

GENOME DEGENERATION IN OBLIGATE PARASITES AND ENDOSYMBIONTS

by

ANNA EVGENYEVNA GANGAEVA

Honors B.Sc., The University of Toronto, 2006

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

(Botany)

THE UNIVERSITY OF BRITISH COLUMBIA

(VANCOUVER)

August 2008

©Anna Evgenyevna Gangaeva, 2008

ABSTRACT

Microorganisms are a goldmine for evolutionary genetics as their genomes can evolve at an extraordinary rate which results in some of the most extravagant adaptations in terms of genome structure and function as well as survival in the most unusual environments. One trend observed in several evolutionary scenarios is genome degeneration. It is most prominent in endosymbionts and obligate intracellular parasites and is a consequence of many constraints encountered in the intracellular environment. The process involves loss of many protein-coding genes, resulting in greater dependence on the host, and loss of non-coding DNA such as intergenic regions, which has a direct impact on regulation of genome function. I have chosen two evolutionarily distinct systems to analyze the stages and functional consequences of genome degeneration, namely the impact of genome compression on transcription in an obligate parasite *Antonospora locustae* (genus Microsporidia), and gene content in the mitochondrion of a diatom endosymbiont found in the dinoflagellate *Durinskia baltica*. I have successfully mapped transcriptional start and termination sites from 14 loci in *Antonospora locustae*, and cloned fragments of two genes that are part of the electron transport chain from the mitochondrion of the diatom endosymbiont in *Durinskia baltica*. My analysis reveals that transcription in *A. locustae* is always initiated immediately upstream of the open reading frame at a single point for every locus, whereas transcriptional termination can occur at several points for a single gene and, in some instances overlaps with a downstream reading frame. The identification of NADH5 and ATPase9 from the mitochondrion of the endosymbiont in *D. baltica* is further evidence for the preservation of function in this enigmatic organelle.

TABLE OF CONTENTS

Abstract	ii
Table of contents	iii
List of tables	iv
List of figures	v
Acknowledgements	vi
Dedications	vii
Co-authorship statement	viii

CHAPTER 1 Introduction

1.1 Genome degeneration	1
1.2 Parasitism	3
1.3 Endosymbiosis	3

CHAPTER 2 Transcription in compressed genomes of microsporidia

2.1 Introduction	6
2.2 Materials and methods	14
2.3 Result	18
2.4 Discussion	22

CHAPTER 3 Mitochondria in diatom endosymbiont of dinoflagellates

3.1 Introduction	26
3.2 Materials and methods	32
3.3 Results and discussion	34
Literature cited	40

List of Tables

- 2.1 Nested primers used in RACE PCR to amplify
3' and 5' cDNA ends of 14 different loci in *A. locustae* 17
- 2.2 Length of untranslated 5' and 3' regions and transcriptional
overlap as determined by alignments of RACE products with
genomic contigs of *Antonospora locustae*. 19
- 3.1 Degenerate primers used to amplify NADH5 and ATPase9
from the mitochondrion of the endosymbiont of *Durinskia baltica*.33

List of Figures

2.1	Transcriptional start and end sites as represented by RACE products for 14 loci in <i>A. locustae</i>	20
3.1	Maximum likelihood phylogeny (ML) of ATP synthase F0 subunit 9 from the endosymbiont of <i>D. baltica</i>	35
3.2	Maximum likelihood phylogeny (ML) of NADH5 from the endosymbiont of <i>D. baltica</i>	36

Acknowledgements

Many thanks to my Supervisor Patrick Keeling for his continued guidance and support. For providing me with the freedom work independently and acquire new skills and knowledge. For helping me cope with setbacks and steering me in the right direction.

I thank members of my committee and examining committee, Naomi Fast, Keith Adams, Brian Leander and Martin Adamson, for helpful questions and guidance throughout meetings and in preparation for graduation, as well as the Sean Graham for chairing my defense committee.

I thank all my labmates, in particular Dr. Nicholas Corradi for his help and collaboration, which made our publication possible. Gillian Gile for helping me learn how to use analytical computer programs employed in the lab and solving general computer-related hick ups.

Dedications

To my parents for their continued emotional blackmail that pushed me to succeed in school (or else...).

To my colleague and partner in crime (literally) Olena V. Morozova for long philosophical conversations about science which reminded me that what I do is not completely meaningless and mundane. For plotting evil schemes with me (see above), which added colour to my life and gave me the will to carry on.

To my friend and mentor Dr. Wenzislawa Ckushumova (Wenzi), my mothership in academia; for her encouragement and help which inspired me to stay in research.

To the Canadian army reserves for their financial support, in particular to all the guys in my unit, the Seaforth Highlanders of Canada for providing ample distraction and entertainment as well as copious amounts of alcohol during my ordeal of writing and studying in preparation for my defense.

Co-authorship statement

All work presented in this thesis was done by me except where stated otherwise. The contents of Chapter two were published in a manuscript, wherein Dr. Nicholas Corradi is the first author, and with the assistance and supervision of Dr. Patrick J. Keeling.

Corradi N., Gangaeva A., Keeling P.J. (2008). Comparative profiling of overlapping transcription in the compacted genomes of microsporidia *Antonospora locustae* and *Encephalitozoon cuniculi*. *Genomics* 91:388-393

Chapter 1: INTRODUCTION

1.1 Genome Degeneration

Microbial systems offer unique opportunities to study evolution as their high reproductive potential and short generation time offers ample opportunity for changes to the genome. Their lifestyle allows for relative ease of uptake and exchange of genetic information via lateral gene transfer. Finally, their adaptability to unusual niches and extreme environments offers a picture of full evolutionary potential of genomes; evolution at its extreme.

Several evolutionary scenarios have produced microbial genomes that are orders of magnitude smaller than those of their relatives. This includes a few free-living organisms with unusually high reproductive rates (Derelle et al. 2006), but is far more extreme in obligate parasites and endosymbionts. For instance, free-living *E.coli* has a genome of 4.6Mb, which is average for bacteria, whereas *Mycoplasma* and *Rickettsia*, both intracellular parasites have genomes of 0.58Mb and ~1Mb respectively. Even further reduction is seen in endosymbionts. Mitochondria, which have a common ancestor with *Rickettsia* have an average genome size of only ~30kbp (Gray et al. 2004). Free-living cyanobacteria which gave rise to chloroplasts have genomes ranging in size from 1.67Mbp to 4.8 Mbp (Cyanobacterial Genome Project) whereas a chloroplast genome of *A. thaliana* is about ~154kb, less than 10% the size of its smallest free-living relative (Sato et al. 1999).

Reduction in genome size can result from loss of non-coding DNA such as intergenic regions, transposons and repeat elements but also loss of genes and shortening of proteins (Andersson & Kurland 1998). As in other genomes, genes in endosymbionts

and parasites can be lost by drift, through accumulation of random mutations. Losses of large DNA fragments can occur through uneven crossing over of DNA molecules and via excision of transposable elements. These phenomena occur in most organisms, however in free living microbes they can be compensated for by uptake of DNA and repair by homologous recombination, or via sexual reproduction (Dale 2003; Andersson & Kurland 1998). The intracellular milieu encountered by parasites and endosymbionts however is not conducive to gene reshuffling. Endosymbionts usually lose their capacity for sexual reproduction, whereas parasites have limited opportunity for interaction among individuals inside the host. In parasites each infection results in a population bottleneck as it is initiated by only one or a few individual cells or spores and does not involve a polymorphic population that would allow for reshuffling of alleles necessary for elimination of deleterious mutations. Finally, the loss of many genes is made possible by alleviation of pressure to preserve these, as they are no longer essential in an intracellular environment (Andersson & Kurland 1998). Smaller genomes offer the advantage of rapid and energetically inexpensive replication, which may also favor loss of pseudogenes and non-coding DNA. Overall the intracellular milieu encountered by endosymbionts and obligate parasites is a highly unusual niche wherein the alleviation of most challenges encountered by free-living organisms is counterbalanced by the challenge to survive inside a host, calling for unique adaptations, biochemical remodeling and genomic reorganization.

1.2 Parasitism

A common adaptation to a parasitic lifestyle is the parasite's ability to exploit the host for acquisition of energy and metabolites. Many genes needed in biosynthetic pathways for the production of nucleotides, carbohydrates, fatty acids and amino acids, have been lost from genomes of parasites like microsporidia, as many of these compounds are obtained from the host (Metenier & Vivares 2001; Katinka et al. 2001). While some proteins are lost altogether others become shorter, usually due to loss of protein-interaction domains upon loss of interacting partners (Katinka et al. 2001; Keeling et al. 2005).

Genome compression involves the loss of non-coding DNA by essentially the same mechanism as the loss of protein coding genes. Whereas "junk" DNA such as transposons and repeat elements can be lost without immediate consequences for the organism, the loss or shortening of intergenic regions can have a profound impact on genome function as these regions bare essential cis-regulatory elements necessary to control transcription. In some species of microsporidia genome compression has resulted in high frequency of transcriptional overlap, making these organisms ideal to study the consequences and mechanisms of coping with genome compression.

1.3 Endosymbiosis

It has been generally accepted that the original endosymbiotic events that gave rise to plastids and mitochondria occurred when a heterotrophic cell engulfed a cyanobacterium and alpha-proteobacterium respectively (Gray & Doolittle 1982). These events are thought to have occurred only once or twice (in the case of plastids) in

evolutionary history and are termed primary endosymbioses. Most of extant plastid diversity came about from secondary and tertiary endosymbioses where a heterotrophic eukaryote engulfed a photosynthetic one. In all cases of endosymbiosis, the prey, instead of being completely digested was incorporated into the host and reduced to a plastid (Archibald & Keeling 2002; 2005).

Genome reduction in endosymbionts is twofold; whereas genes needed for autonomous function such as mitotic cell division and motility are lost altogether, many genes needed for maintenance and replication of the endosymbiont are transferred to the nucleus of the host (Timmis et al. 2004). In the event of secondary or tertiary endosymbiosis, the endosymbiont also undergoes morphological degeneration losing all organelles and ultrastructural features except the plastid. Mitochondria, which are endosymbionts themselves, are almost always lost from eukaryotic cells that undergo reduction to become plastids. Exceptions to this rule are found among the dinoflagellates, some of which contain highly preserved endosymbionts with functional mitochondria. Whereas many studies have been done on mitochondria of free-living organisms, little is known about mitochondria of endosymbionts as these are extremely rare events, looked upon as evolutionary miracles.

