ORGANELLE GENOMES OF LICHENS

BASIL BRITTO XAVIER

Faculty of Life and Environmental Sciences
University of Iceland

ORGANELLE GENOMES OF LICHENS

BASIL BRITTO XAVIER

90 ECTS thesis submitted in partial fulfillment of a
Master of Science degree in Biology

Advisors
Prof. Ólafur S. Andrésson
Dr. Vivian P.W. Miao

Faculty Representatives
Dr. Zophonías O. Jónsson
Dr. Albert V. Smith

Examiner
Prof. Kristinn P. Magnússon

Faculty of Life and Environmental Sciences
School of Engineering and Natural Sciences (SENS)
University of Iceland
Reykjavik, December 2011

ORGANELLE GENOMES OF LICHENS
90 ECTS thesis submitted in partial fulfillment of a *Master of Science* degree in Biology

Faculty of Life and Environmental Sciences
School of Engineering and Natural Sciences
University of Iceland
Askja, Sturlugata 7
107, Reykjavik
Iceland

Telephone: 525 4000

Bibliographic information:
Basil Britto Xavier, 2011, ORGANELLE GENOMES OF LICHENS,
Master's thesis, Faculty of Life and Environmental Sciences, University of Iceland, pp.96

Printing: Haskolaprent hf
Reykjavik, Iceland, December 2011

Abstract

The organelle genomes of three lichen species (*P. membranacea*, *P. malacea* and *C. grayi*; photobiont *Asterochloris* sp) were sequenced, annotated and comparated to its closest non-lichenized counterparts. These organisms are of interest as they serve as models for studying symbiosis. They are difficult to culture these in the laboratory. The mycobiont mitochondrial genomes are approximately 50-60 kb in size, circular molecules containing genes coding for 14 proteins along with one ribosomal protein (*rps3*), two rRNA genes and a set of 26 tRNA genes. In addition the *Peltigera* mtDNAs code for plasmid origin DNA polymerase (*dpoB*) type B2 and *rnpB*, encoding the RNA component of RNaseP. Mitochondrial sequences of *Peltigera* spp. and *C. grayi* are the first reported from *Lecanoromycetes*. The phylogenetic tree of concatenated mitochondrial proteins from the lichenized fungi and other ascomycetes exhibits a pattern significantly different from those derived from nuclear proteins.

The *Asterochloris* sp mitochondrial and chloroplast genomes are amongst the largest organelle genomes in green algae. The mitochondrial genome of *Asterochloris* sp is of the conventional type, it contains a canonical arrangement of a single large rRNA gene along with a 5S rRNA gene, encodes almost a almost full set of tRNAs and has remanants of Group II introns. The chloroplast genome of *Asterochloris* sp is lacking *cysA*, encoding a generally onserved transport protein and its *tilS* (tRNA-Ile lysidine synthase) gene is split, unlike in its chloroplast relatives. *Asterochloris* sp is the first species to be sequenced from a lichen photobiont of the order *Microthamniales* belonging to the class *Trebouxiophyceae.* These organelle genomes will serve as tools to study the function, and evolution of these lichens and its photobionts and the evolutionary cycle of their associated Group I and II introns.

Útdráttur

Erfðamengi frumulíffæra úr þremur tegundum flétta (*Peltigera membranacea*, *Peltigera malacea og Cladonia grayi*) voru raðgreind og gen þeirra skilgreind. Erfðamengin voru borin saman við nánustu ættingja í öðrum lífveruhópum en fléttum. Þessi erfðamengi eru áhugaverð þar sem þau eru í sambýlislífverum. Erfðamengi hvatberanna í svepphlutum fléttanna eru 50-60 kb að stærð, hringlaga með 14 próteingen og 2 rRNA gen eins og dæmigert er í asksveppum auk eins ríbósómpróteingens (*rps3*) og 26 tRNA gena. *Peltigera* hvatberalitningar hafa einnig táknraðir fyrir DNA pólýmerasa B og RNA þátt RNAsa P. Basaraðir þessara hvatberalitninga eru þær fyrstu sem greindar hafa verið í flokknum *Lecanoromycetes.* Skyldleikatré hvatberapróteinanna er marktækt frábrugðið ættartré sem fæst með kjarnapróteinum sömu tegunda.

Grænþörungur af ættkvíslinni *Asterochloris* er ljósbýlingur fléttunnar *C. grayi*. Bæði hvatberalitningur og grænukornslitningur ljósbýlungsins er meðal þeirra stærstu sem þekktir eru meðal grænþörunga. Hvatberalitningurinn er “hefðbundinn”, með eitt rRNA gen auk 5S rRNA gens, með táknraðir fyrir næstum fullt sett af tRNA sameindum og er með leifar af innröðum af flokki II. Sérkenni grænukornalitningsins er fjarvera *cysA* gens (sem táknar flutningsprótein), og tvískipt *tilS* gen (táknar tRNA-Ile lýsidín synþasa).

Þessi *Asterochloris* tegund er hin fyrsta úr ljósbýlungi af ættinni *Microthamniales* (í flokknum *Trebouxiophycea*) þar sem bæði hvatbera- og grænukornslitningar hafa verið raðgreindir.

Basaraðir þessara litninga úr frumulíffærum flétta munu verða nýtt við rannsóknir á starfi, þróun og skyldleika þessara flétta ásamt innraða af flokkum I og II sem þeim fylgj

Dedication

This thesis is dedicated to my family

Basil Britto

TABLE OF CONTENTS

Acknowledgements

This research project would not have been possible without the support of many people. I would like to extend my deepest gratitude especially to the following persons.

First and foremost I wish to thank my supervisor Professor Ólafur S. Andrésson whose patience, kindness, wisdom, commitment as well as his academic excellence have been invaluable to me. He has supported me by not only providing me with this research opportunity but also guided me through the rough roads to finish this thesis.

I gratefully acknowledge my co-supervisor Dr. Vivian P.W. Miao from University of British Columbia, Canada who let me experience lichens in the fields, taught me the state of the art of scientific writing, patiently corrected my writings over and over towards the contribution of the article on *Peltigera* mt genomes.

Deepest gratitude to the members of the supervisory committee Dr. Zophonías O. Jónsson and Dr. Albert V. Smith without whose assistance this would have not been successful. My special thanks to Professor Gudmundur H Gudmundsson for offering me valuable advice and for the encouragement for improving my presentation skills. I would also like to give my appreciation to our collaborator Dr. Daniele Armaleo, Duke University, U.S.A for allowing us to utilize their *Cladonia grayi* genomic data.

I would like to particularly acknowledge the help of Sara Sigurbjörnsdóttir in to getting familiar with the basic lab codes during my initial days and I also wish to thank the members of Lab 387 (ASKJA) at the Faculty of Life and Environmental Sciences, University of Iceland.

This thesis and research would not be realistic without funding, so my thanks to The Icelandic Research Fund for funding this project. Finally, my deep gratefulness goes to my family for their unconditional love and support throughout my life.

Last but not least, I thank the Almighty for the wisdom and perseverance that he has bestowed upon me during this research project, and indeed, throughout my life through all tests in the past three and half years. You have made my life more bountiful. May your name be exalted, honoured, and glorified.

General Summary

The complete mtDNA sequences of the lichens *Peltigera membranacea, Peltigera malacea* and *Cladonia grayi* and the organelle genomes of its photobiont partner *Asterochloris* sp. were generated and compared to those of non-lichenized counterparts. The annotations of organelle genomes were validated using transcriptomic data, and transcriptionally active non-coding regions (TARs) as well as group I and group II intron splice sites were defined.

The lichen mt genomes obtained were the first to be sequenced from *Lecanoromycetes,* the largest class of lichenized fungi, a part of the *Pezizomycotina,* the largest subphylum of Ascomycota. The mt genomes of *Peltigera* spp. and *C. grayi* have similar gene orders, conserved repeat regions, which are different from those of non-lichenized fungi. The unique feature of the *C. grayi* mt genome is the arrangement of the *rps3* gene and the fusion of a group IA intron with the 5' end of the gene. The gene *dpoB,* proposed to be of plasmid origin, and the *rnpB* gene encoding the RNAcomponent of RNAe P were identified in the *Peltigera* spp. but were absent from the *C. grayi* mt genome.

The primary photobiont partner of *C. grayi* is a unicellular green algae, *Asterochloris* sp. It has the largest mt genome and the third largest cp genome known from green algae. These are the first sequenced cp genomes from the order *Microthamniales* of the class *Trebouxiophyceae.* The *tilS* gene of the *Asterochloris* cp genome is split into two parts and is present in two different places in the cp genome. The genes *psbD* and *psbC* are fused and the generally conserved transport protein coding gene *cysA* is absent from the cp genome. The only intron (group IA) in the *Asterochloris* sp cp genome is the RNA gene *rrn23.*

CHAPTER I

1.1.GENERAL INTRODUCTION

Lichens are symbiotic associations of fungi with photosynthetic green algae (e.g. *Asterochloris*) and/or cyanobacteria (*Nostoc*), are found in terrestrial environments and form their own micro-ecosystems. Lichens have been classified as chlorolichens and cyanolichens based on their primary photobiont partners. Almost 20% of all known fungal species are lichenized, and among lichenized species ~ 42% of are to ascomycetes and nearly 1500 species have *Nostoc* as its photobiont. There are over 14,000 lichen species identified (Nash, 2008; Rikkinen et al . 2002). *Lecanoromycetes* represent the predominant lichenized fungal order but these fungi not well studied because they are often recalcitrant to culture. Some of the symbionts (mycobionts) have been cultivated (e.g. *Cladonia grayi*) but others not (e.g. *Peltigera*).

Peltigeras are foliose lichens and have been studied for many years for their morphology, taxonomy, ecology and biogeographical photobiont partner variations (O'Brien et al . 2005). *C. grayi* is a stalked-cup fruticose lichen and has been widely studied for its antibiotic properties and secondary metabolites. The primary photobiont partner of the *Peltigera* spp. is *Nostoc* and that of *C. grayi* it is an *Asterochloris* sp., a unicellular green algae.

Eukaryotic organelle genomes are clearly prokaryotic remnants (Timmis, 2004), reafirming the theory of endosymbiosis, a focal point in studies of the organelle evolution. The organelle genomes of different lineages (plant and animal) evolved in their own unique directions. Organelle genomes such as mitochondria (proteobacteria) and chloroplasts (cyanobacteria) originated from eubacterial cells and became a part of the host cell (Andersson, 2003; Gray et al . 1999; Lynch et al . 2006).

In this genomic era, sequencing of organelle genomes is central theme because organelle genomes play a key role in many studies of evolution and gene transfers in eukaryotes (Kleine et al . 2009), and are also manageable technically. The unique characteristics of mitochondrial uniparental inheritance, lack of genetic recombination and an independent rate

of evolution makes them suitable for many systematic studies (Hausner, 2003; Lang et al . 1997).

The organelle genomes of the five major eukaryotic lineages are quite similar. The main differences in these genomes are between photosynthetic plants and green algae on one hand, and non-photosynthetic animals, fungi and amoebozoa lineages on the other. During evolution of organelle genomes, genes were lost or transferred to the nucleus (Adams, 2000; Brouard, 2010), with most organelle proteins now encoded by nuclear genes and only imported into organelles through translocase complexes (Esaki, 2004). Organelle genomes have primarily retained genes required for their own function and reproduction, and green algal chloroplasts have retained fewer genes than red algal chloroplasts (Green, 2011). Gene transfers are an ongoing active process in plants and algae; algal organelle genomes are one of the available sources for studying these events (Adams & Palmer, 2003).

In general mt genomes and cp genomes vary widely in size, ranging from approximately 5.9 kb (*Plasmodium falciparum* mt genome) to 521 kb (*Floydiella terrestris* cp genome) (Brouard, 2010) (GenBank). This large difference is due to a varying number of genes, the size of large intergenic regions, numbers and sizes of repeats and the invasion of mobile group I and II introns often containing homing endonuclease genes (GIY YIG and LAGLIDADG types) as well as reverse transcriptase. Integrated into protein coding genes and ribosomal RNA genes (Green, 2011; Hausner, 2003). The major differences between mitochondrial and chloroplast genomes are the use of alternative genetic codes in mt (Green, 2011; Griffiths, 1996; Paquin et al . 1997), fewer ribosomal protein genes (Gray et al . 1999) and encoded tRNA genes. The unicellular algal cp genomes do not encode NADH dehydrogenase components and they are only found in higher plant cp genomes.

Organelle genomes are considered to be the best reservoir for self splicing mobile introns (group I and II) (Lambowitz et al . 1999). Group I and II intron types are lineage specific i.e. group I introns are fungal specific and group II introns which are thought to be derived from splicosomal introns are specific to plant and algal mt and cp genomes. The intron types differ in their secondary structure, consensus sequences at their splice sites, and type of reaction involved during splicing (Nedelcu & Lee, 1998). They are capable of homing into intron-less populations as well as transposing to other locations even in different evolutionary lineages (horizontal transfer) (Lambowitz & Belfort, 1993). The events associated with the homing mechanism are different and involve intron-encoded site specific endonucleases

(LAGLIDADG and GIY-YIG) for the group I introns and intron-encoded reverse-transcriptase-like proteins for the group II introns (Nedelcu et al . 2000).

At present, 2937 organelle genomes have been completely sequenced and deposited in GenBank. Among those 2668 are from mitochondria, 89 from fungi and 15 from chlorophyta; 213 are from chloroplast including 22 from chlorophyta. At present, there are no organelle genomes from lichens. (*P. membranacea* (JN088165), *P. malacea* (JN088164) and *C. grayi*).

In order to better understand the biology and genetics of lichen organelle genomes we used a culture-independent *in-situ* (metagenomic) approach towards the whole genome sequencing of *P. membranacea* and *P. malacea;* and a culture dependent approach for *C. grayi* based on a single spore culture of the mycobiont and an axenic culture of *Asterochloris sp*. In this study, we present the complete organelle genomes from these lichenized fungi as well as the organelle genomes of the lichen photobiont *Asterochloris* sp.

Statement: My advisor and co-advisor have contributed in the scientific writing of the attached manuscript (Appendix) and especially towards *dpoB*, *tilS* and tRNA-Ile analysis.

CHAPTER II

2.1.GENERAL MATERIALS AND METHODS

Genomic DNA and total RNA extraction from *P. membranacea* and genomic DNA from *P. malacea* were performed as described (Xavier B.B et al, manuscript submitted). The single spore culture of the mycobiont of the lichen *C. grayi* is and the clonal culture of its primary photobiont partner *Asterochloris* sp. originated at Duke University (D. Armaleo lab).

