm2	impA	8377	1529	1529	1.00	2299.3	29.9
CpNm2	CPARA_2gp316	137	316	368	0.86	156.2	17.3
CpNm2	sufD	9150	1430	1430	1.00	2685.4	29.0
CpNm2	smc3	3814	3007	3101	0.97	516.2	20.4
CpNm2	rpabc5	2507	650	650	1.00	1618.7	27.2
CpNm2	rpabc6	3073	392	392	1.00	3290.0	30.3

CpNm2	CPARA_2gp321	753	173	173	1.00	1826.7	19.0
CpNm2	Gt-orf365	1989	1061	1061	1.00	786.8	26.3
CpNm2	prsA6	3157	719	719	1.00	1842.8	30.0
CpNm2	rpc1	5677	3896	3896	1.00	611.5	28.1
CpNm2	nip7	548	545	545	1.00	422.0	22.3
CpNm2	hsp70	118022	1937	1937	1.00	25571.4	37.0
CpNm2	CPARA_2gp327	2044	323	323	1.00	2655.8	24.1
CpNm3	ubc4	2744	443	443	1.00	2599.6	38.1
CpNm3	BRSK	7782	1076	1076	1.00	3035.3	31.9
CpNm3	rfc3	2073	923	923	1.00	942.6	24.2
CpNm3	Ha-orf448	2821	1253	1253	1.00	944.9	23.8
CpNm3	brx1	674	632	689	0.92	410.5	20.9
CpNm3	rpl35A	6467	320	320	1.00	8481.5	29.6
CpNm3	rps30	4863	167	167	1.00	12221.1	36.9
CpNm3	CPARA_3gp335	1197	593	593	1.00	847.1	23.4
CpNm3	CPARA_3gp336	823	392	392	1.00	881.1	21.6
CpNm3	CPARA_3gp337	1210	596	596	1.00	852.0	25.5
CpNm3	CPARA_3gp338	877	731	731	1.00	503.5	20.4
CpNm3	CPARA_3gp339	125	221	221	1.00	237.4	18.9
CpNm3	CPARA_3gp340	664	398	398	1.00	700.2	22.8
CpNm3	Gt-orf160	1889	485	485	1.00	1634.6	27.4
CpNm3	kin(ABC)	9343	2305	2453	0.94	1598.5	26.0
CpNm3	CPARA_3gp343	124	271	353	0.77	147.4	18.6
CpNm3	sen1	4909	2042	2042	1.00	1008.9	26.8
CpNm3	Gt-orf270	7731	854	854	1.00	3799.3	34.9
CpNm3	CPARA_3gp346	6527	2792	2792	1.00	981.1	22.3
CpNm3	gidA	13616	1943	1943	1.00	2941.0	33.6
CpNm3	sys1	5979	1328	1328	1.00	1889.5	27.0
CpNm3	rpl27A	8811	401	401	1.00	9221.5	34.1
CpNm3	cpn60	39704	1778	1778	1.00	9371.8	35.4
CpNm3	nmt1	1575	1016	1016	1.00	650.6	24.1
CpNm3	rrp3	2777	1190	1190	1.00	979.4	28.5
CpNm3	eif2G	3881	1286	1286	1.00	1266.6	30.5
CpNm3	kin(cdc2)	6207	902	902	1.00	2888.0	33.0
CpNm3	Gt-orf266	2898	827	827	1.00	1470.7	27.1
CpNm3	CPARA_3gp356	275	267	350	0.76	329.8	20.8
CpNm3	rpb4	292	396	398	0.99	307.9	19.5
CpNm3	CPARA_3gp358	861	308	308	1.00	1173.2	24.6
CpNm3	kin(aaB)	5530	788	788	1.00	2945.2	25.0
CpNm3	rpl1	8405	914	914	1.00	3859.3	34.8
CpNm3	tfIIB	2489	1025	1025	1.00	1019.1	28.9
CpNm3	CPARA_3gp362	2239	1184	1190	0.99	789.6	22.2
CpNm3	CPARA_3gp363	950	848	848	1.00	470.2	19.3
CpNm3	CPARA_3gp364	739	1343	1400	0.96	221.5	20.1
CpNm3	dib1	698	269	269	1.00	1089.0	21.5
CpNm3	prsB7	3834	764	764	1.00	2106.1	30.6
CpNm3	prsA7	2370	722	722	1.00	1377.6	28.4