Mitochondrial genomes

Mitochondrial genomes can vary greatly in structure and composition. In most organisms the mitochondrial genome is present in multiple copies of a circular-mapping chromosome (with a few exceptions), with high AT content (60-80%) and high gene density. Over 90% of the mitochondrial genes have been transferred to the nucleus and

are targeted back to the organelle. Genes commonly retained in the mitochondrial genome are key players in energy production, namely proteins that are part of inner-membrane complexes forming the electron transport chain and needed for oxidative phosphorylation (Gray et al. 1998)

Though the full gene complement necessary for mitochondrial function is similar in all eukaryotes, the complement of genes retained in the mitochondrial genome varies significantly between different organisms. Analysis of *Oxyrrhis marina*, an early-branching dinoflagellate revealed the mitochondrial genome to be highly fragmented and reduced, containing only COXI and COX3-COB fusion genes, along with a few ribosomal RNA's (Slamovits 2007). The mitochondria of diatoms *Phaeodactylum tricornutum* and *Thalassiosira pseudonana* (44kbp) on the other hand have retained NADH genes 1-7, 9,11, ATP synthase genes 6,8 and 9, COX1-3, COB and a handful of ribosomal protein genes (Armbrust et al. 2004; Marie-Pierre Oudot le-Seqc, personal communications).

To obtain a full picture of mitochondrial biochemistry and function would require the identification of all nuclear-encoded mitochondrial proteins along with those found in the mitochondrial genome. However the degree of mitochondrial genome conservation and particularly the presence therein of key functional proteins can be used as tell tale signs of functional integrity.

CH 2 : TRANSCRIPTION IN COMPRESSED GENOMES OF MICROSPORIDIA

2.1 INTRODUCTION

Microsporidian Ecology and Pathogenicity

Microsporidia are intracellular eukaryotic parasites with a broad host range, from protists and arthropods, to birds, fish and mammals, including humans. The phylum *Microspora* comprises about 160 genera and 1300 described species (Larsson 1999). While some species have a detrimental impact on economically important animal and insect populations, others are being used (or have potential) as biological control agents against insect pests and disease-bearing vectors such as the malaria mosquito (Weinzierl & Henn 1989). In humans microsporidia can cause infections with severe to lethal outcome for immunocompromised individuals, such as AIDS patients or people on immunosuppressive drugs (Bryan & Schwartz 1999). Transmission usually occurs by ingestion or inhalation. Initial infection causes damage to intestinal or lung epithelia leading to malabsorption and respiratory problems respectively, and can disseminate to other organs such as the CNS, eyes, liver, kidney, spleen etc... Depending on the site of infection the symptoms may include diarrhea, nausea, vomiting, weight loss, headaches, loss of vision and cognition as well as destruction of infected tissues leading to severe and permanent organ damage (Franzen & Muller 2001). The spread of HIV as well as increased use of immunosuppressive drugs renders otherwise opportunistic parasites a growing concern which necessitates greater understanding of basic biology and genetics of these poorly studied systems.

Morphology and Lifestages

Microsporidia persist in the environment as stress-resistant spores about 1-40 μm in diameter. Spore shapes vary between species, ranging from simple spheres, ovals or rods to extravagant spirals. The outer spore layers consist of proteinaceous exospore and a chitin endospore, which protect the spore from mechanical damage and desiccation. The hard spore wall is subtended by the plasma membrane, which encloses a single cytosolic compartment containing one to two nuclei, depending on the species. The posterior of the cell is occupied by a large vacuole. Spanning the length of the spore and anchored at the apex of the cell is a conspicuous coiled tube called the polar filament. Close to the anchoring point of the polar filament, we find a membranous system known as the polaroplast; both of these are key components of the infection apparatus. A striking cytological feature of microsporidia is the absence of membrane-delimited organelles found in most eukaryotes. Microsporidia do not have canonical mitochondria, lysosomes, peroxisomes or secretory granules. A highly reduced membrane system studded with polyribosomes and adjacent to the nucleus is all that remains of the endoplasmic reticulum. There is no prominent golgi apparatus in the spore, though some reduced forms thereof have been detected in intracellular stages of the infection (Vavra & Larsson 1999).

The infectious cycle is initiated when the polar filament everts under the pressure of a swelling posterior vacuole and pierces the host cell. The cytoplasmic contents of the spore enveloped in the polaroplast are transferred through the polar filament into the host cytosol, much like the contents of a syringe through a needle (Cali & Tarkavian 1999). Once inside the host, microsporidia enter an intracellular amoeboid stage of their life cycle termed merogony; a period of high metabolic activity and reproductive rate. The

meronts reproduce in the host cytosol or in parasitophorous vacuoles, utilizing nutrients and macromolecules of the host. After one or several replication cycles the meronts transition into sporogony, forming spores with complete cell wall and infectious apparatus. The infectious cycle terminates with the release of mature spores, causing lysis of the host cell (Cali & Tarkavian 1999).

Cell division occurs by either binary fission (division of a single mother cell into two daughter cells) or schizogony (division of nuclei and subsequent splitting of multinucleate cell into multiple daughter cells) (Cali & Tarkavian 1999). Sexual reproduction occurs during the intracellular stages in some species but appears to be rare; little is known of the process other than light-microscopic observations of homogamous fusion, karyogamy and meiosis (Hazard & Brookbank 1984).

Phylogeny

Early classifications based on morphology placed microsporidia among the sporozoa, a group of spore-forming parasites, which comprised members of apicomplexa, myxosporidia, actinomyxidia and haplosporidia (Kudo 1947). In the 1980's they were reclassified as archeozoa; an ancient eukaryotic branch that predates the endosymbiotic event which lead to the acquisition of mitochondria, and development of complex organellar compartmentalization (Cavalier-Smith 1983).

Several lines of evidence from molecular analyses, such as the reduced size and peculiar structure of macromolecules reinforced the idea that microsporidia are an ancient eukaryotic lineage, transitional between prokaryotes and eukaryotes (Vossprinck et al. 1987; Cury et al. 1980). For instance microsporidial ribosomes appear prokaryotic in

size and composition, having a sedimentation coefficient of 70S and comprising 50S and 30S large and small subunits respectively as well as lower protein content (50% rather than 60%) reminiscent of prokaryotes (Curgy et al. 1980). However, these proved to be derived rather than primitive features which came about as a result of reductive evolution. Despite some contradiction among different gene phylogenies (Vossbrinck et al. 1987; Kamaishi et al. 1996), most evidence now places microsporidia as relatives of fungi (Edlind et al. 1996; Hirt 1999; Keeling 2003; Keeling & Doolittle 1996; Van de Peer 2000).

The final evidence against the conclusion that microsporidia are primitively amitochondriate ancient eukaryotes was the discovery of mitochondrial proteins in the nuclear genome of several microsporidian species, including the heat shock protein 70 (HSP70) (Arisue et al. 2002; Germot et al. 1997; Peyretailade et al. 1998) and pyruvate dehydrogenase E1 (Fast, Keeling 2001). Subsequent immunolocalization studies of HSP70 in meronts of *Trachipleistophora hominis* (Williams et al. 2002) revealed the existence of a relict organelle, the mitosome, which lost its genome and key ultrastructural features such as cristae, but apparently had some nuclear proteins targeted to it. Since then the sequencing of the genome of *E. cuniculi* has revealed the existence of many mitochondrial proteins (Katinka et al. 2001), solidifying the idea that these organisms once had functional mitochondria.

Most of the controversy regarding the phylogenetic placement and identity of microsporidia arose from their extreme reduction at the morphological and molecular levels. However it is becoming increasingly clear that these peculiarities are highly derived rather than primitive features, which arose over evolutionary time due to

pressures of a parasitic lifestyle, making microsporidia highly specialized fungal parasites whose morphology and biochemistry have been streamlined for rapid and effective infection and replication in the host.

Genomics of Microsporidia

Some of the smallest eukaryotic genomes known to date are found among microsporidia. The smallest microsporidian genome is that of *Encephalitozoon intestinalis*; being only 2.3Mbp, it is half the size of the *E.coli* genome. Other *Encephalitozoon* species have slightly larger genomes, with 2.5 Mbp in *E. hellem* and 2.9Mbp in *E. cuniculi*. *Antonospora locustae*, another model microsporidian, has a 5.4 Mbp genome, about half the size of yeast (Metenier & Vivares 2001; Keeling & Fast 2002)

Most of the information we have on microsporidian genome structure and content comes from the complete sequence of the genome of *Ecephalitozoon cuniculi* (Katinka et al. 2001) and partial genome surveys of several other species including *Antonospora locustae* and *Enterocytozoon bieneusi*, as well as much larger genomes of *Brachiola algerae* and *Edhazardia aedis* (Williams et al. 2008).

Karyotype studies reveal microsporidian genomes to be organized on several linear chromosomes, ranging in size from 175 to 2700 kb (Metenier & Vivares 2001; Biderre et al. 1994). Analysis of the smaller genomes, those of *E. cuniculi* and *A. Locustae* revealed similar chromosomal organization typical of most eukaryotes, with chromosome cores dense in protein-coding genes, flanked by ribosomal operons and ending in telomeres (Katinka et al. 2001; Vivares et al. 2002).

Based on genome comparison studies it appears that *A. locustae* and *E. cuniculi* have a similar gene complement of just under 2000 protein-coding genes and that most of the difference in genome size is a result of gene duplication and the presence of gene-sparse regions in *A. locustae*. Genome reduction is manifested in the absence of genes needed for biosynthetic pathways of certain carbohydrates, lipids and amino acids, all of which are obtained from the host (Katinka et al. 2001). Genes remaining are those needed for replication and expression of genetic information, structural genes composing the spore wall and infectious apparatus and transport proteins needed to shuttle essential metabolites from the host to the parasite during intracellular stages. Many remaining genes appear to have been truncated. About 85% of the proteins in *E. cuniculi* are shorter than their yeast homologues, on average by 14.6%, likely due to loss of protein-protein interaction domains following the loss of interacting partners (Katinka et al. 2001; Keeling et al. 2005).

Whereas genome reduction in *E. cuniculi* and *A. locustae* is evident in the loss of functional groups of proteins, genome compression is manifested in the absence of selfish DNA elements such as transposons, significant reduction in number of minisatellite and microsatellite repeats, introns and length of intergenic spaces (Katinka et al. 2001). Interestingly recent genome surveys of larger microsporidian genomes of *Brachiola algerae* and *Edhazardia aedis* revealed these to have a gene complement similar to that of *E. cuniculi* and *A. locustae*, with the majority of genome size difference being accounted for by the presence of large amounts of non-coding DNA. The genomes of *Brachiola algerae* and *Edhazardia aedis* are padded with large intergenic regions that can

be over 2kb long and abundant transposons, resulting in gene densities orders of magnitude lower than in microsporidia with smaller genomes (Williams et al. 2008). Complete genome sequences from microsporidia are needed for adequate comparison of protein content between species, however genomic data obtained thus far indicates that genome reduction is more common among species of microsporidia, whereas genome compression is a unique phenomenon restricted to fewer species. Though the evolutionary forces driving genome compression remain enigmatic the immediate consequences of this process on genome function are readily observable.