C. grayi and *Asterochloris* sp genomic DNA was isolated from cultures and was used for 454 and Illumina sequencing. Similarly, RNA was isolated from cultures at separate reconstitution time points as well as from *in-situ C. grayi* and were used for 454 and Illumina RNA sequencing (personal communication Daniele Armaleo).

The sequences of the *C. grayi* mt genome and the *Asterochloris* sp mt and cp genomes were derived from an ongoing genome project on the lichen *C. grayi* (Armaleo et al, unpublished). The mt and cp genomes of *Asterochloris* sp were represented by individual contigs comprising 85,112 reads with 308-fold coverage and 215,063 reads with 440 fold coverage of the genome respectively. The individual contigs were generated from 2,491,461 reads totaling 0.96 gigabases from 454 Titanium pyrosequencing of the cultured lichen photobiont *Asterochloris* sp. Similarly, the *C. grayi* mt contig was comprised of ~11,000 reads.

2.2. Genome assembly and annotation

Whole genome pyro-sequenced reads were sorted into groups, based on closely related published reference genomes (*Aspergillus niger* and *Pencillium marneffei* mt genomes for mycobiont; *Coccomyxa* and *Micromonas* mt and cp genomes for *Asterochloris* sp). For generating the mycobiont mt genome, sorted reads were assembled using the MIRA assembler (Chevreux, 1999). The assembled mycobiont mt genomes were used for reference and the whole genome reads were mapped in order to verify the genome coverage and smooth assembly. For the photobiont (*Asterochloris* sp) organelle genomes (cp and mt), de novo assembly was performed on the whole 454 reads using the MIRA genome assembler. Semi-automatic annotation of the assembled organelle genomes were carried out using CLC

Genomics Workbench and annotation was verified manually using RNA sequencing data. Illumina reads were mapped to resolve 454 homopolymer errors.

Gene content of the mitochondria was determined by manual annotation using the CLC Genomics Workbench and was compared to fungal mt sequences by blastX searches at NCBI (Altschul, Gish et al. 1990). tRNAs, group I introns, *rns*, 5S *rRNA* and *rnpB* were predicted using RNAweasel . Transcriptionally active regions (TARs) in *P. membranacea* were identified using mRNA sequences generated from different tissues collected at different time points. The mRNA reads were mapped to the *P. membranacea* mt genome using CLC Genomics Workbench. Identification of TARs was based on the following parameters: a minimum coverage of at least four reads/nucleotide for a stretch of at least ~150 nucleotides (nt) and at least ~150 nt away from an annotated gene (Bruno, Wang et al. 2010). Similarily, mapping of mRNA reads on *C. grayi* and *Asterochloris* sp organelle genomes was done to verify splice sites.

2.3. Refining genome assembly

Ambiguity and low quality assembly regions in the genome were resolved by PCR and Sanger sequencing. Primers were designed for the ambigious regions using CLC Genomics Workbench. These regions were amplified and the PCR amplicons were sequenced in both directions using the BigDye Terminator Cycle Sequencing kit (PerkinElmer-Applied Biosystems, Inc., Foster City, CA, USA) and an automated sequencer (ABI 3100; PerkinElmer-Applied Biosystems, Inc.) (data not shown).

CHAPTER III

3.1. General features of the *C. grayi* mt genome

The mt genome of *C. grayi* is 50,836 bp and is circular. The genome size, low G+C content (~29%), gene numbers and open reading frames (ORF) were within the range commonly observed in other fungal mt genomes (Table 1). The *C. grayi* mt genome has 15 predicted ORFs for known protein coding genes: three for ATP synthase F_0 subunits (encoded by *atp6*, *atp8* and *atp9*), three for cytochrome c oxidase subunits (*cox1*, *cox2* and *cox3*), seven for nicotinamide adenine dinucleotide ubiquinone oxidoreductase subunits (*nad1, nad2, nad3, nad4, nad4L, nad5, nad6*), one for cytochrome oxidase b (*cytb*) and ribosomal protein subunit 3 (*rps3*). The genome consists of 42% coding, 13% of intronic and 45% intergenic (non-coding) regions (Fig. 1). The *C. grayi* mt genome has a full set of tRNA genes, as well as genes for the large (*rnl*) and small (*rns*) ribosomal subunit RNAs and also an ORF with no significant match in GenBank (Table 1). Since the genomic DNA was isolated from a single spore culture, no single nucleotide polymorphism (SNP) was detected.

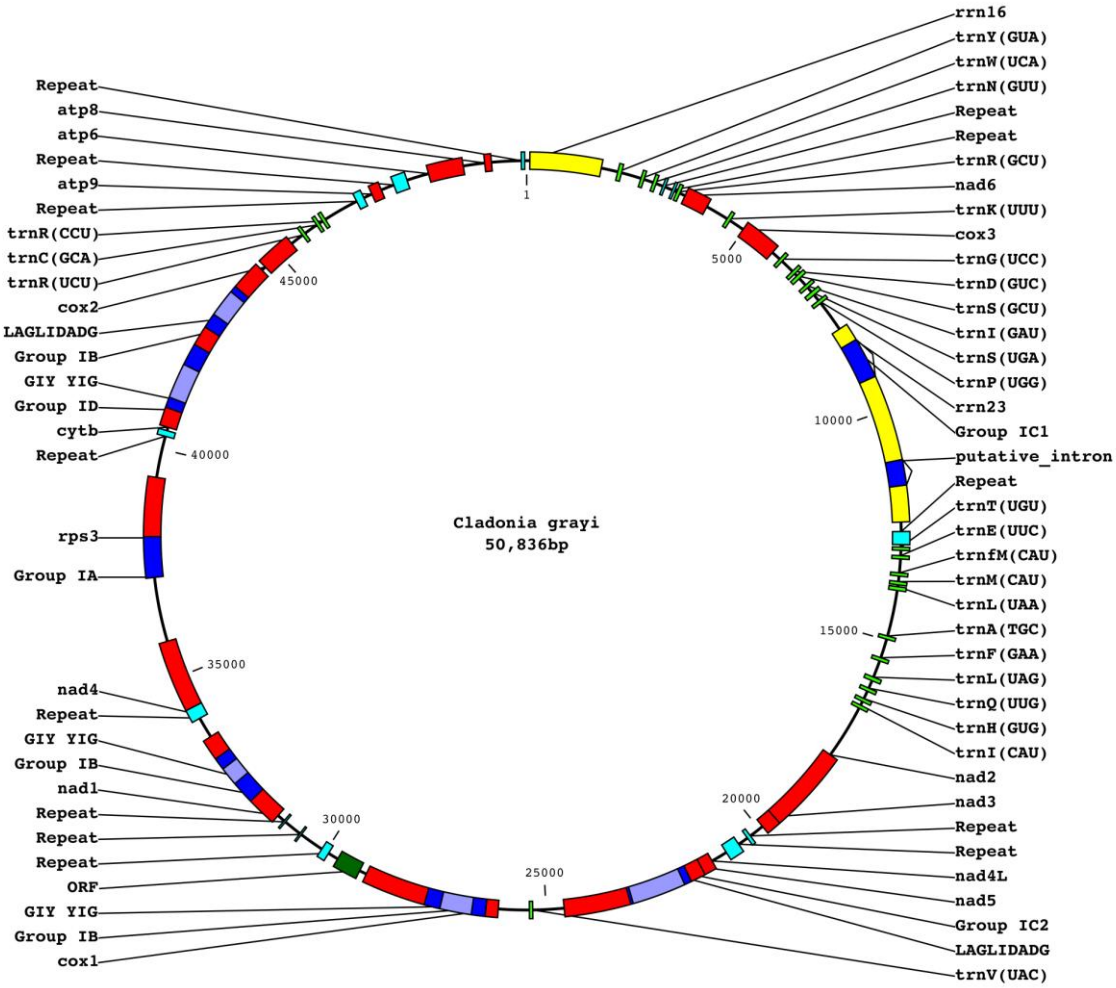


Figure 1. *C. grayi* mitochondrial genome. Circular gene map representing the mt DNA of *C. grayi*. All genes are transcribed clockwise. Genes encoding proteins, ribosomal RNA and repeats are indicated by red, yellow and cyan boxes, respectively; tRNA genes, in green, are specified by standard one-letter amino acid code and anticodon. Introns and homing endonuclease genes (HEG) are shown in blue and light blue, and an unidentified ORF is in dark green.

3.2. Introns and Repeats

There are eight group I introns and five encode intact homing endonucleases (3 GIY-YIG and 2 LAGLIDADG). The *nad1* gene is interrupted by a group IB intron with an intact GIY YIG homing endonuclease ORF. This intron insertion position in *nad1* gene is not seen in the *Peltigera* spp. Moreover, the *nad4L* is not interrupted by introns, whereas the *Peltigera nad4L* is interrupted by introns. The differences in intron numbers and intron positions in protein coding genes vary among these closely related lichen mt genomes. The conserved *cox1* gene is interrupted by a single group IB intron. The intron position class is "D" (FMV * MPA) (Férandon et al . 2010) as in *P. membranacea* and *P. malacea.* This group IB intron has an intact ORF encoding a 225 aa long GIY-YIG homing enodonuclease.

Simple sequence repeats are often called microsatellites. These repeats play an important role in genome evolution, recombination and genetic variation between and within species (Kashi, 2006). Seven simple sequence repeats were found and they varied in sizes from 38-321 bp. Among these repeats a 282 nt long repeat was found very close to the 3' end of the *rnl* gene and nearly the same (87% identity) repeat sequence was also found between the *nad1* and *nad4* genes. The 3' 193 nucleotide of this repeat region overlapped the 5' of the *nad4* gene (Fig 1) (AppendixTable 5).

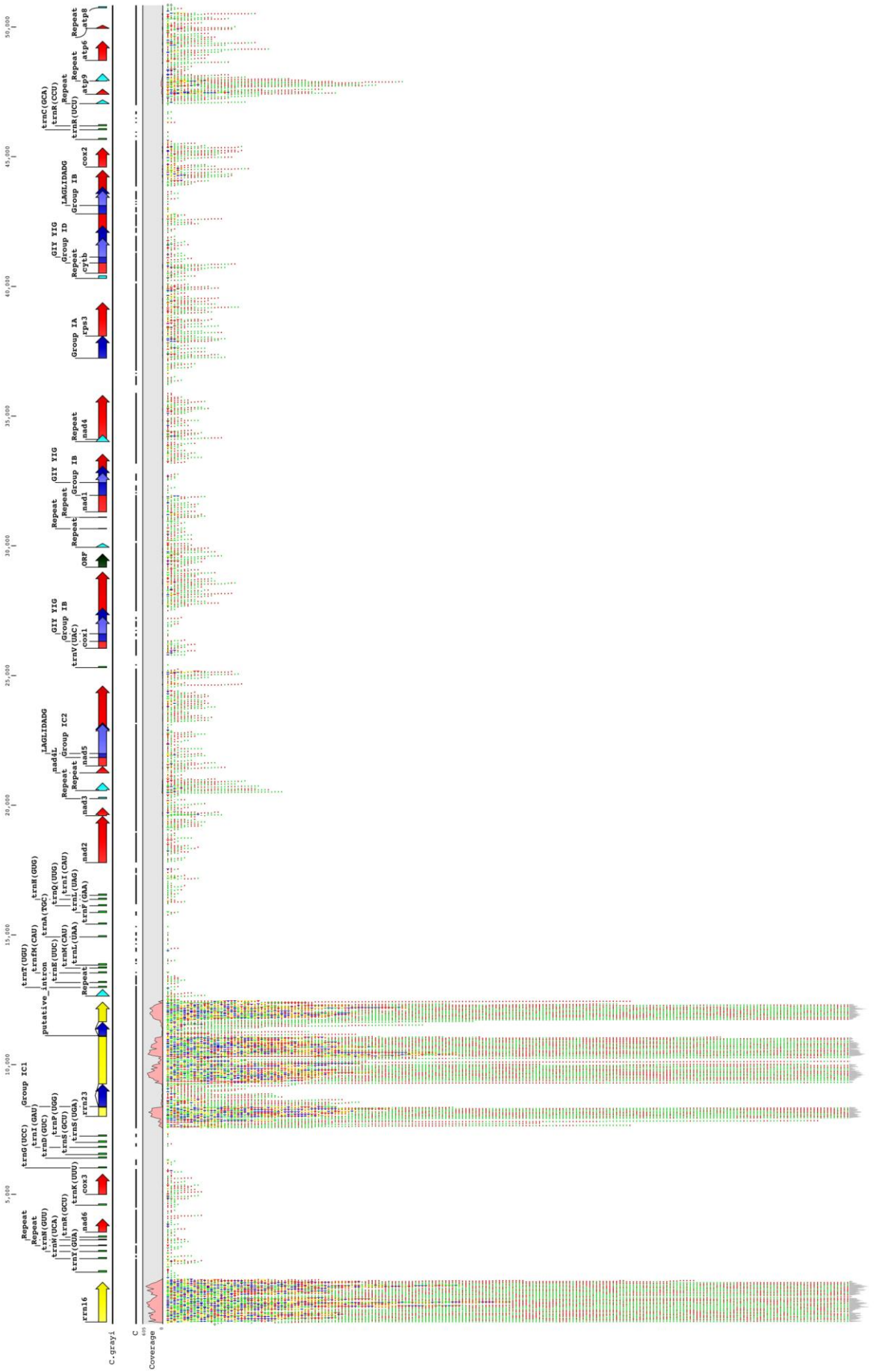


Figure. 2. *C. grayi* mitochondrial transcriptome. Mapped reads are shown below the annotated mt genome. Discontinuous line indicates RNA coverage.

3.3. Phylogeny

The *C. grayi* mt genome is similar to the *Peltigera* spp. mt genomes in general content and order of the genes, and they differ slightly in their coding sequences. The phylogenetic tree generated from the *cox1* gene and mt concatenated proteins correlated with each other and placed *C. grayi* close to the *Peltigera* spp.

Phylogenetic analysis of 12 concatenated mt protein sequences from *C. grayi* along with the two *Peltigera* spp. and 19 representative fungal species (Table 1) using maximum likelihood, maximum parsimony and Bayesian methods returned the same well-supported tree topology with robust bootstrap and posterior probability values (Fig. 3). The same tree topology was obtained using individual proteins Atp6, Nad5 and Cox1. The *C. grayi* and *Peltigera* spp. are coherent groups according to the analysis of concatenated mt proteins.