CpNm3	rps3	6039	632	632	1.00	4010.2	33.2
CpNm3	ebi	2261	1340	1340	1.00	708.1	23.0
CpNm3	CPARA_3gp370	387	651	677	0.96	239.9	19.5
CpNm3	prsS8	5735	1211	1211	1.00	1987.5	28.5
CpNm3	prsS1	4514	2171	2171	1.00	872.6	24.0
CpNm3	taf90	1676	1339	1451	0.92	484.8	25.6
CpNm3	CPARA_3gp374	913	752	752	1.00	509.5	21.0
CpNm3	rps19	5861	467	467	1.00	5267.1	29.1
CpNm3	rpl21	5924	476	476	1.00	5223.1	25.8
CpNm3	rps17	4561	374	374	1.00	5118.1	28.8
CpNm3	prsS12	1978	797	797	1.00	1041.6	24.7
CpNm3	prsB4	2087	581	581	1.00	1507.5	29.6
CpNm3	prsS4	4170	1223	1223	1.00	1431.0	31.5
CpNm3	rpb3	2034	905	905	1.00	943.2	25.9
CpNm3	CPARA_3gp382	414	287	287	1.00	605.4	23.6
CpNm3	CPARA_3gp383	213	365	365	1.00	244.9	18.0
CpNm3	CPARA_3gp384	1062	894	908	0.98	490.9	21.8
CpNm3	CPARA_3gp385	3809	605	605	1.00	2642.3	19.0

CpNm3	rps8	8728	533	533	1.00	6872.4	28.3
CpNm3	Gt-orf139	6636	416	416	1.00	6694.7	34.1
CpNm3	CPARA_3gp388	2727	380	380	1.00	3011.8	22.8
CpNm3	Gt-orf125	1361	413	413	1.00	1383.0	28.0
CpNm3	dnaG	3105	2027	2027	1.00	642.9	24.3
CpNm3	rpl36	1510	200	200	1.00	3168.6	21.9
CpNm3	prsB1	6614	647	647	1.00	4290.2	25.9
CpNm3	Gt-orf419	8319	1451	1451	1.00	2406.2	22.7
CpNm3	rps23	11592	437	437	1.00	11132.6	33.3
CpNm3	rpl15	10306	614	614	1.00	7044.4	31.7
CpNm3	rpl10	8452	587	587	1.00	6042.9	34.2
CpNm3	CPARA_3gp397	4998	644	644	1.00	3257.1	25.6
CpNm3	CPARA_3gp398	3085	527	527	1.00	2456.8	22.0
CpNm3	CPARA_3gp399	3292	566	566	1.00	2441.0	27.2
CpNm3	Gt-orf477	6392	1385	1385	1.00	1936.9	25.0
CpNm3	rps3A	7700	662	662	1.00	4881.5	24.7
CpNm3	taf13	1148	479	479	1.00	1005.8	21.3
CpNm3	snrpG	1869	224	224	1.00	3501.7	21.8
CpNm3	rpl19	4881	458	458	1.00	4472.6	26.6
CpNm3	cdc48a	3656	1976	1976	1.00	776.5	28.3
CpNm3	CPARA_3gp406	684	332	332	1.00	864.6	24.0
CpNm3	CPARA_3gp407	1721	971	971	1.00	743.8	19.1
CpNm3	clpP2	8873	698	698	1.00	5335.0	37.2
CpNm3	rpl18	4656	467	467	1.00	4184.2	28.2
CpNm3	rpl7A	12140	623	623	1.00	8178.1	26.4
CpNm3	eif4E	2982	575	575	1.00	2176.5	28.5
CpNm3	ATP/GTP-bp	2736	1031	1031	1.00	1113.7	27.9
CpNm3	dhm	3738	1625	1625	1.00	965.4	28.6
CpNm3	rcl1	2613	1043	1043	1.00	1051.4	25.2
CpNm3	Gt-orf532	16636	1301	1301	1.00	5366.5	29.0
CpNm3	CPARA_3gp416	1886	320	320	1.00	2473.5	26.2
CpNm3	rfc2	4998	953	953	1.00	2201.0	26.4
CpNm3	CPARA_3gp418	811	1361	1361	1.00	250.1	20.6
CpNm3	rpl12	10453	491	491	1.00	8934.7	35.0