Genome compression and multigene transcripts

Perhaps the greatest impact of genome compression on its function is the shortening of intergenic DNA. The smaller microsporidian genomes are extremely gene-dense; *E. cuniculi* and *A. locustae* have a gene density of about 0.97genes/kb and 0.94genes/kb respectively which is about twice that of yeast. The average length of intergenic regions is only 129bp in *E. cuniculi* and 211bp in *A. locustae* (Katinka et al. 2001). Yeast, which is considered to have a relatively compact genome compared to other eukaryotes has intergenic region of ~600bp on average for divergent ORFs (Dujon 1996).

Intergenic regions in Eukaryotes contain transcriptional control elements such as promoters and terminators, which direct transcriptional machinery to a given ORF and terminate transcription respectively. Higher organisms with complex developmental programs requiring spatial and temporal regulation of gene expression have evolved complex modular promoters that can be activated by different combinations of

transcription factors, which are in turn controlled by a variety of external stimuli. These cis-regulatory elements are harboured in large intergenic regions that can be larger than entire genomes of simpler organisms. Simpler unicellular organisms can get away with relatively minimal promoters but nevertheless necessitate these cis-regulatory elements to control transcription. Whereas disposal of junk DNA such as repeat elements may not have an immediate physiological consequence for the organism, reduction of intergenic region can have a severe impact on genome function by affecting transcription.

Surveys of EST's in *Antonospora locustae* and *Encephalitozoon cuniculi* revealed a peculiar pattern of gene expression in microsporidia. From 871 clones, which had identifiable homologues in the public database, 11% encoded more than one gene whereas 17% contained sequence that is antisense to a known protein-coding gene. These transcripts are not likely to be part of a post-transcriptional control such as RNAi as they are longer, usually containing up to three genes or gene fragments in different orientations (Williams et al. 2002). The alternative explanation is that microsporidia, much like prokaryotes have operons; in other words several genes controlled by the same promoter are transcribed onto a single polycistronic mRNA which is translated into several proteins. However this is an unlikely explanation since the mRNA in *A. locustae* and *E. cuniculi* contains genes oriented in different directions and gene fragments which are unlikely to produce functional protein products.

Rather than being a complex regulatory mechanisms found in higher eukaryotes or a reversion to operon-like transcription seen in bacteria overlapping gene expression in microsporidia appears to be a by-product of genome compression which resulted in displacement of transcriptional control elements into or past neighboring genes.

As a first step in understanding transcriptional regulation in the compressed genomes of microsporidia I have sequenced full-length 5' and 3' cDNA ends from fourteen genetic loci in *Antonospora locustae*, using RLM-RACE PCR (RNA ligase-mediated amplification of cDNA ends). RACE products (cDNA ends) were subsequently aligned with large contigs assembled from the partial genome sequence survey found in the *Antonospora locustae* database, in order to assess the position of transcriptional start and end sites in relation to neighboring genes.

2.2 MATERIALS AND METHODS

Contig assembly

Sequences of *A. locustae* nuclear genome were obtained from the *Antonospora Locustae genome database* (<http://gmod.mbl.edu/perl/site/antonospora01?page=intro>) and EST project found in PepDB (<http://amoebidia.bcm.umontreal.ca/pepdb/searches/welcome.php>). These were aligned in Sequencher (version 4.2) to create larger contigs, ranging in size from 3.5kb to 40kb. Genes selected for amplification by RACE PCR were those known to be expressed (present in the EST database), and positioned more than 500bp away from the ends of a contig. Genomic regions chosen for the analysis were relatively gene-sparse with no overlapping ORFs.

RNA Isolation

Antonospora locustae spores (M&R Durango Inc.) were ground on liquid nitrogen with mortar and pestle. Full RNA extraction was performed using Trizol reagent (Invitrogen).

RLM-RACE PCR

EST libraries are usually constructed by random cDNA amplification using oligo dT primers, which amplify mRNA's from the poly-A tails. This method often results in transcripts which are truncated at the 5' end, and thus does not allow to determine the location of transcriptional start sites. RLM RACE PCR on the other hand allows for the isolation of full-length 5' ends, enabling the identification of transcriptional start sites. First Choice RLM-RACE (Ambion) and Taq DNA polymerase (Invitrogen) were used to amplify full-length 3' and 5' cDNA ends.

For 5' RACE whole RNA was treated with CIP (calf intestinal phosphatase) to remove phosphates from uncapped transcripts (truncated mRNA), as well as tRNA, rRNA and contaminating DNA. The RNA was subsequently treated with TAP (Tobacco Acid Pyrophosphatase) to remove caps from full-length mRNAs exposing 5' monophosphates. A 45bp RNA adapter was ligated to the mRNA using T4 ligase. In this step only phosphorylated mRNA's acquired an adapter, as the ligation can only occur in the presence of a free 5' phosphate. The cDNA ends of the desired genes were subsequently amplified by PCR using adapter-specific and gene-specific nested primers (Table 2.1). To ensure that the products were not a result of DNA or truncated mRNA contamination the RT/PCR steps were repeated using mRNA that was treated with CIP

but not TAP. All loci that yielded the same products in these two analyses were discarded from the data as contaminants.

In 3'RACE an adapter containing a poly-T tail was used with adapter-specific and gene-specific nested primers (Table 2.1) to amplify genes by Reverse-Transcription PCR.

Cloning and sequencing

Secondary PCR was performed using Taq DNA polymerase (Invitrogen).

Secondary PCR products were cloned into TOPO TA vector and TOP-10 competent E.coli cells (Invitrogen). Eight clones from each PCR product were sequenced using NAPS capillary sequencing with BigDye 3.1. The obtained sequences were aligned to pre-constructed contigs of genomic and EST sequence data. All 5'RACE products found to begin downstream of the predicted translation initiation site of the gene of interest were discarded as contaminants from truncated transcripts. All 3'RACE products where the poly (A) tail corresponded to a poly (T) stretch in the genomic DNA were discarded as possible contaminants where the adapter bound a random A-rich region on the mRNA rather than a Poly (A) tail.

Table 2.1 Nested primers used in RACE PCR to amplify 3' and 5' cDNA ends of 14 different loci in *Antonospora locustae*

<i>Antonospora locustae</i> primers			
Locus	Accession number	Primer sense	Primer sequence
Checkpoint protein kinase	AF406785	3'inner	CGA AGA ACA TCT CAA ACG TCT GTT GCA
		3'outer	CGT TGC TGA AGA CTC TGT CAT TGC GTT
		5'inner	CTT CCT TGG TGG ACT TGA AAG TCT TCG A
		5'outer	CCA GAC TCA GCA CAA TGT AGA AGT TGC
CTP synthase		3'inner	GTT GAG GCT GCT GTT CTC AGA GAT CAC
		3'outer	AAG CCA CTT GGA GTT GTC GTG CAC GGA
		5'inner	ATA CCT CGG AGC TAA TGT CGC ATG ATA
		5'outer	TGC ATC TAC CTC GCT TCC ATC AGC A
Cytidine deaminase	AY548885	3'inner	GAG TGC AGT GAG GCT GTT GCA ACG CTT
		3'outer	CCG ATC TTC GGC GGC ACA ACT GTG CTT
		5'inner	AGA TCA CTG CGT CTT CCC GCA CTA CAA
		5'outer	ACG CAA CTA GTT CCG GCT GCG CAA GTG G
DNA-binding Domain containing protein		3'inner	CTT ATA GTG ACA TGA TCT GCT GCG ACT C
		3'outer	CAG AAG GAA CGT TCT GCA AGT GCA GAC
		5'inner	TTG TAG TGC CTC GCC TTG TGC GGA TAC
		5'outer	ACT TCC GCC TGT ACG TTT CGA TGC TTC
WD repeats containing protein		3'inner	AGTATGGAGCTCCTCGGCACGATCGA
		3'outer	TTGTGTCTGCGGACCGGTGCATCAAGA
		5'inner	ACAACGCCATCGCGCAGCACAGTGT
		5'outer	ACGACGAGGCCTCCAAGGCAGTCAAA
Internalin		3'inner	TAT AGA GTT GAA CAT AAG CTC GTG CAA
		3'outer	CGA TAT AAT CGC AAT TGC ATG CAT GAA
		5'inner	CAC TTC TGC CTG TAG ATG CTC AGA CAC
		5'outer	CCT CGC AAT CCA ATG TTA TAT CTG GCT C
Lea domain containing protein		3'inner	CGA GAA GAC GAA GAG CGG TGC AGC AAG
		3'outer	AGC GAC AAC ATG AAG CAC GGA TAC GAG
		5'inner	CTC TCC GGC ACT GCC AGC ACC ATG CTT
		5'outer	CTC GTA TCC GTG CTT CAT GTT GTC GCT
Lipoprotein		3'inner	GGA AGA GGA AGA TAT CGA GTA CAT GAA
		3'outer	GCT AGG CGA CGA GGA GTA TGC AAG AAG
		5'inner	TTC TGG AAG TTC TGT CAG TCG GTT GCC A
		5'outer	GTT GCC ACC AAG ACC AAG CTT CTT CAG
Phosphate transporter		3'inner	GCG CAC ATG TGC AAG AAC AAG AGG CTT
		3'outer	CTT GAC GGT GCC TAT TGA CTG GAT TA
		5'inner	CAC GAG CTG AAC TAG CTG CGA GAC CAT
		5'outer	TCC TCT CCT ACT GCG ACT CCG TTC TG
P-68 like protein		3'inner	ACC GAT ATG GAC AGA AAC AAC GCA CGG
		3'outer	TTA CGT GCA CCG TAT CGG AAG AAC AGC
		5'inner	CAT ACC GAC TGT GCG AGC CTC GAT CT
		5'outer	GCA AGT CCA CGT CAA TCG GAC GCA GC
Ribosomal protein L9	AY548900	3'inner	AGA CTC TAT TGC GCG GCT CGG CAC C
		3'outer	CGA CGG AAA GAC GGT AGT CGT GAA GAA
		5'inner	AAG GTC GCG TAC AGC AGT AGC CCT
		5'outer	GAT GCA CAC GTA ATC AGC TTG CTG C
SnoRNA associated RNP		3'inner	CAT CTC CAA ACT TAT CCT TAA GCG CAC
		3'outer	GTT CAG TGC ATC TCG TTC TCT CTT CGA
		5'inner	GAG AGC CTT CAT GTA CTC ACG TTC TAG
		5'outer	GTA CGT AAG GGC TGA GAC GAA CGG TT
Ubiquitin		3'inner	AGC ATC TGC ACT CGC TAT AAC GCC GCT T
		3'outer	TCA CTG TGG ACG GAG ATC CTT ATG GAT
		5'inner	GAA TCA ACC ATG ACA AGC ACA CGG AAG A
		5'outer	TGT CTT GTA GCG AAA CTG TCT GTT ATC AA

2.3 RESULTS

Thirteen 5'RACE and 3'RACE products from fourteen different loci were successfully cloned and analyzed. The intergenic regions flanking the genes analyzed ranged from 11 nucleotides to 903 nucleotides with an average of ~288 nucleotides, and no overlapping Open Reading Frames (Table 2.2).