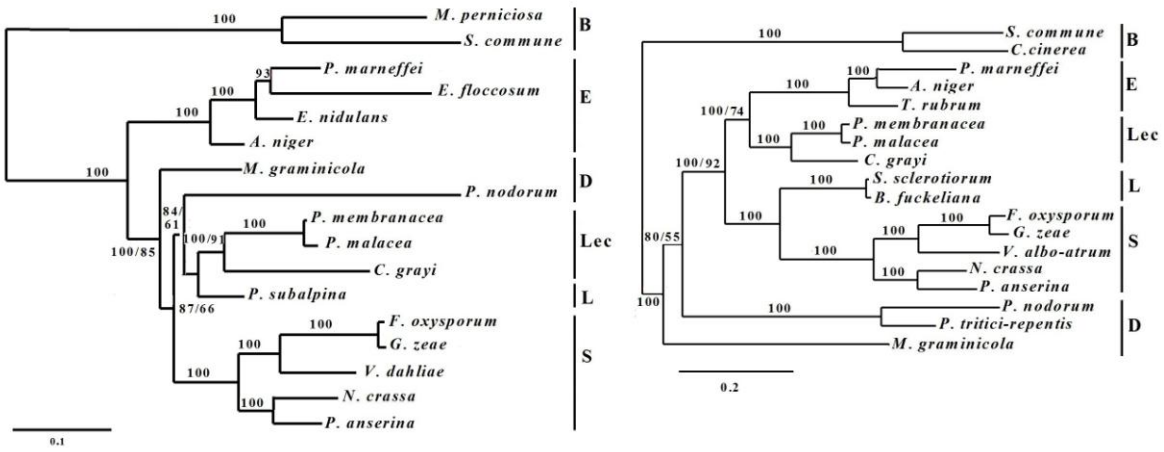


Figure 3. Phylogenetic trees based on mitochondrion encoded proteins and Rpb1, a nuclear encoded protein. a) Twelve proteins encoded by the mitochondria of *P. membranacea, P. malacea* and *C. grayi* were concatenated and aligned to corresponding sequences from representatives of other classes of fungi (Table 1). Branch support value, in percent, indicates Bayesian posterior probability, followed by maximum likelihood bootstrap value if different. Branches are drawn to scale, with the bar indicating 0.1 substitutions per site. b) The complete protein sequences of Rpb1 from the species in a) or their closest relatives available in Genbank, plus *P. nodorum* (Table 1), were aligned and used for building a maximum likelihood tree. Branch support values as in a) with the bar indicating 0.2 substitutions per site. B = *Basidiomycotina*; E = *Eurotiomycetes*; Lec = *Lecanoromycetes*; L = *Leotiomycetes*; D = *Dothidiomycetes*; S = *Sordariomycetes*. For full names and accession numbers see Table 1.

3.4. Comparison with *Peltigera* mt genomes

The *C. grayi* mt genome shares some of the common features of ascomycetes as well as *Peltigera* spp., including that all genes are coded in the same strand and a one base overlap between the *nad4L* and *nad5* genes. The gene order of the *C. grayi* mt genome is the same as in *Peltigera* spp. except for the arrangement of the *rps3* gene. Unlike *Peltigera* spp. mt genomes the *C. grayi* mt genome does not have a *dpoB* gene of presumed plasmid origin, nor the *rnpB* gene that encodes for RNaseP RNA component, the seven nucleotide short exon found in the *nad5* gene, the fused *cytb* and *cox2*, the lack of intergenic region between *nad2* and *nad3*, and an intron interruption in the *nad1* and *nad4L* genes.

3.5. Rps3

The *rps3* gene codes for one of the proteins in the small subunit of the ribosome which plays an important role in ribosome assembly (Loar et al . 2004). The mitochondrial Rps3 protein has conserved N and C terminal domains but is highly polymorphic across the fungal species (Bullerwell et al . 2000; Smits et al . 2007). The U11 region of the organelle *rnl* gene has a site specific target region for group I introns across fungal species. The location of the *rps3* gene within a *rnl* group I intron presumably helps in the transcription of the *rps3* gene and the maintenance of the group I intron (Sethuraman et al . 2009). The same subtype of intron i.e., group IA is also found in the *rnl* of the algal organelle genomes. However, the secondary structure of group IA introns varies and some possess a complete structure while others have a partial structure. In *C. grayi* the mitochondrial *rps3* appears in a novel arrangement, fused to intron sequences and free standing. The *rps3* gene in fungal mt has been found in three contexts. i) free standing ii) within a group I intron in *rnl*, and iii) fused to a HEG (homing endonucleases) in *rnl* (Fig. 5). The free standing *rps3* gene fused with part of a group IA intron at the 5' end reveals a new arrangement in fungal mt genomes. The *rnl* embedded version of the *rps3* gene is thought to be a canonical pattern in *Pezizomycotina* (filamentous ascomycetes) mitochondria, but the pattern seen in *C. grayi* is different from that in previously published fungal mt genomes. The group IA intron hosting the *rps3* gene varies among these lichenized fungal mitochondrial genomes and differs in terms of the secondary structure. The *C. grayi* mt genome group IA intron is partial or derived from a subtype of group IA and this intron degeneration suggests that group IA introns have a strong tendency to associate with the *rps3* gene. The predicted secondary structure of the group IA intron,

which includes 29 nt of the *rps3* gene, reveals that the loop part of the intron structure hosts the start codon of the *rps3* gene. The coverage and the contigous distribution of mRNA reads onto the group IA intron and the *rps3* gene reaffirms the importance of the group IA intron association with the *rps3* gene and its translational ability. The *rps3* gene in *C. grayi* is truncated by a stop codon (Fig. 2) and the phylogenetic tree generated from Rps3 proteins shows that it is separates deeply in the lineage when compared to the other *Lecanaromycetes* (Fig. 4).

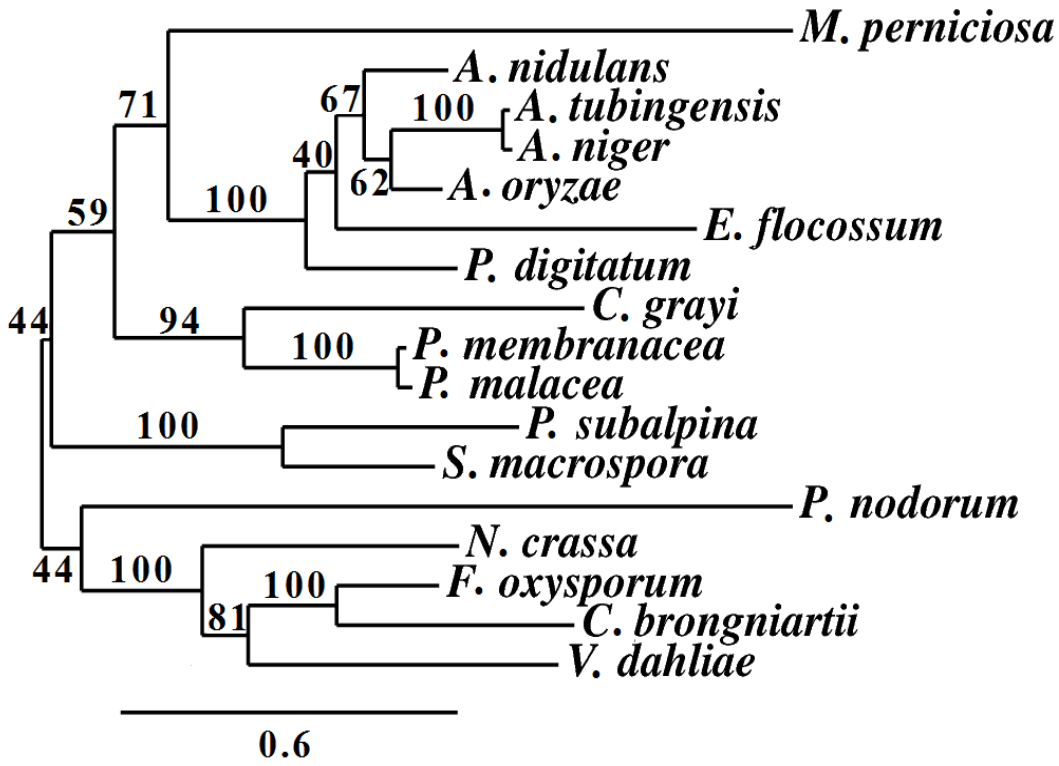


Figure 4. Phylogenetic trees based on mitochondrion encoded *rps3* protein. Rps3 protein encoded by the mitochondria of *P. membranacea*, *P. malacea* and *C. grayi* were aligned to corresponding sequences from representatives of other fungi. Branch support in percent, indicates ML bootstrap value. Branches are drawn to scale, with the bar indicating 0.6 substitutions per site.

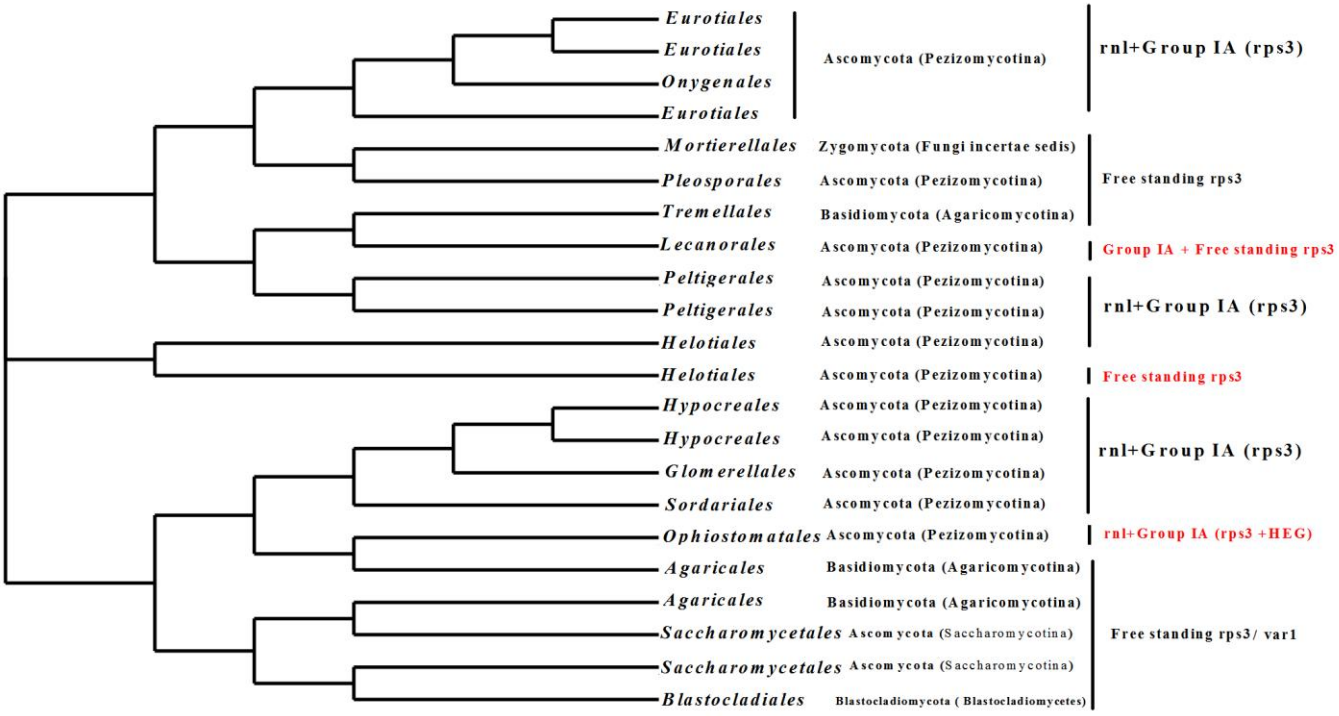


Figure 5. Cladistics tree topology. The fungal order grouping based on the *rps3* gene arrangement in the published fungal mitochondrion genomes.

CHAPTER IV

4.1. Introduction

The green algae mt genomes are classified into two types. First, a conventional type, which features a complex set of respiratory and ribosomal protein coding genes, a complete set of tRNA genes, canonical unfragmented single copy rRNA genes and 5S rRNA (e.g. *Trebouxiophyceae*). Second, a derived compact type that is deficient in genes for ribosomal protein, 5S rRNA, tRNAs, has fragmented and multiple copies of rRNA genes and has a smaller mitochondrial genome size (e.g. *Chlamydomonas* type) (Nedelcu et al . 2000).

To date, 15 published mt genomes from a variety of green algae have been deposited in Genbank. Their genome sizes vary between 12 kb and 95 kb. These size differences are due to i) intergenic spacer, repeats and introns ii) the number of protein coding genes present in these mt genomes varies from 6 to 72 and, iii) the arrangement and number of structural RNAs genes also varies greatly. These include the three ribosomal genes *rrn23*, *rrn16* and *rrn5* and tRNA genes that vary from a single tRNA gene (trnM) to a full set matching all amino acid codons. This suggests that in many cases most of the mt tRNAs are encoded by the nuclear genome and that the tRNAs are imported from the cytoplasm. The arrangement of the ribosomal genes in the mt genomes is completely different from fungal mt genomes. In most cases rRNA genes are split or an extra copy is present in the genome. Some of the algal mtDNAs lack 5S *rRNA* and ribosomal protein genes. Here, we present the largest known green algal conventional type mt genome, from *Asterochloris* sp with a size of ~110 kb.

4.2. General features of the *Asterochloris* sp mt genome

The circular map of the *Asterochloris* sp. mtDNA (Fig 6) consists of 110,932 bp and has a G+C content of 43.7%. , *Asterochloris* sp mtDNA contains nine *nad* genes, five *atp* genes, three cytochrome c oxidase subunits; one for cytochrome oxidase b, nine ribosomal small subunit protein genes, two ribosomal large subunit protein genes, three ribosomal (*rRNA*) genes and *tatC*. Also, the *Asterochloris* sp mt genome has 18 ORFs with no known functional similarity (Fig. 8), a single intronic ORF encoding the LAGLIDADG endonuclease motif (Belfort & Roberts, 1997) in *cox1*, a ribosomal large subunit 6 protein gene (*rpl6*) and a ribosomal protein small subunit gene (*rps11*). All the genes are coded on the same strand except for *tatC*. Ribosomal RNA genes are in a single copy and three introns reside within *rnl*, including a conventional group IA subtype intron which is usually found in fungal mt genomes hosting the *rps3* gene. Gene density is low in the mt DNA (Fig. 9), with known functional genes accounting for only ~50% of the sequence. Coding regions are in six clusters and evenly dispersed throughout the genome.