CpNm3	mcm6	5726	2219	2219	1.00	1083.0	23.5
CpNm3	rpl34	1908	242	242	1.00	3308.9	21.8
CpNm3	CPARA_3gp422	3079	1541	1541	1.00	838.5	21.1
CpNm3	rpb11	2236	305	305	1.00	3076.8	27.8
CpNm3	CPARA_3gp424	5492	1310	1310	1.00	1759.5	25.3
CpNm3	rpb8	2091	455	455	1.00	1928.7	29.2
CpNm3	dbp4	1770	1379	1379	1.00	538.7	24.6
CpNm3	cbf5	11192	1076	1076	1.00	4365.3	24.7
CpNm3	rla1	7575	341	341	1.00	9322.9	27.2
CpNm3	rli1	7582	1799	1799	1.00	1768.8	28.3
CpNm3	gidB	926	755	755	1.00	514.7	23.0
CpNm3	kin(gs)	6823	1064	1064	1.00	2691.3	33.2
CpNm3	rad51	4509	995	995	1.00	1901.9	35.0
CpNm3	Ha-orf239	992	701	701	1.00	593.9	22.8
CpNm3	rad25	3445	1874	1874	1.00	771.5	28.5
CpNm3	pcna	7660	773	773	1.00	4158.8	36.7
CpNm3	CPARA_3gp436	544	924	1091	0.85	209.3	20.0
CpNm3	rps29A	1599	170	170	1.00	3947.5	29.8
CpNm3	cenp-A	2308	338	338	1.00	2865.8	29.5
CpNm3	rpl27	7088	422	422	1.00	7049.1	25.5
CpNm3	rpc9	2048	320	320	1.00	2686.0	27.1
CpNm3	CPARA_3gp441	136	323	332	0.97	171.9	17.1
CpNm3	CPARA_3gp442	175	272	272	1.00	270.0	23.4
CpNm3	rpl37A	6166	278	278	1.00	9308.5	35.1
CpNm3	rpl5	13090	719	719	1.00	7640.7	28.5
CpNm3	CPARA_3gp445	112	458	491	0.93	95.7	18.7
CpNm3	mce	2778	1070	1070	1.00	1089.6	23.4
CpNm3	gblp2	11657	941	941	1.00	5199.0	31.6
CpNm3	rpl24	611	317	317	1.00	808.9	19.8
CpNm3	rps28	3120	197	197	1.00	6646.7	36.9
CpNm3	rps13	5767	449	449	1.00	5390.4	29.8

CpNm3	tubA	9673	1352	1352	1.00	3002.7	36.1
CpNm3	CPARA_3gp452	1104	701	701	1.00	661.0	20.4
CpNm3	CPARA_3gp453	506	491	491	1.00	432.5	21.1
CpNm3	CPARA_3gp454	419	449	449	1.00	391.6	19.8
CpNm3	sen34	1319	749	749	1.00	739.1	26.0
CpNm3	rpl13	6914	443	443	1.00	6550.1	30.4
CpNm3	cdc28	5914	1873	1889	0.99	1313.9	27.0
CpNm3	Gt-orf938	1637	3059	3173	0.96	216.5	20.4
CpNm3	CPARA_3gp459	1022	446	446	1.00	961.7	21.9
CpNm3	rps15	5624	431	431	1.00	5476.3	32.6
CpNm3	eif6	4462	677	677	1.00	2766.1	32.4
CpNm3	CPARA_3gp462	12912	1853	1853	1.00	2924.4	23.6
CpNm3	nop1	3916	749	749	1.00	2194.2	31.7
CpNm3	CPARA_3gp464	374	875	914	0.96	171.7	20.7
CpNm3	CPARA_3gp465	1813	1142	1142	1.00	666.3	23.3
CpNm3	ubc4	3043	443	443	1.00	2882.8	38.1
	Sum	2382754			**Average**	2767.8	25.8

Mapping rates were calculated by gene lengths and mapped regions.
RPKM (Reads Per Kilobase of exon per Million mapped reads) was calculated by the following formula;
=(number of reads for each gene)*1000000/(total number of reads)*1000/(gene length)