5'RACE

Most 5'RACE products began immediately upstream of the translation initiation site ATG, with the 5' untranslated region ranging in length from 0 to 9 nucleotides. A single 5' RACE product was isolated for every gene analyzed, implying that transcriptional start site is always the same for a given gene. In two instances, where the intergenic region upstream of the gene of interest was particularly short, 5' RACE PCR resulted in amplification of transcripts containing upstream ORFs. This includes the 5'RACE product of Cytidine deaminase, which contains an upstream hypothetical protein, and that of Ubiquitin Carboxy terminal hydrolase, containing VID24 gene (Fig2.1).

3'RACE

3'RACE products showed greater degree variation in length than 5'RACE; 3' untranslated regions ranged in size from 14 to 299 nucleotides with an average length of ~180 nucleotides. Most transcripts ended in intergenic regions however that of Checkpoint kinase ran through the 110 nucleotide intergenic region and into the downstream gene, PDH-E1 α with a ~100 nucleotide overlap. Most genes in this study

had a single 3'RACE product, except for Internalin which had two distinct products, with 10 and 262 nucleotide 3' untranslated regions.

Table 2.2 Length of untranslated 5' and 3' regions and transcriptional overlap as determined by alignments of RACE products with genomic contigs of *Antonospora locustae*.

Gene Name	Initiation Site ^a	5'Overlap	Poly (A) site ^b	3'overlap
P68-like protein	1	None	104	None
Phosphate transporter	0	None	299	None
Lea protein	0	None	74	None
Rpl9	0	None	83	None
Sno-RNA associated Protein	9	None	38	None
DNA-binding protein	1	None	258	None
CTP synthase	0	None	14	None
WD-repeat protein	4	None	—	—
Cytidine deaminase	505	485	673	None
Checkpoint kinase	5	None	211	PDH-E1 α
Phosphoesterase	—	—	374	None
Internalin	6	None	10/262	None
Lipoprotein	0	None	74	None
Ubiquitin carboxy Terminal Hydrolase	549	525	53	None

^aBase pairs upstream of translation initiation site ATG of the gene of interest

^b Base pairs downstream of translation termination site TAA/TAG/TGA

Fig 2.1 Transcriptional start and end sites as represented by RACE products for 14 loci in *Antonospora locustae*. Gray arrows represent ORFs; black arrows, position of gene-specific primers; light grey arrows, 5'RACE products; dotted lines with poly (A) stretch, 3'RACE products.

A. P68-like protein



B. Phosphate Transporter



C. Lea 4 domain containing protein



D. Ribosomal Protein L9



E. Sno RNA associated Ribonucleoprotein



F. DNA-binding Domain related cluster



G. CTP synthase

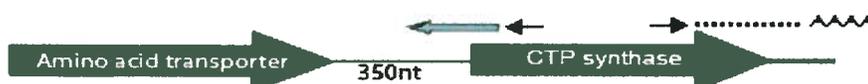
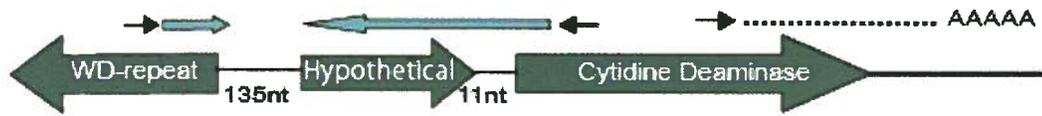
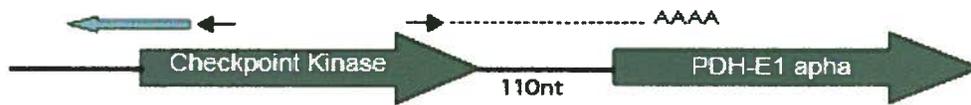


Fig 2.1 continued

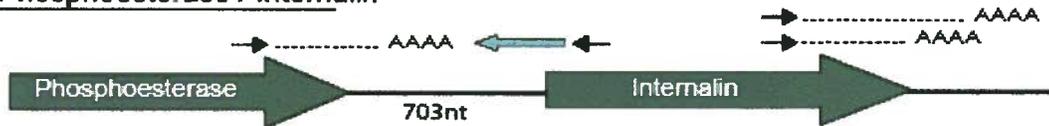
H. WD-repeat / Cytidine deaminase



I. Checkpoint Protein kinase



J. Phosphoesterase / Internalin



K. Putative Lipoprotein



L. VID25 / Ubiquitin



Ch 2.4 DISCUSSION

Strict transcriptional initiation sites and loose termination sites; functional and evolutionary implications.

Analysis of 5' RACE products indicates that in *A. locustae* transcription is initiated a few nucleotides upstream or immediately at the beginning of the open reading frame. There was no variation among clones of 5'RACE products for a given gene indicating that every ORF has a single transcription initiation point. Extremely short intergenic regions upstream of the genes analyzed resulted in amplification of mRNA containing the upstream ORF. The above indicates the presence of promoter sequences upstream of genes, which direct transcriptional machinery to a strict initiation site. Transcription was never initiated inside another gene in the loci I analyzed, however, previous work with the Photolyase gene in *A. locustae* indicates that this scenario is possible (Slamovits & Keeling P.J. 2004)

In the case of Ubiquitin and Cytidine deaminase, where 5'RACE products included the upstream genes, there are several possible explanations. One possibility in both cases is that we are looking at the 3' ends of the transcripts from the upstream genes, which run into the genes of interest downstream. This is consistent with the observation that transcripts can end in the middle of downstream genes (fig 2.1, I). In this case there may be transcripts for Ubiquitin and Cytidine deaminase beginning in their short intergenic regions, which have not been detected in the analysis due to preferential amplification of the 3'end of their neighbors. There is also a possibility that the two genes are co-transcribed with their upstream neighbors and translated from the same mRNA. To confirm this would require sequencing of the entire mRNA molecule from start to finish,

ensuring that both reading frames are complete, and isolation of the corresponding proteins to verify that both are successfully translated. More 5'RACE products must be analyzed to determine if the observed pattern holds throughout the genome, before further work is undertaken to determine how promoters are incorporated into compressed genomes.

Eukaryotic promoters

Eukaryotic promoters can be extremely complex and diverse, having combinations of cis-regulatory elements and enhancers, kilobases away from the beginning of the open reading frame which they control. These cis-elements are necessary for spatial and temporal regulation of gene expression, controlled by wide array of transcription factors which are themselves regulated by external stimuli and developmental cues. Simpler organisms can get away with simpler cis-regulation, however minimal promoters and sufficient intergenic space to prevent collision of transcriptional machinery are still necessary for gene expression.

A prevalent feature in eukaryotic minimal promoters is a consensus sequence known as the TATA box; TATAAA or a similar T/A rich region. This element may function alone or in conjunction with other sequences in positioning of transcriptional machinery upstream of the gene to be transcribed. During initiation of transcription these sequences are bound by TATA-binding protein (TBP), a subunit of a multimeric transcription factor TFIID, which recruits proteins of the pre-initiation complex as well as RNA polymerase II, required to melt the target DNA and initiate RNA synthesis. The TATA box is found about 25-30 nucleotides upstream of transcriptional start sites in

drosophila and humans, and 40-200 nucleotides in yeast. These distances are crucial for appropriate positioning of transcriptional machinery catalytic centre, and initiation of transcription at the desired location (Smale & Kadonaga 2003).

TBP is highly conserved among organisms at the functional level, and is present in *A. locustae* (Fast et al. 1999). Evidence from other organisms suggests a minimal distance requirement for positioning of transcriptional machinery upstream of the transcription initiation site is dictated by the binding of TBP and its interaction with other proteins in the complex (Smale & Kadonaga 2003). Extremely short intergenic regions in *A. locustae* may not accommodate the most basic promoters, as they do not provide sufficient space between the TATA box and transcriptional initiation site for the assembly of transcriptional machinery.

Unlike transcriptional initiation, transcriptional termination in *A. locustae* appears to be sloppy, with transcripts ending in several locations and within neighboring coding regions (Fig 2.1 I, J). Interestingly the opposite scenario was observed in another microsporidian *Encephalitozoon cuniculi*, where transcription termination was always at the same point for a given gene whereas transcriptional initiation was loose, starting at different points, often inside upstream open reading frames (Corradi et al. 2008). Contrasting data from these two species indicates that these organisms are independently evolving transcriptional regulatory elements that can be incorporated into neighboring genes, which allows them to cope with genome compression and elimination of intergenic regions. To fully understand how transcriptional regulation is evolving in compressed genomes a detailed analysis of promoters and terminators would be necessary. Future directions would involve identifying core promoter and terminator

regions as well as characterization of transcriptional components responsible for binding these sequences.

Untranslated regions in Eukaryotes

Another noteworthy feature of *A. locustae* transcripts is the extremely short length or in some cases absence of untranslated regions (UTR's). The average length of 5'UTR in eukaryotic mRNA is 100-200 nucleotides whereas 3'UTR can range from 200-1000 nucleotides. These untranslated regions contain sequences responsible for regulation of translation and sub-cellular localization, and can form complex secondary structures that recruit proteins controlling translation and mRNA export from the nucleus (Pesole et.al. 2001) These extra levels of regulation are essential in complex organisms to allow fine tuning of gene expression in response to environmental stimuli. *A. locustae* appears to get away without long UTR's as seen in other eukaryotes; it seems to have disposed of that level of regulation of gene expression and mRNA localization. Whether the morphological simplicity and dormancy of microsporidian spores have rendered this higher level of regulation unnecessary or it has been replaced by other regulatory mechanisms is to be determined.