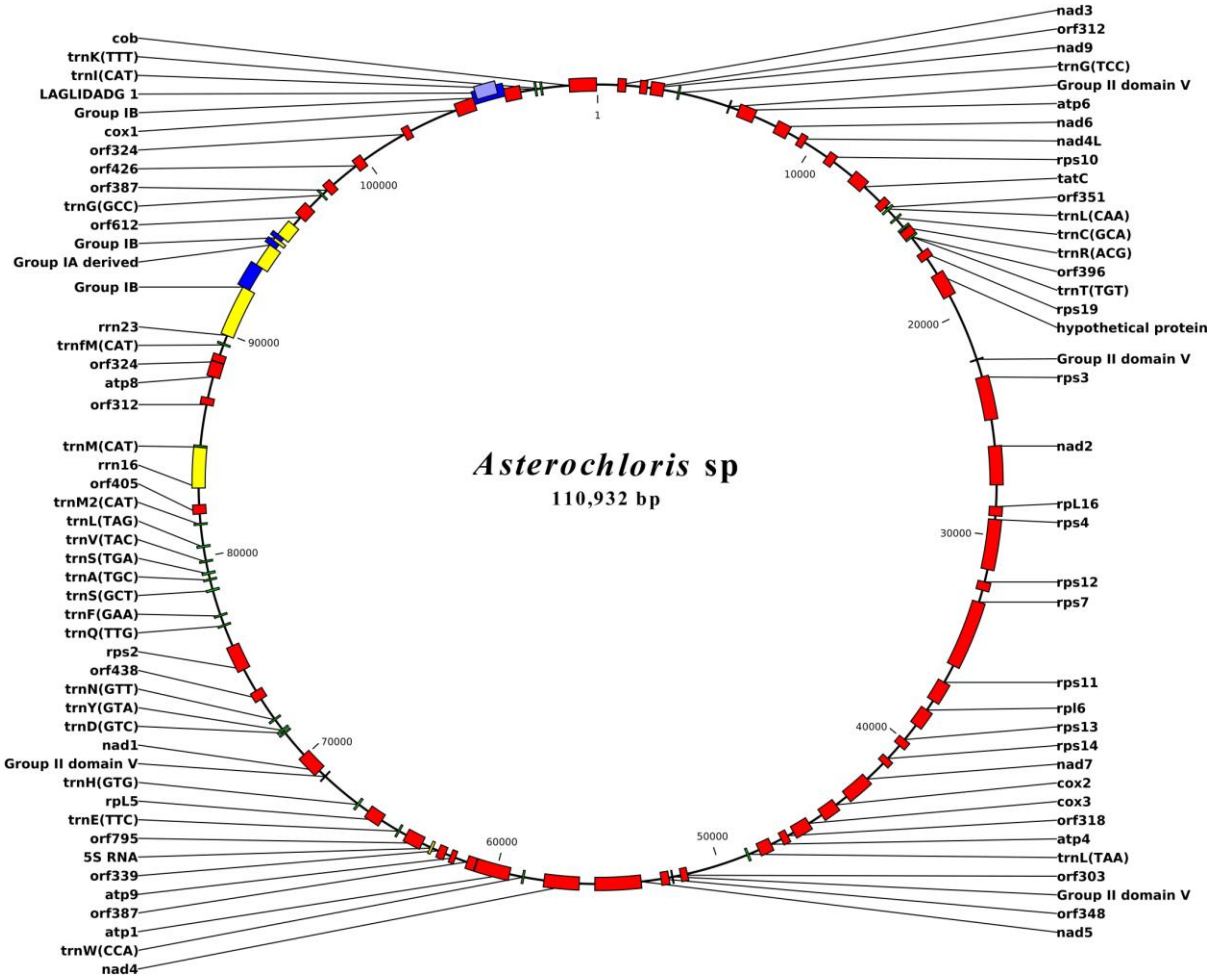


Figure 6. *Asterochloris* sp mitochondrial genome. Circular gene map representing the mt DNA of *Asterochloris* sp. All genes are transcribed clockwise (except *tatC*). Genes encoding proteins and ORFs are marked in red; ribosomal RNAs in yellow; tRNA genes, in green, are specified by standard one-letter amino acid code and anticodon. Introns and homing endonuclease genes (HEG) are shown in blue and light blue, and putative remnants of Group II domain V are marked in red.

4.3. tRNA gene loss in mt genomes

The tRNA genes have a patchy distribution across the eukaryotic organelle genomes. Most of the fungal mtDNA genomes contain 22-30 tRNAs, enough for their organelle function (Lang et al . 1999). Contrary to tRNA genes in fungal mtDNA, the organelle genomes of algae and plants vary largely in numbers. The *Asterochloris* sp. mt genome encodes 25 tRNAs and lacks two tRNA genes (*trnP* and *trnI* (GAU)) (Fig. 7), presumably compensated by import of nuclear encoded cytoplasmic tRNAs. The *Asterochloris* sp. mt genome is one among the algal mt genomes that encodes most of tRNAs in its genome. The tRNA gene transfers are usually one way, i.e., transfers are only from organelle to nucleus and there is no evidence of a vice versa transfer (Adams & Palmer, 2003). Although these genes are functionally origin specific most of the tRNAs have dual role in their importation functions ie., nucleus to chloroplast and nucleus to mitochondria (Börner, 1996; Gray et al . 1998). Compared to eubacterial tRNA genes most of the organelle tRNAs lack introns. Loss of tRNA genes from the organelles is probably due to less or no selection on the gene and partial RNA editing of the anticodon (Börner, 1996).

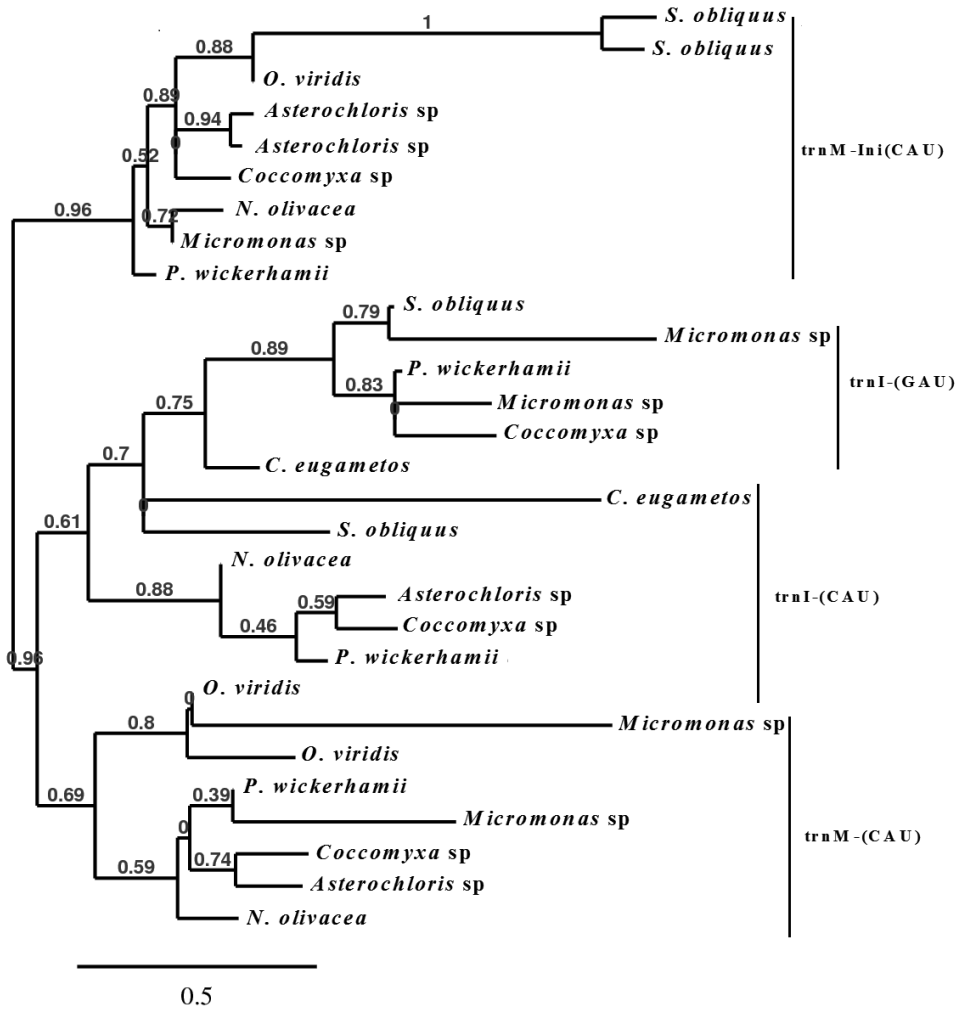


Figure 7. Algae mt Met and Ile tRNAs. The phylogenetic tree of mt Met and Ile tRNAs was constructed using the maximum likelihood method implemented in the PhyML program (v3.0 aLRT) using the substitution model HKY85. Reliability of internal branching was assessed using the aLRT test and the likelihood ratios are shown above the branches.

4.4. Introns

The *Asterochloris* sp. mt genome is populated by total of four group I introns inserted into *cox1* and rRNA (*rnl*). Interestingly, the introns present in *cox1* and the U11 region of *rnl* are located at the same sites as in fungal genes. The *Asterochloris* sp. *cox1* intron contains an open reading frame *orf930,* that displays two LAGLIDADG motifs, indicative of a endonuclease function. The insertion point of this intron is conserved, the position class is "V" (Fig. 12) as described (Férandon et al . 2010).

4.4.1. Group II introns

Certain intergenic sequences of the *Asterochloris* sp. mt genome are derived from degenerate group II introns and are identified as domain V which is the most conserved secondary structure and a key component of the catalytic core of group II introns (Lang et al . 2007; Seetharaman et al . 2006) . There are four 42 nt sequences which are structurally similar to the functionally important domain V of group II introns but the functionality of the intergenic regions is unknown (Fig. 6). This suggests that the *Asterochloris* sp. mt genome has harboured group II introns previously and that the domain V part of the group II intron is strongly selected to remain in the genome even though other parts of the group II introns have disappeared.

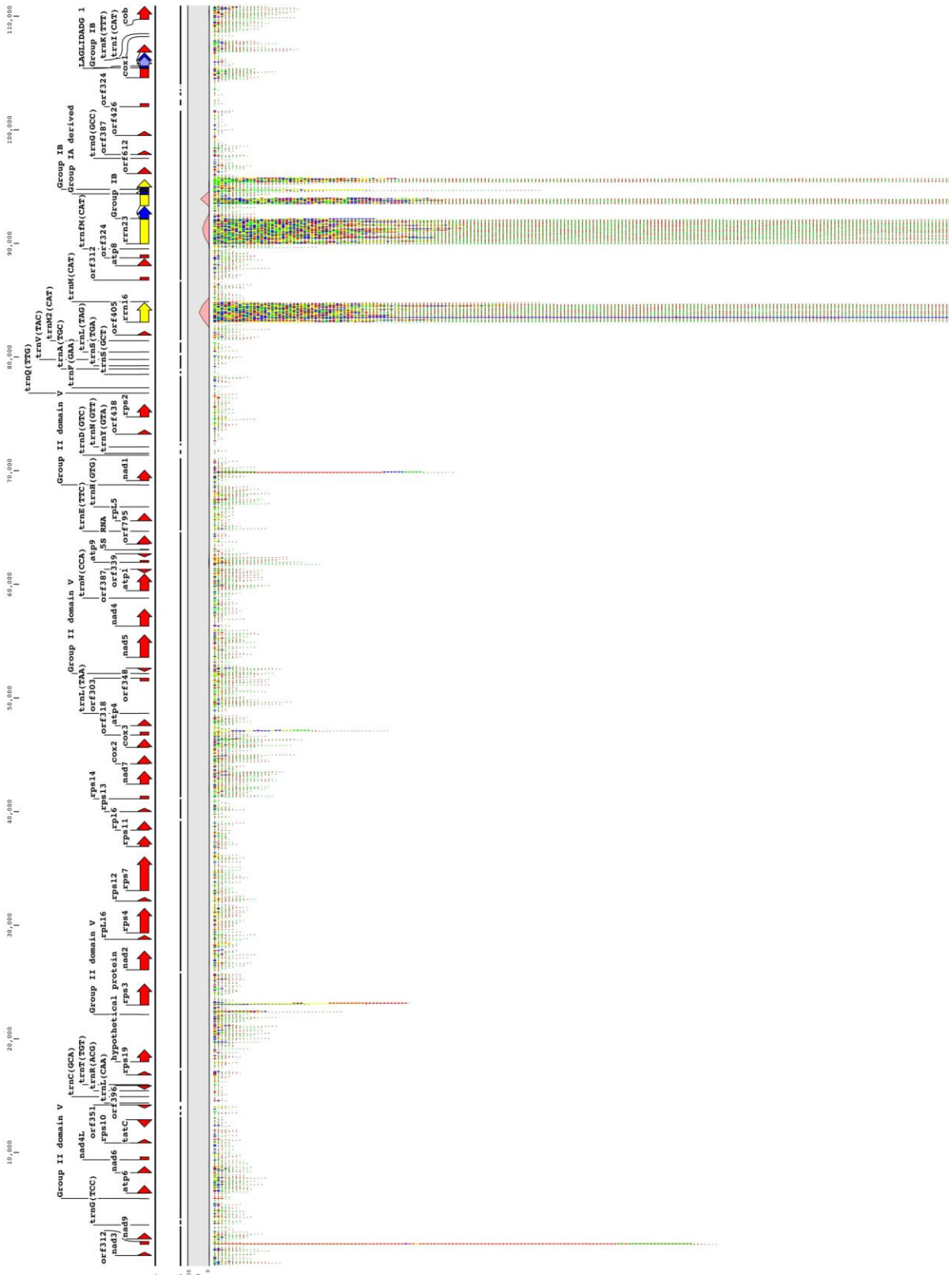


Figure 8. ***Asterochloris* sp mitochondrial transcriptome.** Mapped reads are shown below the annotated mt genome. Discontinuous line indicates RNA coverag