Ch 3. MITOCHONDRIA IN DIATOM ENDOSYMBIONT OF DINOFLAGELLATES

3.1 INTRODUCTION

Dinoflagellate Ecology

Dinoflagellates are planktonic, predominantly unicellular protists, having one or more motile lifestages with two dissimilar flagella (Fensome et al. 1993). Dinoflagellates are found in benthic and pelagic habitats of marine and freshwater ecosystems, all around the globe, from the arctic to tropical seas. About half the known species have plastids and are either photosynthetic or mixotrophic. Many species are strictly heterotrophic, grazing on other planktonic populations. A few species lead parasitic or symbiotic lifestyles in association with other organisms (Taylor 1987).

Dinoflagellates have a great impact on aquatic ecosystems, being significant contributors to primary production, and as endosymbionts of reef-building corals (Taylor 1987). Many species of dinoflagellates produce toxic algal blooms known as red tides, as well as toxins which accumulate in consumers, spreading through the food chain, resulting in fish kills and toxic effects in birds and humans. Syndromes affecting humans include neurotoxic and paralytic shellfish poisoning (NSP) and ciguatera which are manifested in gastrointestinal, neurological and, less frequently cardiovascular effects, with 10% lethality (Steidinger 1984).

The tremendous ecological impact and diversity of dinoflagellates is reflected in their wide variety of adaptations, complex lifestages and unique biochemistry, making these organisms a goldmine for discovery in molecular biology and evolution.

Cell morphology

Dinoflagellates exhibit some of the most extravagant and diverse cell shapes. Different lifestages of the same organism were often mistaken for different species due to their distinct cell morphologies. At their vegetative stage most dinoflagellate species are motile haploid cells, which reproduce by mitosis. The cell wall is composed of membrane-bound amphiesmal vesicles or alveoli, found beneath the plasma membrane. In some species these are filled with cellulose granules, forming rigid plates of armor or theca whose arrangements can be used for taxonomic classification. The cell wall can form intricate protrusions giving the cells their extravagant shapes (Netzel & Durr 1984). Two dissimilar flagella, one transverse one longitudinal, emerging from a pore on the ventral side of the cell and running perpendicular to each other, propel the cell in a unique spiral motion (Taylor 1987).

Dinoflagellates possess all canonical features of eukaryotic cells, including the nucleus, golgi bodies, endoplasmic reticulum and mitochondria. Photosynthetic dinoflagellates have either secondary or tertiary plastids (or both) depending on the species. Dinoflagellates are unique among eukaryotes in their plastid diversity, as different species can have plastids of different origins that were acquired at different points in evolutionary history.

Dinoflagellate Plastids

Dinoflagellates' unique ability to take up endosymbionts resulted in extraordinary plastid diversity among species, with different combinations of chlorophylls and

accessory pigments. Most dinoflagellates have a peridinin-containing, three-membrane bound plastid derived from red algae; an ancestral plastid, characteristic of the chromalveolate group (Dodge, Jeffrey 1989). However a few lineages have replaced this plastid with secondary and tertiary endosymbionts of various origins. A secondary plastid derived from a green alga has been found in *Lepidodinium* where it replaced the red algal peridinin plastid (Watanabe et al. 1990). Cases of tertiary endosymbiosis have been observed in different genera with endosymbionts of haptophyte (Tengs et al. 2000), cryptophyte (Hackett et al. 2003; Schnepf & Elbrächter 1988), and diatom origins. These groups of organisms are chromists, a sister group to the alveolates, containing a red algal plastid themselves.

Canonical cases of endosymbiosis involve complete integration of an endosymbiont into the host, resulting in loss of organelles and leaving little more than the plastid surrounded by a series of membranes (McFadden 2001). However, dinoflagellates exhibit different degrees of plastid integration. From transient associations such as kleptoplasty, where a photosynthetic prey is used as a plastid by its heterotrophic predator prior to being digested (Fields & Rhodes 1991; Schnepf & Elbrächter 1992; 1999; Skovgaard 1998) to complete integration and reduction of an endosymbiont to an organelle, entirely dependent on the host. Dinoflagellates' promiscuity and relative ease of plastid uptake resulted in more frequent endosymbiotic events throughout evolutionary history. As a result these organisms offer a unique opportunity to observe different stages of plastid integration and different degrees of endosymbiont reduction at the ultrastructural and genomic levels.

Dinoflagellates with diatom endosymbionts

Durinskia baltica and the closely related *Kryptoperidinium foliaceum* are two dinoflagellate species with endosymbionts derived from pennate diatoms, which have conserved many cellular components normally lost in plastids (Chesnick et al. 1996; 1997). The diatom endosymbiont lost its cell wall and motility and has no prominent ultrastructural features of its free-living relatives. The endosymbiont is present in all lifestages of the host and their division is synchronized. It is separated from the host cytoplasm by a single membrane (Tomas & Cox, 1973; Chesnick et al. 1996; 1997). Interestingly the diatom occupies most of the host cytoplasm, and appears to be taking over the host from the inside. Furthermore it has preserved its nucleus and mitochondria with intact cristae; the endosymbiont mitochondria outnumber those of the host. The endosymbiont plastid is surrounded by four membranes; the outermost being continuous with the diatom ER. The host cytosolic compartment has all typical features of dinoflagellates, with the dinokaryotic nucleus, an eyespot (remainder of the red-algal plastid), and mitochondria (Tomas & Cox 1973). Ultrastructural observations indicate that the endosymbiont may be at its early stages of integration; consequently we expect to be able to observe the initial stages of genomic deterioration, the end product of which we see in most extant plastids. Another possibility is that the diatom is resisting degeneration and preserving functional autonomy to a greater degree than seen in other endosymbionts, which would make this a unique event in evolutionary history.

Dinoflagellate Genomes

When referring to the dinoflagellate genome one thinks of the nuclear or dinokaryotic genome. However dinoflagellates are mosaics of genetic information acquired by lateral gene transfer and gene reshuffling between the host and endosymbionts. Dinoflagellate nuclear genomes are among the largest known to date, ranging from about 3000 Mbp to 200 000 Mbp (65 times larger than the human genome) (Spector 1984). This vast amount of DNA is found on chromosomes which are permanently condensed during the cell cycle. Rather than having canonical nucleosomes the chromatin is held together by histone-like proteins (HLP), which resemble those of bacteria (Dodge 1966). The extraordinary genome sizes in these organisms have rendered whole genome sequencing and assembly very challenging; thus most molecular data comes from EST surveys. There is very little genomic data available from the diatom-bearing *D. baltica* and *K. foliaceum* and thus we cannot say with certainty how much gene reshuffling occurred between the different genomes, or which genes have been transferred to the dynokaryon. Organellar genomes on the other hand are relatively small, more amenable to analysis and can yield insight into functional conservation of the organelle.

Dinoflagellates with tertiary endosymbionts contain perhaps the greatest collection of genomes in a single cellular environment. *D. baltica* and *K. foliaceum* have two nuclear and two mitochondrial genomes belonging to the host and tertiary endosymbiont. They also have a plastid genome belonging to the tertiary endosymbiont. It isn't clear how much remains of the red-algal plastid genome from secondary

endosymbiosis, however studies in other dinoflagellate species indicate that we are likely to find remnants thereof in the dinokaryotic nucleus (Fast 2001; Patron 2006).

The peculiarity of these species is the unusual degree of preservation of the tertiary endosymbionts at the ultrastructural and genetic levels. The diatom nucleus has retained HSP90, actin and tubulin genes. The latter two are particularly surprising as actin is a component of the cytoskeleton whereas tubulin is needed for motility and mitotic spindle formation (McEwan and Keeling 2004); however the endosymbiont is no longer motile and divides amitotically. The mitochondria of both host and endosymbiont have been found to express genes from the electron transport chain, indicating that both organelles are likely to be functional. Cytochrome oxidase genes, COX1-3 as well as COB are expressed in the endosymbiont along with the small subunit rRNA (LSUrRNA); two of these (COX1 and COB) are also expressed in the host (Imanian & Keeling 2007).

Mitochondrial genome of the endosymbiont of *D. baltica*

Previous work in the lab revealed the presence of COX1-3, COB (complex II and IV of electron transport chain) and LSURNA genes in the mitochondrial genome of the endosymbiont of *Durinskia baltica*. My goal was to sequence new markers from the mitochondrial genome of the endosymbiont, in order to determine the degree of genome degeneration in this enigmatic organelle, with the hope to further our evaluation of its functional integrity. Here I report the sequencing of two more markers which are components of complexes I and V in the electron transport chain. Functional components of these complexes have not been previously found in *D. baltica*. Identification of these genes supports the notion that the endosymbiont mitochondria have a functional electron

transport chain capable of generating ATP, making these organelles unique phenomena in the history of evolution.

3.2 MATERIALS AND METHODS

Culture growth and nucleic acid isolation

Cultures of *Durinskia baltica* CS-38 obtained from CSIRO Microalgae Supply Service (CSIRO Marine and Atmosphere Research Laboratories, Tasmania, Australia) were maintained in f/2-Si medium on a 12/12 light/Dark cycle at 22 °C. A combination of antibiotics, including 25 µg/ml penicillin G, 25 µg/ml streptomycin, 240 µg/ml gentamicin and 46 µg/ml ciprofloxacin was added to all cultures used for molecular work to eliminate bacterial contamination. Cells were harvested by several rounds of centrifugation at 8°C for 5min @4000Xg. For nucleic acid extraction ~100mg of cells were crushed on liquid nitrogen with mortar and pestle. Whole cell DNA was purified using DNeasy Plant DNA isolation kit (Qiagen, Mississauga, Ontario). RNA was extracted using TRIZOL reagent (Invitrogen).

Primer design and Cloning

Sequence of the mitochondrial genome of *Phaeodactylum tricornutum* was graciously provided by Marie-Pierre Oudot-Le Seqc. Protein sequences from this genome were aligned with those of a related diatom *Thalassiosira pseudonana* (CCMP1335) using CLUSTALW. Degenerate primers were designed to most conserved stretches of protein sequence alignments (table 3.1). PCR and RT PCR were performed using Taq DNA polymerase and Superscript III one-step RT system respectively (Invitrogen).

Products of the appropriate size were gel-purified using the Qiagen gel extraction kit. These were cloned into TOPO TA cloning vector and TOP10 competent cells (Invitrogen). DNA purified from eight separate clones was sequenced using Big dye 3.1 capillary sequencing.

Table 3.1 Degenerate primers used to amplify NADH5 and ATPase9 from the mitochondria of the endosymbiont of *D. baltica*.

GENE	Primer sense	Primer sequence
NADH5	Outer Forward	5'-CAAATGTTTCGTGGGATGGGARGGNGTNGGN-3'
	Outer Reverse	5'-AACGCCAGCAGTAACCATNGTNGCNGCRTG-3'
	Inner forward	5'- CAGGCTAATAARGCGGCNATHAARGCNATG-3'
	Inner reverse	5'-TACAGGAGTTGGTCCYTCCATNGCRTCNGG-3'
ATP9	Outer Forward	5'-GGAGCAGGCCTGGCNACNATHGGNMTN-3'
	Outer Reverse	5'-CATGAGCGCGAANAKNGCDATNGCYTC-3'
	Inner forward	5'-GCTGGTGCTGGNGTNGGNATHGGN-3'
	Inner reverse	5'- GGTGAGGGCRAANCCNAKDATNGC-3'

Phylogeny of mitochondrial proteins

The obtained DNA sequences were translated in and aligned with protein sequences of other organisms using CLUSTALW. Phylogenetic analysis was performed using PHYML version 2.4.4, using maximum likelihood analysis with a bootstrap value of 100. William and Goldman mode of substitution was used, with 4 gamma rate categories and a portion of invariable sites. Gamma parameter alpha and proportion of invariable sites were estimated from the data.

3.3 RESULTS AND DISCUSSION

The genes ATP synthase F0 subunit 9 and NADH dehydrogenase subunit 5 found in the mitochondrial genome of the endosymbiont of *Durinskia baltica*.

Fragments of the genes ATPase9 and NADH5 from the mitochondria of the endosymbiont of *Durinskia baltica* were successfully sequenced from 4 and 6 different clones respectively. NADH5 of the closely related *P. tricorutum* is ~670 amino acids long, whereas ATPase9 is 75 amino acids long; the fragments cloned from the respective genes of *D. baltica* were 77 and 52 amino acids long, representing only small, mid-portions of these genes. NADH5 was also amplified from whole RNA using Reverse-Transcription PCR, indicating that it is expressed.

Several lines of evidence suggest these markers are found in the mitochondrial genome of the endosymbiont. Phylogenetic analysis consistently placed the gene fragments with diatom mitochondrial genes. Low bootstrap values in both phylogenetic trees can be accounted for by the relatively short length of these gene fragments, which results in weaker support for their placement in the tree. Unfortunately dinoflagellate genes could not be included in the analysis as homologues of mitochondrial NADH5 and ATPase9 were not found in the NCBI, or PepDB databases (Fig 3.1, 3.2).

Figure 3.1 ATP synthase F0 subunit 9_maximum likelihood phylogeny (ML). Major lineages are delineated by square brackets and labeled on the right. Genes from all lineages included in the analysis were identified as ATPase9 in NCBI blast. Gene from *Durinskia baltica* shown in black box, grouping with diatoms.

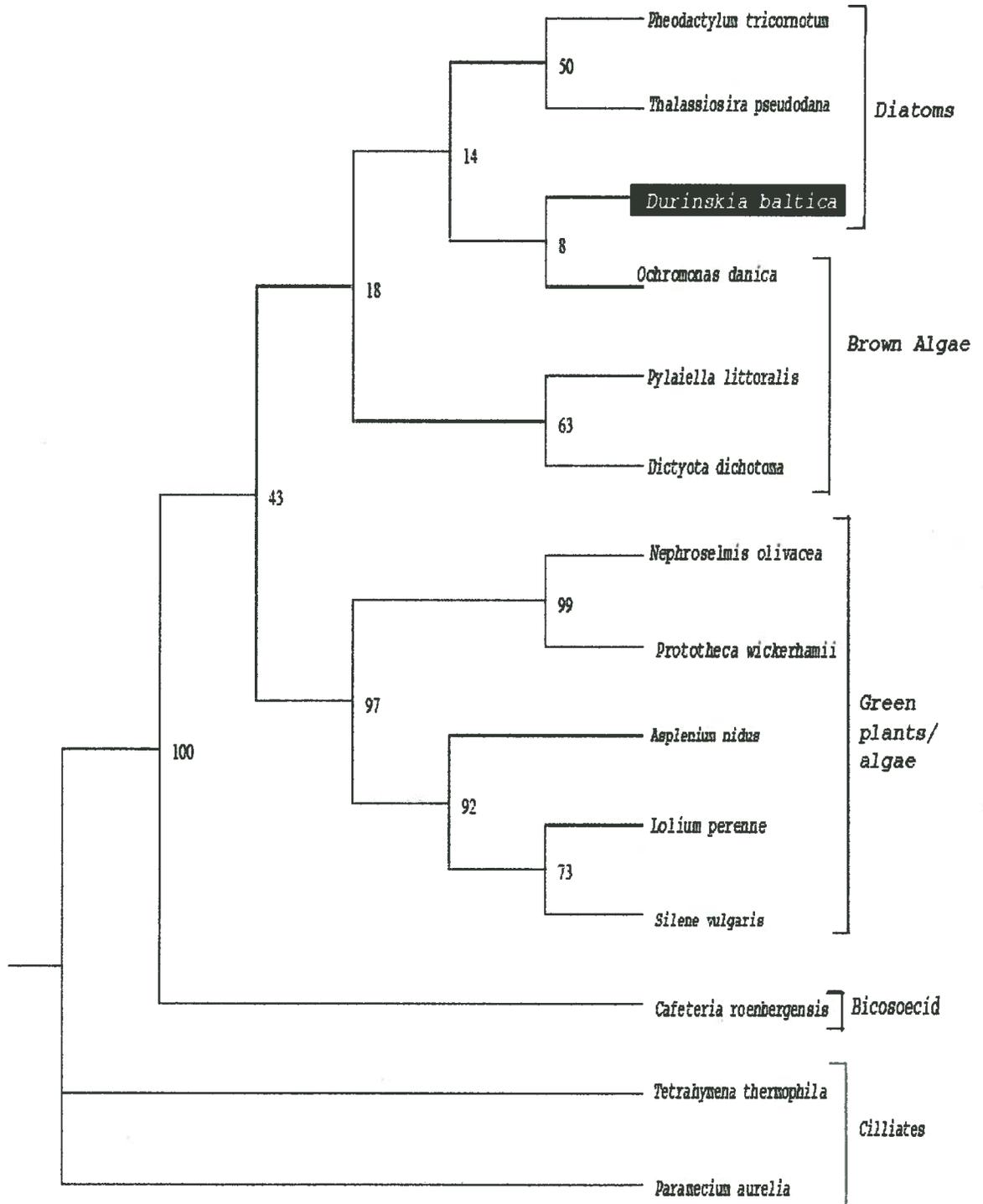
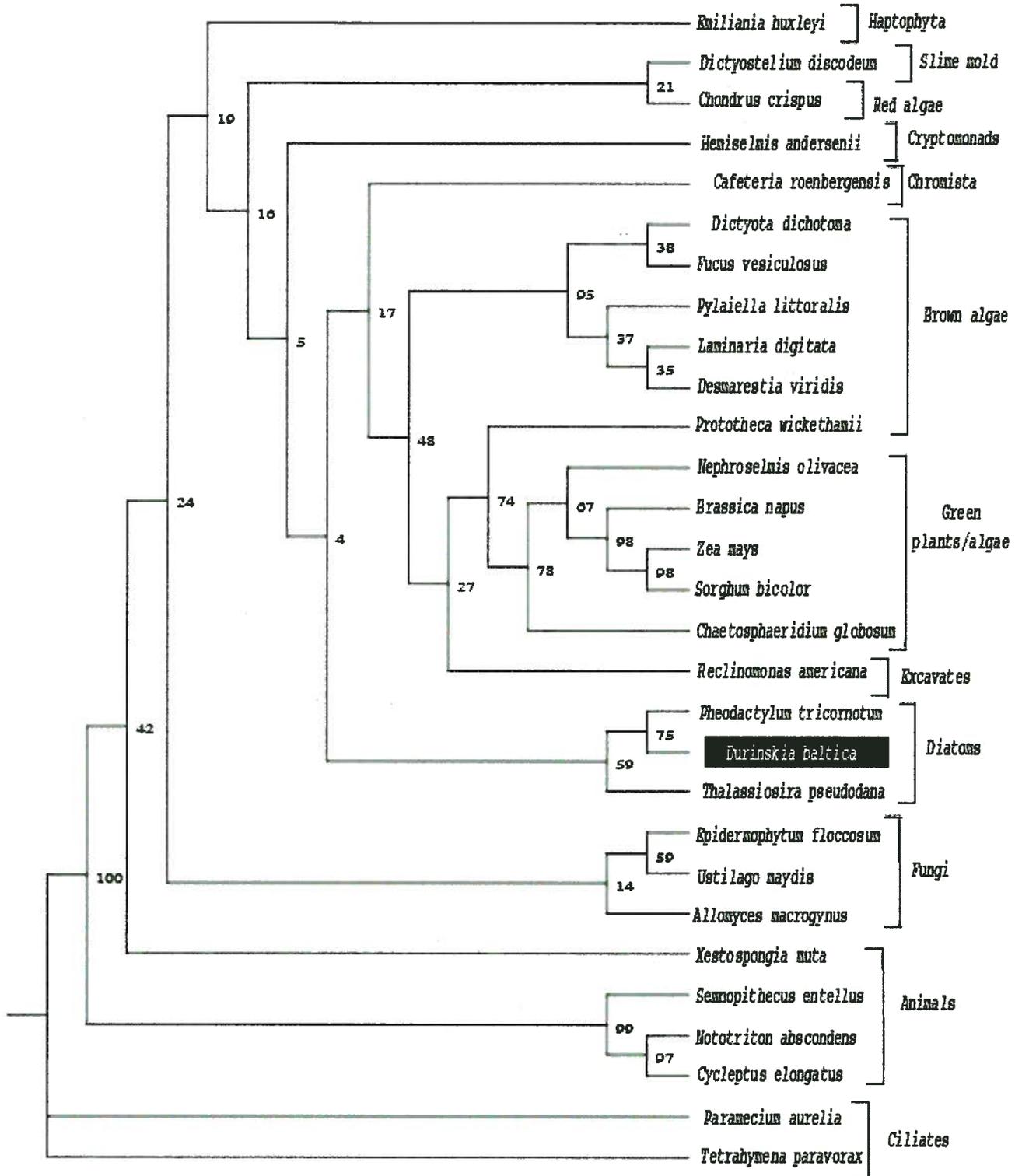


Figure 3.2 NADH dehydrogenase subunit 5 maximum likelihood phylogeny (ML). Major lineages are delineated by square brackets and labeled on the right. Genes from all lineages included in the analysis were identified as NADH5 in NCBI blast. Gene from *Durinskia baltica* shown in black box, grouping with diatoms.