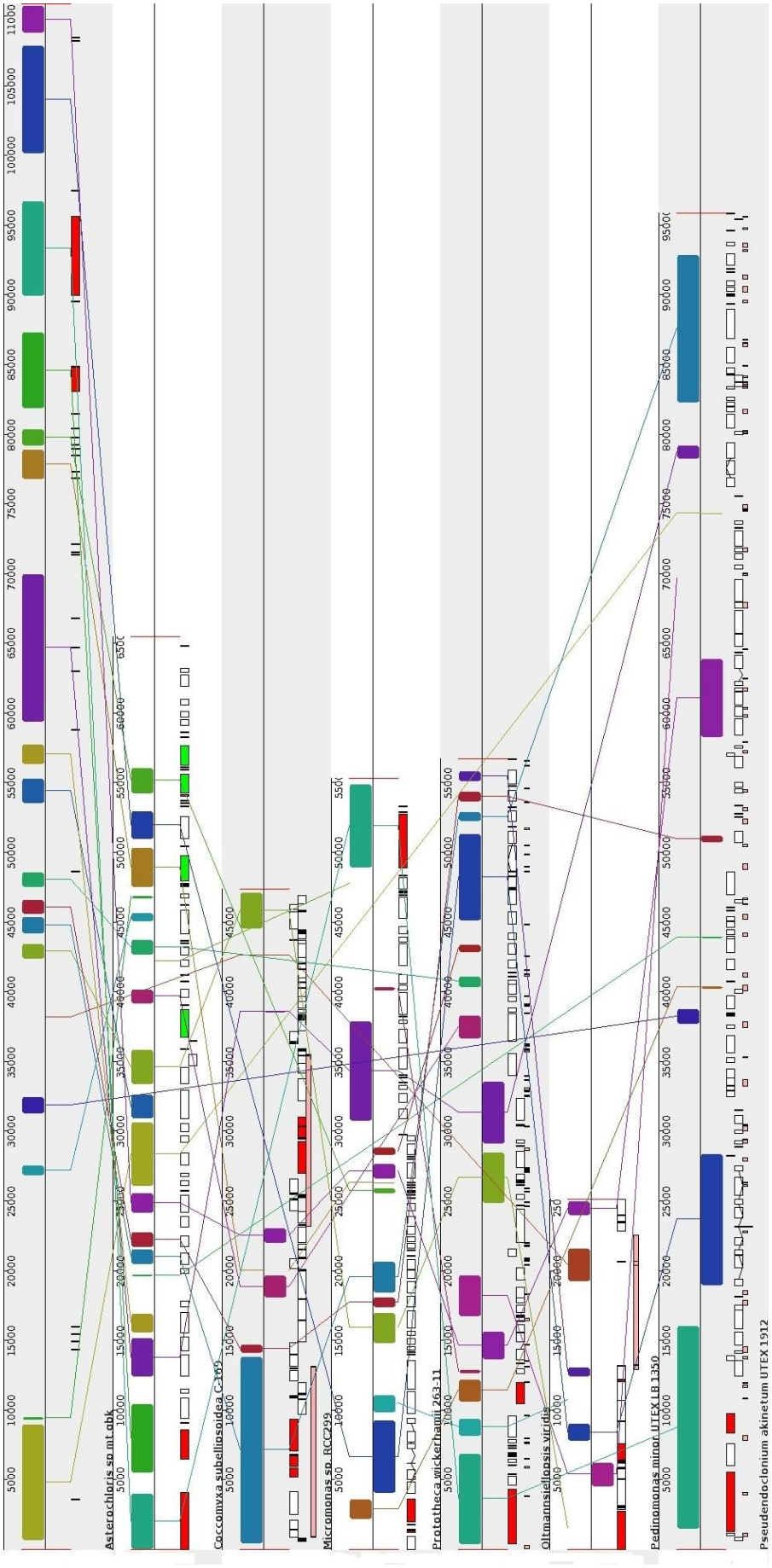


Figure 9. ***Asterochloirs* sp mitochondrial gene order**. Comparison of mtDNA of *Asterochloris sp* and green algae was aligned using mauve online tool to check order of genes.

CHAPTER V

5.1.General features of the cp genome of *Asterochloris* sp.

The circular map of the *Asterochloris* sp. cp genome consists of 217,546 bp, has an overall G+C content of 39%, has 74 ORFs and is one of the largest cp genomes known in green algae. The genome has 16S, 23S, 5S rRNA genes and 30 tRNA genes covering all amino acid codons and four subunits of a prokaryotic-type RNA polymerase (*rpoA, rpoB, rpoC1* and *rpoC2*) responsible for transcription. It also has four genes that encode the polypeptides of Photosystem I (*psaA, psaB, psaC* and *psaJ*), 14 genes for Photosystem II (*psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbJ, psbk, psbL, psbM, psbN, psbT* and *psbZ*), four genes for the cytochrome complex (*petA, petB, petD* and *petG*), nine genes for ribosomal protein small subunits (*rps2,rps3,rps7, rps8, rps9,rps11, rps12, rps14* and *rps18*), nine genes for ribosomal protein large subunit (*rpl2, rpl5, rpl12, rpl14, rpl15, rpl16, rpl19, rpl20, rpl23* and *rpl36*), six genes for ATP synthase (*atpA, atpB, atpE, atpF, atpH* and *atpI*), five genes for metabolic functions (*chIB, chIN, chlL, chlI* and *accD*), two genes for translation associated functions (*tufA, infA*), four genes for assembly and membrane insertion related functions (*ycf3, ycf4, ccsA, minD* and *cemA*), two genes for protein quality control related functions (*ftsH* and *clpP*) and 6 ORFs encoding hypothetical proteins. One ORF encodes a protein of 506 aa and is not similar to any known proteins. The genes are encoded in both strands (Fig. 10).

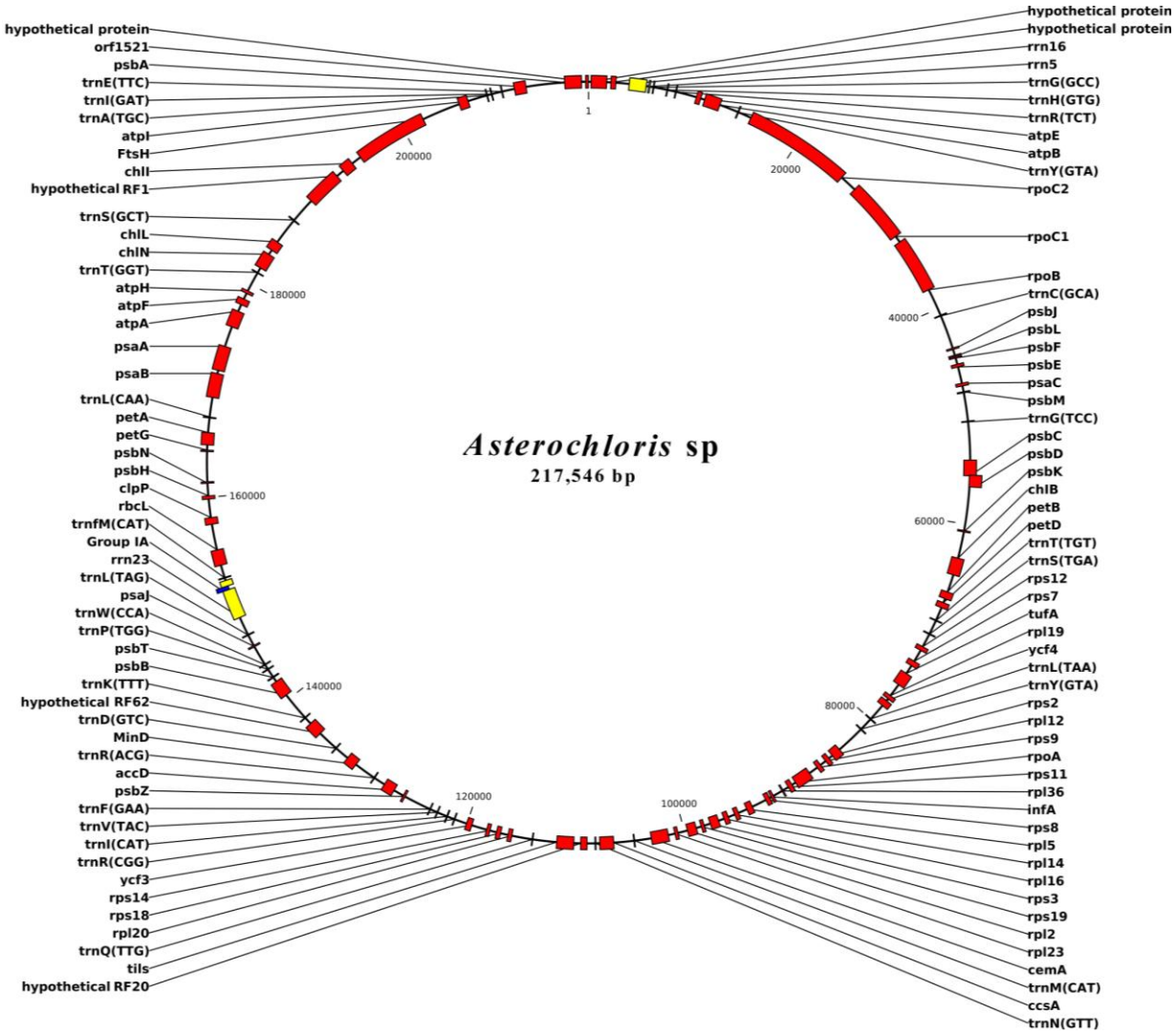


Figure 10. *Asterochloris* sp chloroplast genome. Circular gene map representing the cp DNA of *Asterochloris* sp. Genes are transcribed in both stands. Genes encoding proteins, ORFs, and ribosomal RNAs are indicated by red and yellow boxes, respectively; tRNA genes, in green, are specified by standard one-letter amino acid code and anticodon. Introns and homing endonuclease genes (HEG) are shown in blue and light blue.

The conserved gene cluster *accD* and *cysA* is not seen in the *Asterochloris* sp. Cp and moreover, the *cysA* gene is not present. The *psbD* and *psbC* genes appears fused and is supported by transcript mapping (Fig. 11).

The *Asterochloris sp* cp genome contains a single group I intron inserted into the large rRNA gene (*rrn23*) (Fig. 11) and the large amount of intergenic regions in the *Asterochloris* sp. cp genome probably plays important role in the indirect way to the evolution of multicellularity (Smith & Lee, 2010) in the lineage that gave rise to *Asterochloris* sp. was a major contributor to this large intergenic regions.

When compared to green algae cp genomes, rRNA genes are present as single and unfragmented copies in the *Asterochloris* sp. cp genome (Table 3). The conserved gene cluster *accD* and *cysA* is not found in the *Asterochloris* sp. cp and moreover, the *cysA* gene is not present. The *psbD* and *psbC* genes appears fused and this is supported by transcript mapping (Fig. 11).

The *Asterochloris* sp. cp genome contains a single group I intron inserted into the large rRNA gene (*rrn23*) (Fig. 11) and the large amount of intergenic regions in the *Asterochloris* sp. cp genome probably plays important roles in the indirect way to the evolution of multicellularity in the lineage that gave rise to *Asterochloris*.

When compared to green algae cp genomes, rRNA genes are present as single and unfragmented copies in the *Asterochloris* sp. cp genome (Table 3).

5.2. TilS

The TilS enzyme is responsible for differentiating tRNA-Ile2 (CAU) from tRNA-Met (CAU) by converting the codon specificity from AUG to AUA as well as converting the aminoacylation specificity from methionyl-tRNA synthetase to that of isoleucyl-tRNA synthetase in eubacteria and organelle genomes of eukaryotes (Nakanishi et al . 2005). A recent study suggests that the lack of the *tilS* gene and associated function is compensated by taking up an other anticodon (TAT) in bacteria (Fabret et al . 2011). In *A. thaliana*, a mutation in the *tilS* gene homolog is lethal at the embryo stage. Some other studies reveal that lack of *tilS* and associated function leads to ATA(Ile) codon-deficient translation (Soma et al . 2003). Interestingly, the *tilS* gene in *Asterochloris* sp. is found in two different parts in the cp genome and they are present on different strands; orf1542 (N terminal) and orf-RF62 (C terminal). The *tilS* gene shows patchy distribution in algae cp genomes (chlorophyta) (Table 3). The *tilS* gene in the class *Trebouxiophyceae* species has two different forms; complete and a split, and some green algal cp genomes completely lack the *tilS* gene. The presence of tRNA-Ile2 along with the split *tilS* gene and lack of its nuclear counterpart in the *Asterochloris* sp. nulcear genome indicated that the split *tilS* gene parts are functional. From the evolution point of view, degradation or reduction of the selection of the *tilS* gene has resulted in splitting of the gene or loss in some genomes. There are several factors which are involved in the non-maintanence of the *tilS* gene in the organelles which include the importance of redox-associated functions in organelles that might increase the free-radical-induced mutagenic load for genes in organelles thus favouring their transfer to the nucleus (Allen & Raven, 1996). The nuclear genome undergoes recombination during sexual reproduction and it creates a lot of variations thus acting as a source for natural selection in response to the environment. Organelle genomes have a higher mutation rate and are prokaryotic in origin and is therefore difficult for nuclear DNA to get integrated into the organelle DNA as well as to maintain a different code for translation.

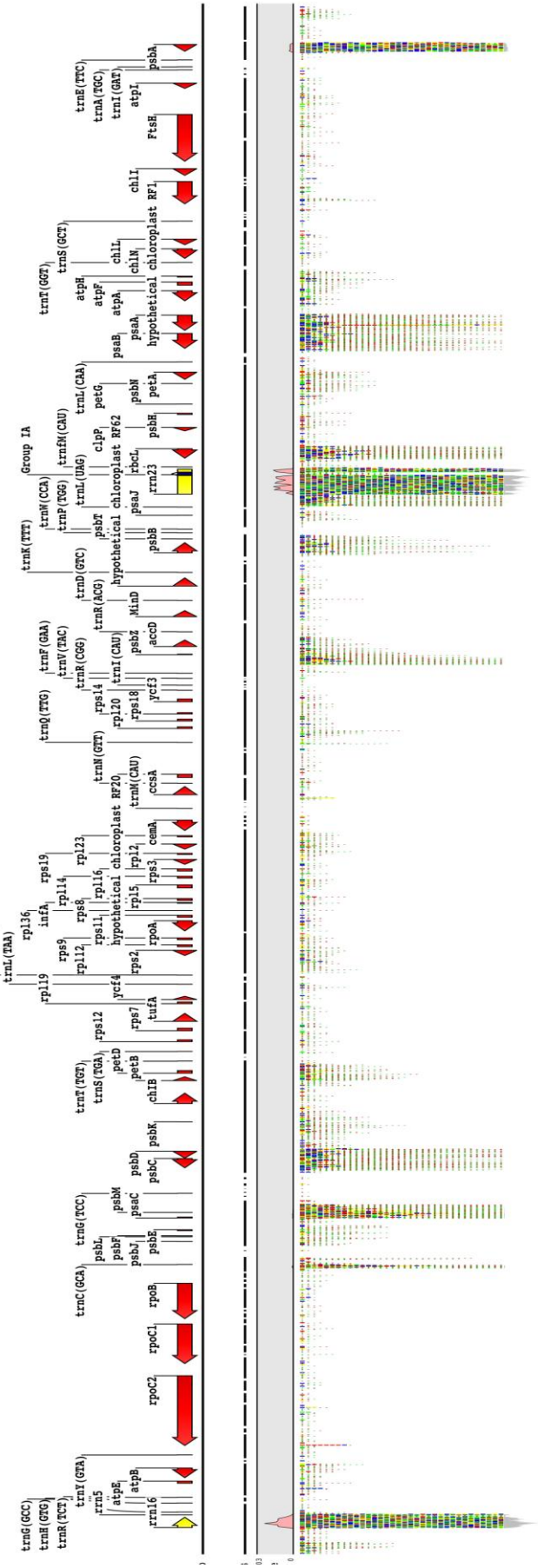


Figure 11. ***Asterochloirs* sp chloroplast transcriptome.** Mapped reads are shown below the annotated cp genome. Discontinuous line indicates RNA coverage.