Having identified the genes isolates as belonging to the diatom endosymbiont, we must consider whether they are in fact found in the mitochondrial genome or as many mitochondrial genes, the endosymbiont nucleus. Nuclear encoded mitochondrial genes have N-terminal transit peptides that target them back to the organelle as well as higher abundance of introns and higher GC content characteristic of the host nucleus (Gray et al. 1998). Unfortunately having only small fragments of these genes there is no telling if transit peptides are present therein, and though no introns were found in either NADH5 or ATP9 it doesn't exclude the possibility that they may be present in the remaining sequence. However A/T content in both markers is characteristic of the mitochondrial rather than nuclear genomes, with ATPase9 having 53.5% A/T and NADH5 having 69.7% A/T. Nuclear genes of both dinoflagellates and diatoms have A/T content of ~46-48%, (McEwan & Keeling 2004) significantly lower than the range for mitochondrial genomes which is about 60-80% A/T content (Gray et al 1998).

Studies in the early branching dinoflagellate *Oxyrrhis marina* indicate that dinoflagellate mitochondria only contains COX, COB and LSU genes, excluding the possibility that the markers above could have originated from the host mitochondrial genome. This fact, along with the phylogenetic analyses and sequence characteristics of the NADH5 and ATPase9 are consistent with the notion that these genes are found in the mitochondria of the endosymbiont. Although to confirm their location beyond any doubt the sequence of the entire mitochondrial genome and placement of these markers therein is required.

Functional and evolutionary implications for the presence of electron transport chain proteins in the endosymbiont mitochondria.

Mitochondrial energy production is mediated by membrane bound protein complexes I-V of the electron transport chain. Previous work in the lab identified COX1-3 as well as COB genes, (which are part of complexes III and IV) in the mitochondrion of the endosymbiont. NADH5 is part of complex I whereas ATPase9 is found in complex V. The presence of these genes in the endosymbiont mitochondrial genome, along with intact cristae, furthers the notion that this organelle is functional and participates in energy production.

Two scenarios can account for such functional conservation. Either the endosymbiosis with the diatom occurred very recently in evolutionary history and the diatom is in the process of degenerating or it is resisting degeneration, preserving autonomous energy production. One possible driving force that may prevent mitochondrial genome degeneration in this scenario is the distinct compartmentalization of the endosymbiont, namely its large volume, and the membrane separating it from the host, which may hinder shuttling of ATP from the host to the endosymbiont.

Much work is needed to fully understand the dynamics of genome degeneration in this non-canonical system. Understanding whether the diatom endosymbiont is currently degenerating or resisting the process would require more genetic data, and an assessment of how recently in evolutionary history the endosymbiotic event took place. This would require drawing comparisons among species with endosymbionts that were incorporated into their host at different times in history and comparing the degree of genomic and ultrastructural degeneration of endosymbionts among these species. Such an analysis

would provide a snapshot of different stages of endosymbiont integration, yielding unique insight into the dynamics of genome remodeling in endosymbiosis.

Dinoflagellates' promiscuity and relative ease of plastid uptake makes this group unique in offering an opportunity to study genome degeneration in endosymbionts of different origin across evolutionary time, providing scientists with universal understanding of the evolutionary underpinnings and mechanisms of genome degeneration.

LITERATURE CITED

- Andersson S.G. & Kurland C.G. (1998). Reductive evolution in resident genomes. *Trends in Microbiology* 6 (7): 263-268
- Archibald J.M. & Keeling P.J. (2002). Recycled plastids: a “green movement” in eukaryotic evolution. *Trends in Genetics* 18: 577-584
- Archibald, J.M. & Keeling, P. J. (2005). On the origin and evolution of plastids. In: Sapp J. (ed.), *Microbial Evolution: Concepts and Controversies* Oxford University Press, Oxford. pp. 238-260.
- Arisue N., Sanchez L.B., Weiss L.M., Muller M. & Hashimoto T. (2002). Mitochondrial-type hsp70 genes of the amitochondriate protists, *Giardia intestinalis*, *Entamoeba histolytica* and two microsporidians. *Protistology International* 51: 9-16.
- Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, Putnam NH, Zhou S, Allen AE, Apt KE, Bechner M, Brzezinski MA, Chaal BK, Chiovitti A, Davis AK, Demarest MS, Detter JC, Glavina T, Goodstein D, Hadi MZ, Hellsten U, Hildebrand M, Jenkins BD, Jurka J, Kapitonov VV, Kröger N, Lau WW, Lane TW, Larimer FW, Lippmeier JC, Lucas S, Medina M, Montsant A, Obornik M, Parker MS, Palenik B, Pazour GJ, Richardson PM, Rynearson TA, Saito MA, Schwartz DC, Thamatrakoln K, Valentin K, Vardi A, Wilkerson FP & Rokhsar DS. (2004). The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* 1;306(5693): 79-86.
- Biderre C., Pages M., Méténier G., David D., Bata J., Prensier G. & Vivarès C. P. (1994). On small genomes in eukaryotic organisms: molecular karyotypes of two microsporidian species (Protozoa) parasites of vertebrates. *Comptes Rendus del' Academie des Sciences Ser. III* 317: 399-404.
- Brossard N., Delage E., Littlejohn T.G., Plante I., Rioux P., Saint-Louis D., Zhu Y. & Burger G. (1998). Genome structure and gene content in protist mitochondrial DNAs. *Nucleic Acid Research* 15;26(4): 865-78.
- Bryan R.T. & Schwartz D.A. (1999). Epidemiology of microsporidiosis. In: Wittner M. & Weiss L.M. (ed). *The Microsporidia and Microsporidiosis*. American Society for Microbiology, Washington, D.C pp. 502–516.
- Burley S.K. & Roeder R.G. (1996). Biochemistry and structural biology of transcription factor IID (TFIID). *Annual Review Biochemistry* 65: 769-799.

- Cali A. & Tarkavian P.M. (1999). Microsporidian development, morphology and life cycles. In: Wittner M., Weiss L.M. (ed.) *The Microsporidia and Microsporidiosis*. American Society for Microbiology, Washington, D.C. pp: 85-128.
- Cavalier-Smith T. (1983) A six-kingdom classification and a unified phylogeny. In: Schenk H.E.A. & Schwemmler W.S. (ed.) *Endocytobiology II: Intracellular Space as Oligogenetic*. De Gruyter, Berlin, pp. 1027-34
- Chesnick J.M., Kooistra W.H.C., Wellbrock U. & Medlin L.K. (1997). Ribosomal RNA analysis indicates a benthic pennate diatom ancestry for the endosymbionts of the dinoflagellates *Peridinium foliaceum* and *Peridinium balticum*. *Journal of Eukaryotic Microbiology* 44: 314-320
- Chesnick J.M., Morden C.W. & Shmieg A.M. (1996). Identity of the endosymbiont of *Peridinium foliaceum*: analysis of rbcLS operon. *Journal of Phycology* 32: 850-857
- Corradi N., Gangaeva A. & Keeling P.J. (2008). Comparative profiling of overlapping transcription in the compacted genomes of microsporidia *Antonospora locustae* and *Encephalitozoon cuniculi*. *Genomics* 91: 388-393
- Curgy J., Vavra J. & Vivares C. P. (1980). Presence of ribosomal RNA's with prokaryotic properties in microsporidia, eukaryotic organisms. *Biologie Cellulaire* 38: 49-52.
- Cyanobacteria genome project
<http://www.people.vcu.edu/~elhaij/cyanonews/V16/GenomeProjects.html>
- Dale C., Wang B., Moran N. & Ochman H. (2003). Loss of DNA recombinational repair enzymes in the initial stages of genome degeneration. *Molecular Biology and Evolution* 20(8): 1188-1194
- Derelle, E. Ferraz C, Rombauts S, Rouzé P, Worden AZ, Robbens S, Partensky F, Degroeve S, Echeynié S, Cooke R, Saeys Y, Wuyts J, Jabbari K, Bowler C, Panaud O, Piégu B, Ball SG, Ral JP, Bouget FY, Piganeau G, De Baets B, Picard A, Delseny M, Demaille J, Van de Peer Y & Moreau H. (2006). Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. *Proceedings of the National Academy of Sciences of the United States of America*. 103: 11647-52.
- Dodge J.D. (1989). Phylogenetic relationships of dinoflagellates and their plastids. In: Green J.C., Leadbeater B.S.C. & Diver W.L. (ed.) *The Chromophyte algae; Problems and perspectives*. Clarendon Press, Oxford, pp: 207-227

- Dodge J.D. (1966). The Dinophyceae In: Godward M.B.E (ed.) The chromosomes of algae. St. Martin's Press, New York, USA. pp: 96-115
- Dujon B. (1996) The yeast genome project: what did we learn? *Trends in Genetics* 12(7): 263-270.
- Edlind T.D., Li J., Visvesvara G.S., Vodkin M.H., McLaughlin G.L. & Katiyar S.K. (1996) Phylogenetic analysis of β -tubulin sequences from amitochondriate protozoa. *Molecular Phylogenetics Evolution* 5: 359-67.
- Fast N.M. & Keeling P.J. (2001). Alpha and beta subunits of pyruvate dehydrogenase E1 from microsporidian *Nosema locustae*: mitochondrion derived carbon metabolism in microsporidia. *Molecular Biochemistry Parasitology* 117: 201-209.
- Fast N.M., Kissinger J.C., Roos D.S., Keeling P.J. (2001). Nuclear-Encoded, Plastid-Targeted Genes Suggest a Single Common Origin for Apicomplexan and Dinoflagellate Plastids. *Molecular Biology and Evolution* 18: 418-426
- Fensome R.A., Taylor F.J.R., Norris G., Sarjeant W.A.S., Wharton D.I. & Williams G.L. (1993). A classification of fossil and living Dinoflagellates. *Micropaleontology Press* 7: 351
- Fields, S.D. & Rhodes, R.G. (1991). Ingestion and retention of *Chromonas* spp. (Cryptophyceae) by *Gymnodinium acidotum* (Dinophyceae). *Journal of Phycology* 27: 525– 529.
- Franzen C. & Muller A. (2001). Microsporidiosis: human disease and diagnosis *Microbes and Infection* 3: 389-400
- Gast, R. J., Moran, D. M., Dennett, M. R. & Caron, D. A. (2007). Kleptoplasty in an Antarctic dinoflagellate caught in evolutionary transition? *Environmental Microbiology*. 9 (1): 39-45
- Gray, M.W. & Doolittle, W. F. (1982). Has the endosymbiont hypothesis been proven? *Microbiology Review* 46: 1–42
- Gray MW, Lang BF, Cedergren R, Golding GB, Lemieux C, Sankoff D, Turmel M, Brossard N, Delage E, Littlejohn TG, Plante I, Rioux P, Saint-Louis D, Zhu Y & Burger G. (1998) . Genome structure and gene content in protist mitochondrial DNAs. *Nucleic acid research* 15;26(4): 865-78.
- Gray M.W., Lang B.F. & Burger G. (2004). Mitochondria of Protists. *Annual review Genetics* 38: 477-524