5.2.1. Intermediate stage of gene transfers –"split gene"

In general, studies have shown that a gene split could occur during gene transfer events. So far, mitochondrial genes of plants and algae have been reported to show gene splitting. The processes of gene splitting and transfers have been described in angiosperms (Adams, 2000; Adams & Palmer, 2003). For instance, an angiosperm mt *rpl2* gene is split and 3' end of the gene is first transferred to the nucleus followed by the 5' end of the gene (Adams et al . 2001). A similar split gene pattern and transfer is also noticed in algae. But, gene transfer events are observed in two different lineages of green algal mt genomes (Adams & Palmer, 2003). In *Scenedesumus* the mt-*cox2* gene is split and the 3' part of the gene is transferred to the nuclear genome (Kück et al . 2000) whereas in *Chlamydomonas* and *Polytomella* the 5'part of the mt-*cox2* gene is transferred to the nucleus (Pérez-Martínez et al . 2001). These observations lead to the proposal that the split pattern of the *tilS* gene in the *Asterochloris* sp. cp genome is the "intermediate stage of gene transfer event" and a second or an extra copy of *tilS* gene is not found in the nuclear genome of *Asterochloris* sp and therefore this strengthens the argument of an intermediate stage. A similar pattern of a split *tilS* gene is also observed in other closely related algal cp genomes (Table 3). If the above mentioned instances of split mt genes are to be considered as a part of the eukaryotic genome evolution then this cp split *tilS* gene can also be treated in the same manner.

CHAPTER VI

6. GENERAL CONCLUSIONS

The mitochondrial genomes of *Peltigera membranacea*, *Peltigera malacea* and organelle genomes of *Cladonia grayi* represent the first completely annotated genomes of lichens. A feature among organelle genomes are the intron numbers, type of intron and intron insertion points which vary widely between species. These observations on introns reveal that group I introns are not specific to fungal mt genomes (Haugen et al . 2005) because intron sequences are not conserved, so they are unlikely to remain in specific lineage and most likely they would spread across the tree of life via horizontal transfer (Bernard, 1989). The comparative study of the organelle genomes allows us to explore diversity in overall organization, unusual gene features like a split *tilS* gene, fusions and tRNA evolution. Information on organelle genome evolution also helps us to study the peculiar distribution of genetic code variants and its interplay with tRNA types. The competition of intracellular variation in these genomes presumably leads to organelle evolution. Natural selection is a real force for organelle gene evolution but there are no specific arguments for determining the architectural features of organelle genomes . The genome-wide transcriptomic analysis of these organelles provides a rich source of data for validating annotations, unusual gene features and TARs (Fig. 13). All these findings will enhance analysis of the forthcoming *P. membranacea*, *P. malacea*, *C. grayi* and *Asterochloris* sp. nuclear genomes. The complete characterization of the organelle genomes has improved our knowledge of their function and evolution in chlorolichens and cyanolichens.

REFERENCES

1. Adams, 2000. Repeated , recent and diverse transfers of a mitochondrial gene to the nucleus in flowering plants. *Nature* 408: 354.

2. Adams KL, Ong HC, Palmer JD, 2001. Mitochondrial Gene Transfer in Pieces: Fission of the Ribosomal Protein Gene rpl2 and Partial or Complete Gene Transfer to the Nucleus. *Molecular Biology and Evolution* 18: 2289-2297.

3. Adams KL, Palmer JD, 2003. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. *Molecular Phylogenetics and Evolution* 29: 380-395.

4. Allen J, Raven J, 1996. Free-radical-induced mutation vs redox regulation: Costs and benefits of genes in organelles. *Journal of Molecular Evolution* 42: 482-492.

5. Andersson S, 2003. On the origin of mitochondria : a genomics perspective. *Philosophical Transactions : Biological Sciences* 358: 165.

6. Belfort M, Roberts RJ, 1997. Homing endonucleases: keeping the house in order. *Nucleic Acids Research* 25: 3379-3388.

7. Bernard D, 1989. Group I introns as mobile genetic elements: Facts and mechanistic speculations — a review. *Gene* 82: 91-114.

8. Börner, 1996. RNA editing changes the identity of a mitochondrial tRNA in marsupials. *EMBO Journal* 15: 5949.

9. Brouard, 2010. The exceptionally large chloroplast genome of the green alga Floydiella terrestris illuminates the evolutionary history of the Chlorophyceae. *Genome Biology and Evolution* 2: 240.

10. Bullerwell CE, Burger G, Lang BF, 2000. A novel motif for identifying Rps3 homologs in fungal mitochondrial genomes. *Trends in Biochemical Sciences* 25: 363-365.

11. Chevreux TP, Wetter T, Suhai S, 1999. Genome sequence assembly using trace signals and additional sequence information In Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB) (Germany) pp. 45-56.

12. Esaki M, 2004. Mitochondrial Protein Import. *Journal of Biological Chemistry* 279: 45701.

13. Fabret C, Dervyn E, Dalmais B, Guillot A, Marck C, Grosjean H, Noirot P, 2011. Life without the essential bacterial tRNAIle2–lysidine synthetase TilS: a case of tRNA gene recruitment in *Bacillus subtilis*. *Molecular Microbiology* 80: 1062-1074.

14. Férandon C, Moukha S, Callac P, Benedetto J-P, Castroviejo M, Barroso G, 2010. The *Agaricus bisporus* cox1 gene: the longest mitochondrial gene and the largest reservoir of mitochondrial group I introns. *PLoS One* 5: e14048.

15. Gray MW, Burger G, Lang BF, 1999. Mitochondrial evolution. *Science* 283: 1476-1481.

16. 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 Acids Research* 26: 865-878.

17. Green BR, 2011. Chloroplast genomes of photosynthetic eukaryotes. *Plant Journal* 66: 34-44.

18. Griffiths AJF, 1996. Mitochondrial inheritance in filamentous fungi. *Journal of Genetics* 75: 403-414.

19. Haugen P, Simon DM, Bhattacharya D, 2005. The natural history of group I introns. *Trends in Genetics* 21: 111-119.

20. Hausner G, 2003. Fungal mitochondrial genomes, plasmids and introns. In Applied Mycology and Biotechnology, K A Dilip, and G K George, eds. (Elsevier)pp. 101-131.

21. Kashi, Y, 2006. Simple sequence repeats as advantageous mutators in evolution. *Trends in Genetics* 22: 253.

22. Kleine T, Maier UG, Leister D, 2009. DNA Transfer from Organelles to the Nucleus: The Idiosyncratic Genetics of Endosymbiosis. *Annual Review of Plant Biology* 60: 115-138.

23. Kück U, Jekosch K, Holzamer P, 2000. DNA sequence analysis of the complete mitochondrial genome of the green alga Scenedesmus obliquus: evidence for UAG being a leucine and UCA being a non-sense codon. *Gene* 253: 13-18.

24. Lambowitz AM, Belfort M, 1993. Introns as Mobile Genetic Elements. *Annual Review of Biochemistry* 62: 587-622.

25. Lambowitz AM, Caprara MG, Zimmerly S, Perlman PS, 1999. Group I and group II ribozymes as RNPs: clues to the past and guides to the future. In: The RNA World, Second Edition (R.F. Gesteland, T.R. Cech, and J.F. Atkins, Editors) edn (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York).

26. Lang BF, Burger G, O'Kelly CJ, Cedergren R, Golding GB, Lemieux C, Sankoff D, Turmel M, Gray MW, 1997. An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. *Nature* 387: 493-497.

27. Lang BF, Gray MW, Burger G, 1999. Mitochondrial genome evolution and the origin of eukaryotes. *Annu Rev Genet* 33: 351-397.

28. Lang BF, Laforest MJ, Burger G, 2007. Mitochondrial introns: a critical view. *Trends in Genetics* 23: 119-125.

29. Loar JW, Seiser RM, Sundberg AE, Sagerson HJ, Ilias N, Zobel-Thropp P, Craig EA, Lycan DE, 2004. Genetic and Biochemical Interactions Among Yar1, Ltv1 and RpS3 Define Novel Links Between Environmental Stress and Ribosome Biogenesis in Saccharomyces cerevisiae. *Genetics* 168: 1877-1889.

30. Lynch M, Koskella B, Schaack S, 2006. Mutation Pressure and the Evolution of Organelle Genomic Architecture. *Science* 311: 1727-1730.

31. Nakanishi K, Fukai S, Ikeuchi Y, Soma A, Sekine Y, Suzuki T, Nureki O, 2005. Structural basis for lysidine formation by ATP pyrophosphatase accompanied by a lysine-specific loop and a tRNA-recognition domain. *Proceedings of the National Academy of Sciences of the United States of America* 102: 7487-7492.

32. Nash TH, 2008. Lichen Biology 2nd edn (Cambridge University Press).

33. Nedelcu AM, Lee RW, 1998. A degenerate group II intron in the intronless mitochondrial genome of Chlamydomonas reinhardtii: evolutionary implications. *Molecular Biology and Evolution* 15: 918-922.

34. Nedelcu AM, Lee RW, Lemieux C, Gray MW, Burger G, 2000. The Complete Mitochondrial DNA Sequence of Scenedesmus obliquus Reflects an Intermediate Stage in the Evolution of the Green Algal Mitochondrial Genome. *Genome Research* 10: 819-831.

35. O'Brien HE, Miadlikowska J, Lutzoni F, 2005. Assessing host specialization in symbiotic cyanobacteria associated with four closely related species of the lichen fungus *Peltigera*. *European Journal of Phycology* 40: 363-378.

36. Paquin B, Laforest M-J, Forget L, Roewer I, Wang Z, Longcore J, Lang BF, 1997. The fungal mitochondrial genome project: evolution of fungal mitochondrial genomes and their gene expression. *Current Genetics* 31: 380-395.

37. Pérez-Martínez X, Antaramian A, Vázquez-Acevedo M, Funes S, Tolkunova E, d'Alayer J, Claros MG, Davidson E, King MP, González-Halphen D, 2001. Subunit II of Cytochrome c Oxidase in Chlamydomonad Algae Is a Heterodimer Encoded by Two Independent Nuclear Genes. *Journal of Biological Chemistry* 276: 11302-11309.

38. Rikkinen J, Oksanen I, Lohtander K, 2002. Lichen guilds share related cyanobacterial symbionts. *Science* 297: 357-357.

39. Seetharaman M, Eldho NV, Padgett RA, Dayie KT, 2006. Structure of a self-splicing group II intron catalytic effector domain 5: Parallels with spliceosomal U6 RNA. *RNA* 12: 235-247.

40. Sethuraman J, Majer A, Iranpour M, Hausner G, 2009. Molecular Evolution of the mtDNA Encoded rps3 Gene Among Filamentous Ascomycetes Fungi with an Emphasis on the Ophiostomatoid Fungi. *Journal of Molecular Evolution* 69: 372-385.

41. Smith DR, Lee RW, 2010. Low Nucleotide Diversity for the Expanded Organelle and Nuclear Genomes of Volvox carteri Supports the Mutational-Hazard Hypothesis. *Molecular Biology and Evolution* 27: 2244-2256.

42. Smits P, Smeitink JAM, van den Heuvel LP, Huynen MA, Ettema TJG, 2007. Reconstructing the evolution of the mitochondrial ribosomal proteome. *Nucleic Acids Research* 35: 4686-4703.

43. Soma A, Ikeuchi Y, Kanemasa S, Kobayashi K, Ogasawara N, Ote T, Kato J-i, Watanabe K, Sekine Y, Suzuki T, 2003. An RNA-Modifying Enzyme that Governs Both the Codon and Amino Acid Specificities of Isoleucine tRNA. *Molecular Cell* 12: 689-698.

44. Timmis K, 2004. Endosymbiotic gene transfer : organelle genomes forge eukaryotic chromosomes. *Nature Reviews - Genetics* 5: 123.

Appendix

Table 1. Fungal species, genes, mt genomes, GenBank accession numbers.

Species	Class	Mitochondrial features												Nuclear proteins	
		Size (~kb)	GC %	Protein genes	ORF	tRNAs	*dpoB*	*rnpB*	*rps3*	*cox1* (bp)	Cox1 (aa)	*cox1* introns	Mt genome Accession	TilS Accession	Rpb1 Accession
Peltigera membranacea	Lecanoromycetes	62.8	27	14	1	26	1	1	1	10198	631	6	JN088165	JN032401	JN596954
Peltigera malacea	Lecanoromycetes	63.4	27	14	1	26	1	1	1	12690	627	9	JN088164	JN032402	JN639885
Cladonia grayi	Lecanoromycetes	50.8	29	14	1	26	0	0	1	2966	552	1			
Aspergillus niger	Eurotiomycetes	31.1	26	14	2	25	0	0	1[c]	5059	566	3	NC_007445	XP_001398711	XP_001389676
Aspergillus tubingensis	Eurotiomycetes	33.6	26	14	2	25	0	0	1[c]	5073	565	4	NC_007597		
Aspergillus nidulans	Eurotiomycetes	33.2[a]	25	14	0	28	0	1[d]	1[d]	4973	567	3	Footnote [a]		
Epidermophyton floccossum	Eurotiomycetes	30.9	23	14	4	25	0	0	1	2703	528	1	NC_007394		
Paracoccidioides brasiliensis	Eurotiomycetes	71.3	20	14	0	25	0	0	1	12428	710	8	NC_007935		
Penicillium marneffei	Eurotiomycetes	35.5	24	14	10	28	0	0	1	10027	561	7	NC_005256	XP_002144809	XP_002146907
Penicillium digitatum	Eurotiomycetes	28.9	25	14	0	27	0	0	1	1775	591	0	NC_015080		
Trichophyton rubrum	Eurotiomycetes														XP_003231777
Fusarium oxysporum	Sordariomycetes	34.4	31	14	0	25	0	0	1	1593	530	0	AY945289		EGU79617
Gibberella zeae	Sordariomycetes	95.5	31	14	1	28	0	0	1	16914	530	12	NC_009493	XP_384534	XP_381092
Glomerella lindemuthiana	Sordariomycetes									1380	460	2	HM450146		
Neurospora crassa	Sordariomycetes	64.8	36	14	6	28	0	0	1	7972	557	4	footnote [b]	XP_956207	XP_964097
Podospora anserina	Sordariomycetes	100	30	13	37	27	0	0	1	24506	541	14	NC_001329		XP_001912461
Verticillium dahliae	Sordariomycetes	27.2	27	14	0	25	0	0	1	1623	540	0	NC_008248		
Verticillium albo-atrum	Sordariomycetes														XP_003002835
Mycosphaerella graminicola	Dothideomycetes	43.9	31	14	8	27	0	1[c]	0	2010	669	0	NC_010222	EGP84297	EGP87916
Phaeosphaeria nodorum	Dothideomycetes	49.7	29	12	3	27	0	0	1	1668	556	0	NC_009746		XP_001801796
Phialocephala subalpina	Leotiomycetes	43.7	27	15	4	28	0	0	1	1722	573	0	JN031566		
Sclerotinia sclerotiorum	Leotiomycetes													XP_001589873	XP_001593006
Coprinopsis cinerea	Agaricomycetes														XP_001828577
Moniliophthora perniciosa	Agaricomycetes	109	31	14	54	26	1	0	1	8929	533	6	NC_005927		
Schizophyllum commune	Agaricomycetes	49.7	21	14	5	27	0	0	1	1583	527	0	NC_003049		XP_003037684
Mortierella verticillata	Zygomycetes	58.7	27	14	8	26	0	1	1	5811	534	3	NC_006838		
Rhizopus oryzae	Zygomycetes	54.1	26	14	9	24	0	1	0	4994	505	3	NC_006836		
Glomus intraradices	Glomeromycetes	70.6	37	14	0	26	0	0	0	7990	529	12	NC_012056		

a) http://megasun.bch.umontreal.ca/People/lang/species/asp/. [Incomplete genome]

b) http://www.broadinstitute.org/annotation/genome/neurospora/MultiDownloads.html

c) Identified in this study using the deposited genomes: *M. graminicola rnpB* (mt24998-24821), *A. tubingensis rps3* (mt16536-17753), *A. niger rps3* (mt16403-17620).

d) Identified in this study using *A. nidulans* mt genome: *rnpB* (mt20776-20962), *rps3* (mt2957-4189).

Features	*P. membranacea*	*P. malacea*	*C. grayi*	*Asterochloris* (mt)	*Asterochloris* (cp)
size	62,785 bp	63,363 bp	50,836 bp	110,932 bp	217,546 bp
Genes	15	15	15	32	67
Introns	17	20	8	4	1
ORFs	1	1	1	18	7
dpoB	Pseudogene	1	0	0	0
tRNA	26	26	26	25	30
rRNA	2	2	2	3	3
AT%	72.9	72.6	70.5	47.3	61.0
rnpB	1	1	0	0	0
Genes interrupted by introns	5	7	4	2	1
(Non –ATG)	*3* *nad6, cox2, cytb*	3 *nad6,cox2,cytb*	1 *cox2*	0	0

Table 2. Features of organelle genomes

Table 3. Algal cp genome features

Organisms	Order	Class	*tilS*	trnI(CAU)	*cysA*	*accD*	*ftsH*	*minD*	*infA*	*psbD-psbC*	rRNA copies	introns (*rrn23*)	tRNA
Asterochloris sp	Microthamniales	Trebouxiophyceae	split	+	-	+	+	+	+	fused	single	single intron	30
Coccomyxa sp	Trebouxiophyceae incertae sedis	Trebouxiophyceae	split	+	+	+	+	+	+	fused	single	+	32
Helicosporidium	Trebouxiophyceae incertae sedis	Trebouxiophyceae	split	+	-	+	+	-	-	-	single	-	25
Chlorella variabilis	Chlorellales	Trebouxiophyceae	complete	+	+	+	+	+	+	fused	single	-	32
Chlorella vulgaris	Chlorellales	Trebouxiophyceae	complete	+	+	+	+	+	+	fused	single	single intron	33
Parachlorella	Chlorellales	Trebouxiophyceae	complete	+	+	+	+	+	+	fused	single	-	32
Leptosira	Ctenocladales	Trebouxiophyceae	split	+	+	+	+	+	+	not fused	single	-	28
Pedinomonas	Pedinomonadales	Pedinophyceae	-	-	+	+	+	+	-	fused	two	-	34
Micromonas pusilla	Mamiellales	Mamiellophyceae	-	-	-	-	-	-	-	-	two	-	7
Oltmannsiellopsis viridis	Oltmannsiellopsis	Chlorophyta incertae sedis	-	-	-	+	+	+	+	fused	two	++	28
Chlamydonmonas reinhardtii	Chlamydomonadales	Chlorophyceae	-	-	-	-	-	-	-	not fused	two	single intron	34

Table 4. Codon usage.

P. membranacea	62,785 bp	*P. malacea*	63,363 bp	*C. grayi*	50,836 bp
Total coding sequences	13,440 bp	Total coding sequences	13,440 bp	Total coding sequences	13,176 bp
Total codon count	4,480	Total codon count	4,480	Toal codon count	4341

All codons	G+C	29.2%	All codons	G+C	29.2%	All codons	G+C	30.53%
1st position	G+C	36.6%	1st position	G+C	36.5%	1st position	G+C	36.87%
2nd position	G+C	35.0%	2nd position	G+C	35.1%	2nd position	G+C	34.98%
3rd position	G+C	16.0%	3rd position	G+C	16.1%	3rd position	G+C	19.75%

Codon	AA	Fraction	Number	Codon	AA	Fraction	Number	Codon	AA	Fraction	Number
GCA	A	0.277	74	GCA	A	0.277	74	GCC	A	0.134	38
GCC	A	0.135	36	GCC	A	0.135	36	GCG	A	0.042	12
GCG	A	0.041	11	GCG	A	0.045	12	GCT	A	0.475	135
GCT	A	0.547	146	GCT	A	0.543	145	TGC	C	0.262	11
TGC	C	0.132	5		C	0.1	4	TGT	C	0.738	31
TGT	C	0.868	33	TGT	C	0.9	36	GAC	D	0.203	24
GAC	D	0.193	23	GAC	D	0.19	22	GAT	D	0.797	94
GAT	D	0.807	96	GAT	D	0.81	94	GAA	E	0.716	78
GAA	E	0.836	93	GAA	E	0.835	91	GAG	E	0.284	31
GAG	E	0.164	18	GAG	E	0.165	18	TTC	F	0.169	64
TTC	F	0.139	53	TTC	F	0.149	57	TTT	F	0.831	315
TTT	F	0.861	329	TTT	F	0.851	326	GGA	G	0.268	74
GGA	G	0.299	88	GGA	G	0.29	86	GGC	G	0.091	25
GGC	G	0.048	14	GGC	G	0.051	15	GGG	G	0.105	29
GGG	G	0.085	25	GGG	G	0.084	25	GGT	G	0.536	148
GGT	G	0.568	168	GGT	G	0.576	171	CAC	H	0.25	22
CAC	H	0.195	16	CAC	H	0.167	14	CAT	H	0.75	66
CAT	H	0.805	66	CAT	H	0.833	70	ATA	I	0.607	292
ATA	I	0.585	309	ATA	I	0.582	306	ATC	I	0.064	31
ATC	I	0.059	31	ATC	I	0.067	35	ATT	I	0.328	158
ATT	I	0.356	188	ATT	I	0.352	185	AAA	K	0.726	98
AAA	K	0.839	94	AAA	K	0.841	95	AAG	K	0.274	37
AAG	K	0.161	18	AAG	K	0.159	18	CTA	L	0.114	77
CTA	L	0.144	95	CTA	L	0.145	96	CTC	L	0.01	7
CTC	L	0.015	10	CTC	L	0.014	9	CTG	L	0.034	23
CTG	L	0.024	16	CTG	L	0.029	19	CTT	L	0.132	89
CTT	L	0.124	82	CTT	L	0.12	80	TTA	L	0.599	404
TTA	L	0.621	411	TTA	L	0.622	413	TTG	L	0.111	75
TTG	L	0.073	48	TTG	L	0.071	47	ATG	M	1	115
ATG	M	1	112	ATG	M	1	111	AAC	N	0.313	57
AAC	N	0.2	36	AAC	N	0.236	43	AAT	N	0.687	125
AAT	N	0.8	144	AAT	N	0.764	139	CCA	P	0.228	37
CCA	P	0.255	40	CCA	P	0.253	40	CCC	P	0.111	18
CCC	P	0.083	13	CCC	P	0.076	12	CCG	P	0.049	8
CCG	P	0.057	9	CCG	P	0.057	9	CCT	P	0.611	99
CCT	P	0.605	95	CCT	P	0.614	97	CAA	Q	0.814	70
CAA	Q	0.871	81	CAA	Q	0.862	81	CAG	Q	0.186	16
CAG	Q	0.129	12	CAG	Q	0.138	13	AGA	R	0.714	65
AGA	R	0.726	69	AGA	R	0.75	72	AGG	R	0.099	9
AGG	R	0.116	11	AGG	R	0.094	9	CGA	R	0.077	7
CGA	R	0.042	4	CGA	R	0.052	5	CGC	R	0.011	1
CGC	R	0	0	CGC	R	0	0	CGG	R	0.011	1
CGG	R	0.011	1	CGG	R	0.01	1	CGT	R	0.088	8
CGT	R	0.105	10	CGT	R	0.094	9	AGC	S	0.102	41
AGC	S	0.086	35	AGC	S	0.086	35	AGT	S	0.279	112
AGT	S	0.287	117	AGT	S	0.291	118	TCA	S	0.2	80
TCA	S	0.226	92	TCA	S	0.234	95	TCC	S	0.072	29
TCC	S	0.054	22	TCC	S	0.052	21	TCG	S	0.037	15
TCG	S	0.025	10	TCG	S	0.022	9	TCT	S	0.309	124
TCT	S	0.322	131	TCT	S	0.315	128	ACA	T	0.445	105
ACA	T	0.423	105	ACA	T	0.417	103	ACC	T	0.076	18
ACC	T	0.069	17	ACC	T	0.061	15	ACG	T	0.081	19
ACG	T	0.06	15	ACG	T	0.057	14	ACT	T	0.398	94
ACT	T	0.448	111	ACT	T	0.466	115	GTA	V	0.399	120
GTA	V	0.407	121	GTA	V	0.408	119	GTC	V	0.027	8
GTC	V	0.051	15	GTC	V	0.048	14	GTG	V	0.116	35
GTG	V	0.101	30	GTG	V	0.103	30	GTT	V	0.458	138
GTT	V	0.441	131	GTT	V	0.442	129	TGG	W	1	7
TGG	W	1	9	TGG	W	1	9	TAC	Y	0.23	47
TAC	Y	0.2	44	TAC	Y	0.195	43	TAT	Y	0.77	157
TAT	Y	0.8	176	TAT	Y	0.805	178	TGA	W	0.794	54
TGA	W	0.803	53	TGA	W	0.8	52	subtotal codon count			4327
subtotal codon count			4467	subtotal codon count			4467	TAA	*	0.147	10
TAA	*	0.167	11	TAA	*	0.169	11	TAG	*	0.059	4
TAG	*	0.03	2	TAG	*	0.031	2				4341
			4480				4480				

Table 5. Comparison of Direct Repeats

First Sequence	Second Sequence	Third sequence	Matches	Matches %	GC %	Location of first sequence	Location of second sequence	Comment
	***P. membranacea* repeats**							
1517-1717	47421-47260		186/203	92%	32%	Intergenic region *nad6* - trnK	Intergenic region *cox1 -cytb*	
1686-1767	54399-54480		78/82	95%	46%	Intergenic region *nad6* - trnK	Intergenic region *cox2 -rnpB*	Partial overlap with first repeat
5788-5954	7369-7534		155/168	92%	37%	Group IC1 intron GIY YIG region in *rnl*	Group IC1 intron region of *rnl*	
9130-9253	10502-10622		123/125	98%	37%	*rnl*	*rps3* contains intron in *rnl*	98% identity between the species
35358-35408	35409-35458	35459-35508		94-100%	47%	TAR2	TAR2	Tandem repeats
	***P. malacea* repeats**							
303-377	368-442		67/76	88%	34%	Part of the trnY gene	Part of the trnN gene	Partly overlap repeats present in trn genes
1885-2008	56749-56877		117/129	91%	42%	Intergenic region *nad6*- trnK	Intergenic region *cox2-rnpB*	
7161-7361	7847-8036		166/205	81%	39%	Group IC1 intron region of *rnl*	Group IC1 intron GIY YIG region in *rnl*	
10189-10310	11566-11687		122/122	100%	38%	*rnl*	rps3 contains intron in *rnl*	98% identity between the species
	***C. grayi* repeats**							
20552-20872	47905-48209		278/321	87%		Intergenic region between nad3 & nad4L	Intergenic region between atp9 & atp6	
12629-12910	34004-34285		249/285	87%		Between rnl and trnT	Intergenic region between nad1 & nad4, part of the sequence overlap with nad4 gene	
29932-30103	47029-47200		146/173	84%		Intergenic region between orf1 & nad1	Intergenic region between trnR & atp9	
40294-40379	40335-40420		80/86	93%		Intergenic region between rps3 & cytb		Repeat partly overlaps
20242-20320	50724-50796		69/79	87%		Intergenic region between nad3 & nad4L	Intergenic region between atp8 & rns	
3009-3061	3231-3283		49/54	91%		Intergenic region between trnN & trnR	Intergenic region between trnN & trnR	
30645-30683	31083-31122		38/40	95%		Intergenic region between orf1 & nad1	Intergenic region between orf1 & nad1	