- Germot A., Phillipe H. & Le Guyader H. (1997). Evidence for loss of mitochondria in microsporidia from mitochondrial-type HSP70 in *Nosema locustae*. *Molecular Biochemistry Parasitology*. 87: 159-168.
- Hackett J. D. L., Maranda H., Yoon S. & Bhattacharya D. (2003). Phylogenetic evidence for the cryptophyte origin of the plastid of *Dinophysis* (Dinophysiales, Dinophyceae). *Journal of Phycology* 39: 440-448
- Hazard EI. & Brookbank J.W. (1984). Karyogamy and meiosis in *Amblyospora spp.* in the mosquito *Culex salinarius*. *Journal of Invertebrate Pathology* 44(3): 9-11
- Hirt R.P., Logsdon J.M., Healy B., Dorey M.W., Doolittle F.W. & Embley M.T. (1999) Microsporidia are related to fungi: evidence from the largest subunit of RNA polymerase II and other proteins. *Protocol Of National Academy of Science* 96: 580-585
- Imanian B., Carpenter K. J., & Keeling P.J. (2007). Mitochondrial genome of a tertiary endosymbiont retains genes for electron transport proteins. *Journal of Eukaryotic Microbiology* 54(2): 146-153
- Imanian B. & Keeling P.J. (2007). The Dinoflagellates *Durinskia baltica* and *Kryptoperidinium foliaceum* retain functionally overlapping mitochondria from two evolutionarily distinct lineages *BMC Evolutionary Biology* 7: 72-183
- Jeffrey S.W. (1989). Chlorophyll c pigments and their distribution in the chromophyte algae. In: Green J.C., Leadbeater B.S.C. & Diver W.L. (ed.) *The Chromophyte algae; Problems and perspectives*. Clarendon Press, Oxford, pp 13-36
- Katinka MD., Duprat S., Cornillot E., Meternier G., Thomarat F., Prensler G., Barbe V., Peyretallade E., Brottler P., Wincker P., Delbac F., Alaoul E.H., Peyret P., Saurin W., Gouy M., Weissenbach J. & Vivares C.P. (2001). Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Letters to Nature* 414(6862): 450-3.
- Keeling P.J. (2003). Congruent evidence from alpha-tubulin and beta-tubulin gene phylogenies for a zygomycete origin of microsporidia. *Fungal Genetics Biology* 38: 298-309.
- Keeling P.J. & Doolittle W.F. (1996). Alpha-tubulin from early diverging eukaryotic lineages and the evolution of the tubulin family. *Molecular Biology Evolution* 13: 1297-1305
- Keeling P.J. & Fast N.M. (2002). Microsporidia: Biology and evolution of highly reduced intracellular parasites. *Annual Review Microbiology* 56: 93-116.

- Kudo R.R. (1947). A biologic and taxonomic study of the microsporidia. In: Thomas (ed.) *Protozoology*. Springfield, Illinois pp 778
- Larsson J.I. R., (1999). Identification of microsporidia. *Acta Protozoologica* 38: 161-197. McEwan M. & Keeling P.J. (2004). HSP90, Tubulin and actin are retained in tertiary endosymbiont of *Kryptoperidinium foliaceum*. *Journal of Eukaryotic Microbiology* 51(6): 651-659
- McFadden G.I. (2001). Primary and secondary endosymbiosis and the origin of plastids. *Journal of Phycology*. 37: 951-959
- Metenier G. & Vivares C.P. (2001). Molecular characteristics and physiology of microsporidia. *Microbes and Infection* 3(5): 407-415.
- M&R Durango, Inc. Copyright © 1999 All rights reserved. Revised: February 07, 2006
M&R Durango Inc., Insectary P.O. BOX 886 Bayfield, CO 81122
- Netzel, H. & Durr, G. (1984). Dinoflagellate cell cortex. In: Spector D.L (ed.) *Dinoflagellates*. Academic Press, Orlando, Florida pp 43-105
- Patron N.J., Waller R.F. & Keeling P.J. (2006). A tertiary plastid uses genes from two endosymbionts. *Journal of Molecular Biology* 357(5): 1373-1382
- Pesole G., Mignone F., Gissi C., Grillo G., Lucciulli F. & Liuni S. (2001). Structural and functional of eukaryotic mRNA untranslated regions. *Gene* 276: 73-81
- Peyretailade E, Brousoole V, Peyret P, Metenier G, Gouy M & Vivares CP (1998). 70-kDa heat shock protein gene of mitochondrial evolutionary origin. *Molecular Biology Evolution* 15: 683-689.
- Sato S., Nakamura Y., Kaneko T., Asamizu S. & Tabata S. (1999). Complete structure of the chloroplast genome of *Arabidopsis thaliana*. *DNA Research* 6(5): 283-290
- Schnepf E. & Elbrächter M. (1988). Cryptophycean-like double membrane-bound chloroplast in the Dinoflagellate, *Dinophysis ehrenb*: evolutionary, phylogenetic and toxicological implications. *Botanica Acta* 101:196-203.
- Schnepf, E. & Elbrächter, M. (1992). Nutritional strategies in Dinoflagellates. *European Journal of Protistology* 28: 3-24.
- Skovgaard A. (1998). Role of chloroplast retention in a marine Dinoflagellate. *Aquatic Microbiological Ecology* 15: 293-301.
- Slamovits C.H. & Keeling P.J. (2004). Class II photolyase in a microsporidian intracellular parasite. *Journal of Molecular Biology* 341: 713-721

- Slamovits C.H., Saldarriaga J.F., Laroque A. & Keeling P.J. (2007). A highly reduced and fragmented mitochondrial genome of the early-branching dinoflagellate *Oxyrrhis marina* shares characteristics with both apicomplexan and dinoflagellate mitochondrial genomes. *Journal of Molecular Biology* 372: 356-368.
- Slamovits C. H. & Keeling P. J. (2008). Plastid-derived genes in the non-photosynthetic alveolate *Oxyrrhis marina*. *Molecular Biology Evolution* 25: 1297-1306.
- Smale S.T. & Kadonaga J.T. (2003). The RNA polymerase II core promoters. *Annual review Biochemistry* 72:449-479
- Sogin M.L., Morrison H.G., and Vossbrinck. C. *Antonosporea Genome Database* Connecticut Agricultural Experiment Station, New Haven
<http://gmod.mbl.edu/perl/site/antonosporea01?page=intro>
- Spector D.L. (1984). Dinoflagellate Nuclei In: Spector D.L. (ed.) *Dinoflagellates*. Academic Press Inc, Orlando, Florida, USA pp: 107-147
- Steidinger, K.A. & Baden D.G. (1984). Toxic Marine Dinoflagellates. In: Spector D.L. (ed.) *Dinoflagellates*. Academic Press, Inc., New York pp: 201-262.
- Taylor F.J.R. (1987). *The biology of dinoflagellates*. Oxford, Blackwell Scientific.
- Tengs T. O. J. Dahlberg K. Shalchian-Tabrizi D. Klaveness K. Rudi C. F. & Delwiche K. S. Jakobsen (2000) Phylogenetic analyses indicate that the 19'hexanoyloxy-fucoxanthin-containing dinoflagellates have tertiary plastids of haptophyte origin. *Molecular Biology and Evolution* 17: 718-729
- Timmis J.N., Ayliffe M.A., Huang C.Y., & Martin W. (2004). Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nature Reviews Genetics* 5: 123-135
- Tomas R.N. & Cox E. R. (1973). Observations on the symbiosis of *Peridinium balticum* and its intracellular alga. *Journal of Phycology* (9):304-323
- Van de Peer Y., Ben Ali A. & Meyer A. (2000). Microsporidia: accumulating molecular evidence that a group of amitochondriate and suspectedly primitive eukaryotes are just curious fungi. 246:1-8.
- Vavra J. & Larsson J.I. R (1999). Structure of microsporidia. In; Wittner M. & Weiss L.M. (ed.) *The Microsporidia and Microsporidiosis*. American Society for Microbiology, Washington, D.C. pp: 7-84.
- Vivarès C. P., Gouy M., Thomarat F. & Méténier G (2002). Functional and evolutionary analysis of a eukaryotic parasitic genome. *Current Opinion in Microbiology* 5: 499-505

- Vossbrinck C.R., Maddox J.V., Friedman S., Debrunner-Vossbrinck B.A. & Woese C.R. (1987). Ribosomal RNA sequence suggests microsporidia are extremely ancient eukaryotes. *Nature* 326: 411-414.
- Watanabe M. M. Y., Suda S., Inouye I., Sawaguchi M. & Chihara (1990). *Lepidodinium viride* (Gymnodiniales, Dinophyta), a green dinoflagellate with a chlorophyll *a*- and *b*-containing endosymbiont. *Journal of Phycology* 26: 741-751
- Watanabe M. M. Y. Takeda T. Sasa S. Suda S. Sawaguchi M. & Chihara (1987). A green dinoflagellate with chlorophylls *a* and *b*: morphology, fine structure of the chloroplast and chlorophyll composition. *Journal of Phycology* 23: 382-389
- Weidner E., Findley AM., Dolgikh V & Sokolova J. (1999). Microsporidian biochemistry and physiology. In: Wittner M. & Weiss L.M. (ed.) *The Microsporidia and Microsporidiosis*. American Society for Microbiology, Washington, D.C pp 85-128
- Weinzierl R. & Henn T. (1989). Alternatives in insect management: Microbial insecticides. Cooperative Extension, University of Illinois, Circular 1295. 12 pp.
- Williams B.A.P., Hirt RP, Lucocq JM. & Embley TM. (2002). A mitochondrial remnant in the microsporidian *Trachipleistophora hominis*. *Nature* 418: 865-869.
- Williams B.A.P., Lee R.C.H., Becnel J.J., Weiss L.M., Fast N.M. & Keeling P.J. (2008). Genome sequence surveys of *Brachiola algerae* and *Edhazardia aedis* reveal microsporidia with low gene densities. *BMC Genomics* 9: 200-208