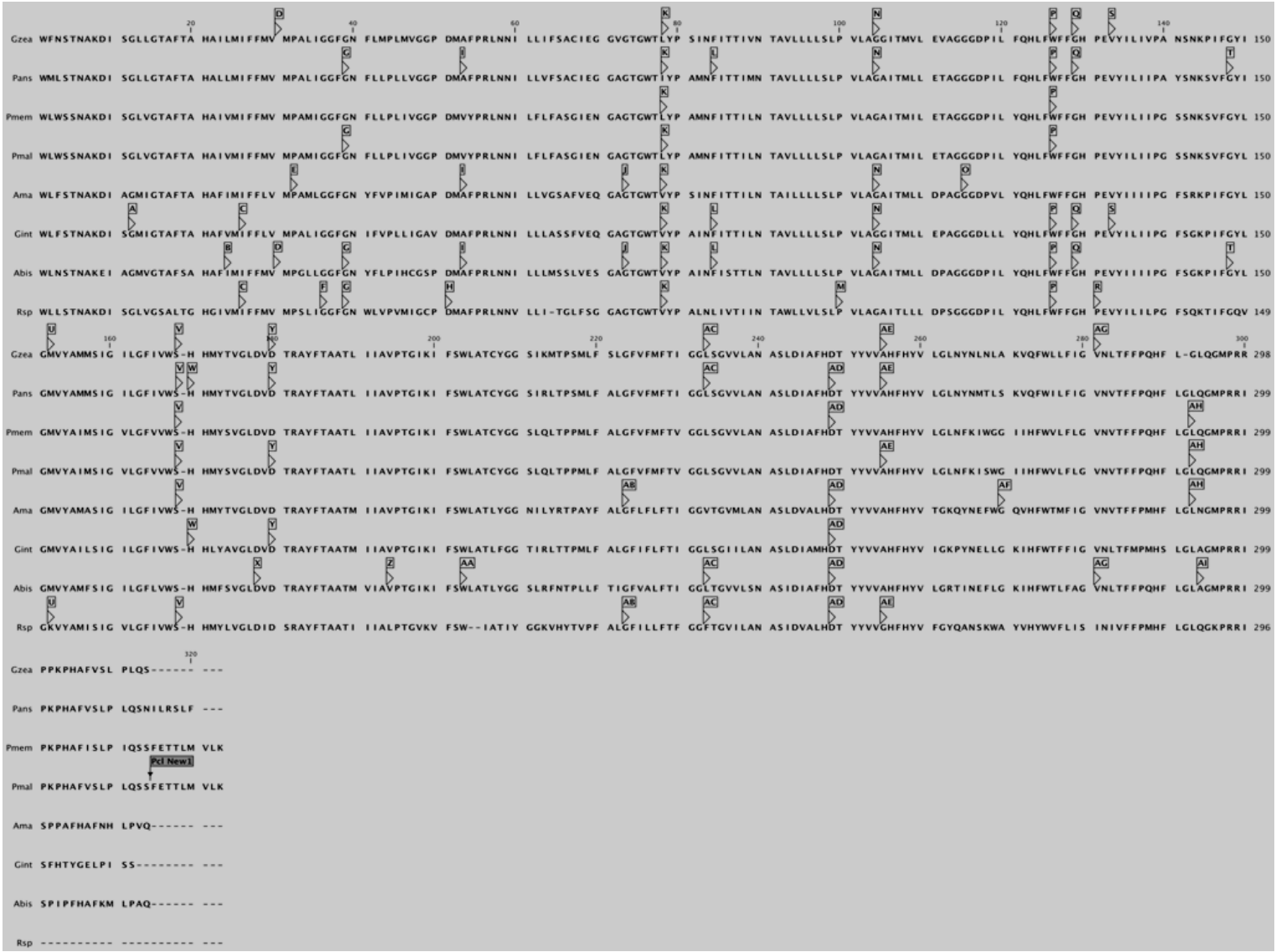


Figure 12 cox1 intron insertion class

Abstract

Mitochondrial genomes from the fungal partners of two terricolous foliose lichen symbioses, *Peltigera membranacea* and *Peltigera malacea*, have been determined using metagenomic approaches, including RNA-seq. The roughly 63 kb genomes show all the major features found in other *Pezizomycotina*, such as unidirectional transcription, 14 conserved protein genes, genes for the two subunit rRNAs and for a set of 26 tRNAs used in translating the 62 amino acid codons. In one of the tRNAs a CAU anticodon is proposed to be modified, via the action of the nuclear-encoded enzyme, tRNA Ile lysidine synthase, so that it recognizes the codon AUA (Ile) instead of AUG (Met). The overall arrangements and sequences of the two circular genomes are similar, the major difference being the inversion and deterioration of a gene encoding a type B DNA polymerase. Both genomes encode the RNA component of RNAse P, a feature seldom found in ascomycetes. The difference in genome size from the minimal ascomycete mitochondrial genomes is largely due to 17 and 20 Group I introns, respectively, most associated with homing endonucleases and all found within protein-coding genes and the gene encoding the large subunit rRNA. One new intron insertion point was found, and an unusually small exon of seven nucleotides was identified and verified by RNA sequencing. Comparative analysis of mitochondrion-encoded proteins places the *Peltigera* spp., representatives of the class *Lecanoromycetes*, close to *Leotiomycetes, Dothidiomycetes* and *Sordariomycetes*, in contrast to phylogenies found using nuclear genes.

1. INTRODUCTION

Mitochondrial (mt) DNA has been used extensively in evolutionary and population studies of all types of eukaryotes and is almost exclusively inherited in a uniparental manner (Galtier et al . 2009). Analysis of mt DNA has proven useful in determining population structures and phylogenies of fungi, first by using length polymorphisms of particular sections, later by sequencing sections between highly conserved PCR priming sites, and more recently by sequencing of whole mt genomes (Anderson & Kohn, 1998). Mt DNA accumulates mutations faster than nuclear DNA, making it more suitable for differentiating closely related organisms, and part of the mt *cox1* gene has been suggested as a universal "barcode" for rapid identification of eukaryotic organisms (Hebert et al . 2003). Information from mt DNA can be compared to that from nuclear DNA from the same fungal cells, and in the case of lichenized fungi (mycobionts) that are found in nature only as symbionts with green algae or cyanobacteria, both types of fungal phylogenies can be compared to those of their symbionts. Comparisons of mt genomes may be particularly valuable in understanding evolution and gene function in this large and diverse group of organisms: Lichen mycobionts, estimated to include up to 40% of all fungi (Nash, 2008) form deeply branching groups in this kingdom (Gargas et al . 1995; Lutzoni et al . 2001) with the greatest numbers of lichenized species within the ascomycete class *Lecanoromycetes*, in the subphylum *Pezizomycotina*. Currently there are mt genomes from over 90 fungal species in GenBank, mostly from ascomycetes and basidiomycetes. Although lichens have been studied using mt DNA markers (Nelsen et al . 2009; Werth, 2010), no lecanoromycete mt genome has been published yet.

The genus *Peltigera* includes about 90 species of mainly terricolous foliose lichens with widespread distribution. They have been the subject of many studies ranging from biogeography, evolution and taxonomy, ecophysiology to symbiosis and population biology. In addition to the mycobiont, all *Peltigera* have a cyanobacterial symbiont (*Nostoc* sp.) and some species also have a chlorophyte alga as a major photobiont; in such cases the *Nostoc* is confined to small spots (cephalodia). Besides photosynthesis, the *Nostoc* symbiont can carry out nitrogen fixation and thus *Peltigera* lichens channel both carbon and nitrogen into the ecosystem.

Here we present an analysis and comparative study of the fungal mt genomes from *Peltigera membranacea* and *Peltigera malacea* of the family *Peltigeraceae* in the class *Lecanoromycetes*. *P. membranacea* is ubiquitous in mixed birchwood communities, the climax community of prevalent lowland lavafields and glaciofluvial plains in Iceland. For comparison and stronger inference of common results we used a geographically distant sample of another species, *P. malacea,* which has similar major characters, but is known not to be phylogenetically close to *P. membranacea* within the genus (Miadlikowska & Lutzoni, 2004). Since *Peltigera* mycobionts have proven refractory to cultivation attempts (E. Stocker and M. Grube, pers. comm), the mt genomes were assembled from data generated by massive parallel sequencing of natural metagenomic isolates represented by whole thalli from field collections. Transcriptomics (RNA-seq) data were also used in addition to the DNA sequence data for annotation and expression analysis.

These first complete mt genomes from lichens, including coding sequences for 15 conserved proteins, now makes it possible to apply more complete molecular information to questions such as the phylogeny of the *Pezizomycotina*, which has proven difficult to resolve (James et al . 2006). These mt genome sequences and further delineation of the component genomes of the symbionts, including that of associated microbiota, the *Nostoc* photobiont, and mycobiont nucleus, will provide a natural genetic framework for extending research on these lichen species.

2. METHODS AND MATERIALS

2.1. Source of lichens and DNA extraction

P. membranacea (accession XBB013, LA-31632 (IINH)) was collected at Keldnagil, Reykjavik, Iceland within a 12 meter span (coordinates 41^{o} 64.7' N, 21^{o} 46.6' W). *P. malacea* (accession DB3992, UBC*)* was collected near Little Fort, British Columbia (51° 37.5' N, 120° 9' W). Samples were extensively washed with water before DNA extraction as previously described (Sinnemann et al . 2000). Genomic DNA (>50 kb) was processed for sequencing via the Roche 454 and the Illumina/Solexa methodology. Approximately 1.76 and 1.64 gigabases of 454 data, and 1.4 and 0.3 gigabases of Illumina data, were obtained for *P. membranacea* and *P. malacea*, respectively.

2.2. Mt genome assembly & annotation, nuclear genes

After an initial Newbler assembly of the 454 reads, contigs having high homology with fungal mt genomes and high coverage were sorted out and their component reads reassembled using the MIRA assembler (Chevreux, 1999). Subsequent assembly and annotation were performed using CLC Genomics Workbench (CLC bio). Illumina reads were used to resolve uncertainties due to homopolymer runs in 454 reads. The average fold coverage with 454 reads was 267 and 372, and with Illumina reads 47 and 5, for *P. membranacea* and *P. malacea*, respectively.

The genome assemblies were validated in two ways. Illumina mate-pair sequences with mean insert sizes of 1950 bp and 3600 bp were mapped onto the mt genome. The sequence pairs were distributed evenly throughout the genome and no discrepancies were found. The TAR2 region in *P. membranacea* and the *cox2-rnpB* regions in both mt genomes were verified by PCR and Sanger sequencing.

For annotation of the mt DNA, the BLAST suite of tools (Altschul et al . 1990) was used for searches in GenBank and comparisons to known fungal mt sequences and identification of repeats. RNAweasel (Lang et al . 2007) was used for identification of tRNA genes, Group I introns, *rns* and *rnpB* (including new annotations of *rnpB* and *rps3* in several published mt

genomes, see Table 1). The large subunit ribosomal RNA gene (*rnl*) was annotated based on homology to the closest relatives in GenBank and from RNA-seq reads in *P. membranacea* (below). Single nucleotide polymorphisms were called for the Illumina data with CLC Genomics Workbench using the standard quality filter and a minimum setting of 3 reads for each minor allele. Comparative physical maps of the two genomes were constructed using the CG View online server (Fig. 1) (Grant & Stothard, 2008).

The nuclear genes, *rpb1* and *tilS*, were found by BLAST searches of the draft assemblies of *P. membranacea* and *P. malacea* metagenomes (unpublished), and annotated using RNA-seq (below).

2.3. RNA analysis of *P. membranacea*

Total RNA was isolated from thalli, apothecia and rhizines using naturally moist, fresh material collected directly into RNAlater solution (Ambion) to inactivate RNAases. RNA was then extracted using the Totally RNA kit (Ambion) and processed for sequencing according to Illumina protocols. RNA sequences from different tissues and seasons were combined and mapped to the *P. membranacea* mt genome using CLC Genomics Workbench. RNA sequence mapping served two purposes: Confirmation of gene candidate annotation, and identification of extragenic, transcriptionally active regions (TAR), defined as any tract of >150 nucleotides (nt) with a minimum coverage of at least four reads/nt, at least 150 nt away from an annotated gene (Bruno et al . 2010). The minimum free energy of predicted RNA structures was calculated with the RNAfold program (Hofacker et al . 1994; McCaskill, 1990; Zuker & Stiegler, 1981).

2.4. Phylogenetic analysis

For phylogenetic analysis, protein sequences from *P. membranacea*, *P. malacea* and other fungi were aligned using ClustalX and ClustalW (Larkin et al . 2007; Thompson et al . 1994) as implemented in the CLC Genomics Workbench (CLC bio) and in MacVector 12.0.2 (MacVector Inc.). All alignments were verified manually and only unambiguous regions were considered for further analysis. A ModelGenerator (Keane et al . 2006) was used to determine the best fit model for maximum likelihood analysis. Maximum likelihood bootstrap support (Felsenstein, 1985) was calculated with 100 or 1000 replicates using PhyML3.0 (Guindon, 2003; Guindon et al . 2010). Maximum parsimony analysis (Swofford, 1993) was conducted in MEGA5 (Tamura et al . 2011). Bayesian analysis was performed using MRBayes 3.1 (Huelsenbeck & Ronquist, 2001).

3. RESULTS AND DISCUSSION

3.1. General features of the mt genomes

The mt genomes of *P. membranacea* and *P. malacea* are circular in topology and comprised of nearly the same number of basepairs, 62,785 and 63,363, respectively (Fig. 1). The genome sizes, low G+C content (~27%) and numbers of genes and open reading frames (ORF) are within the range commonly observed among fungi (Table 1). Each of the *Peltigera* mt genomes has 15 predicted ORFs for known protein coding genes: three for ATP synthase F0 subunits (encoded by *atp6*, *atp8* and *atp9*), three for cytochrome c oxidase subunits (*cox1*, *cox2* and *cox3*), seven for nicotinamide adenine dinucleotide ubiquinone oxidoreductase subunits (*nad1, nad2, nad3, nad4, nad4L, nad5, nad6*) and one each for cytochrome oxidase b (*cytb*), and ribosomal protein subunit 3 (*rps3*). The genomes share 98% identity in the coding regions, and 77% overall, with divergence associated primarily with intronic regions (Fig. 1). Both genomes have a full set of tRNA genes, as well as genes for the large (*rnl*) and small (*rns*) ribosomal subunit RNAs, the RNA component of mt RNAse P (*rnpB*), an ORF with no significant match in GenBank, and a gene or pseudogene for a B-type DNA polymerase (*dpoB*) (Table 1). As in most ascomycetes, the *Peltigera* mt genes are all oriented in the same direction. Since the genomic DNA was not clonal, considerable single nucleotide sequence polymorphism was detected $\geq$5% at 77 sites in *P. membranacea*, and 935 sites in *P. malacea*. These polymorphisms are the subject of a separate study. The mt genome accessions in GenBank are JN088165 for *P. membranacea* and JN088164 for *P. malacea*.

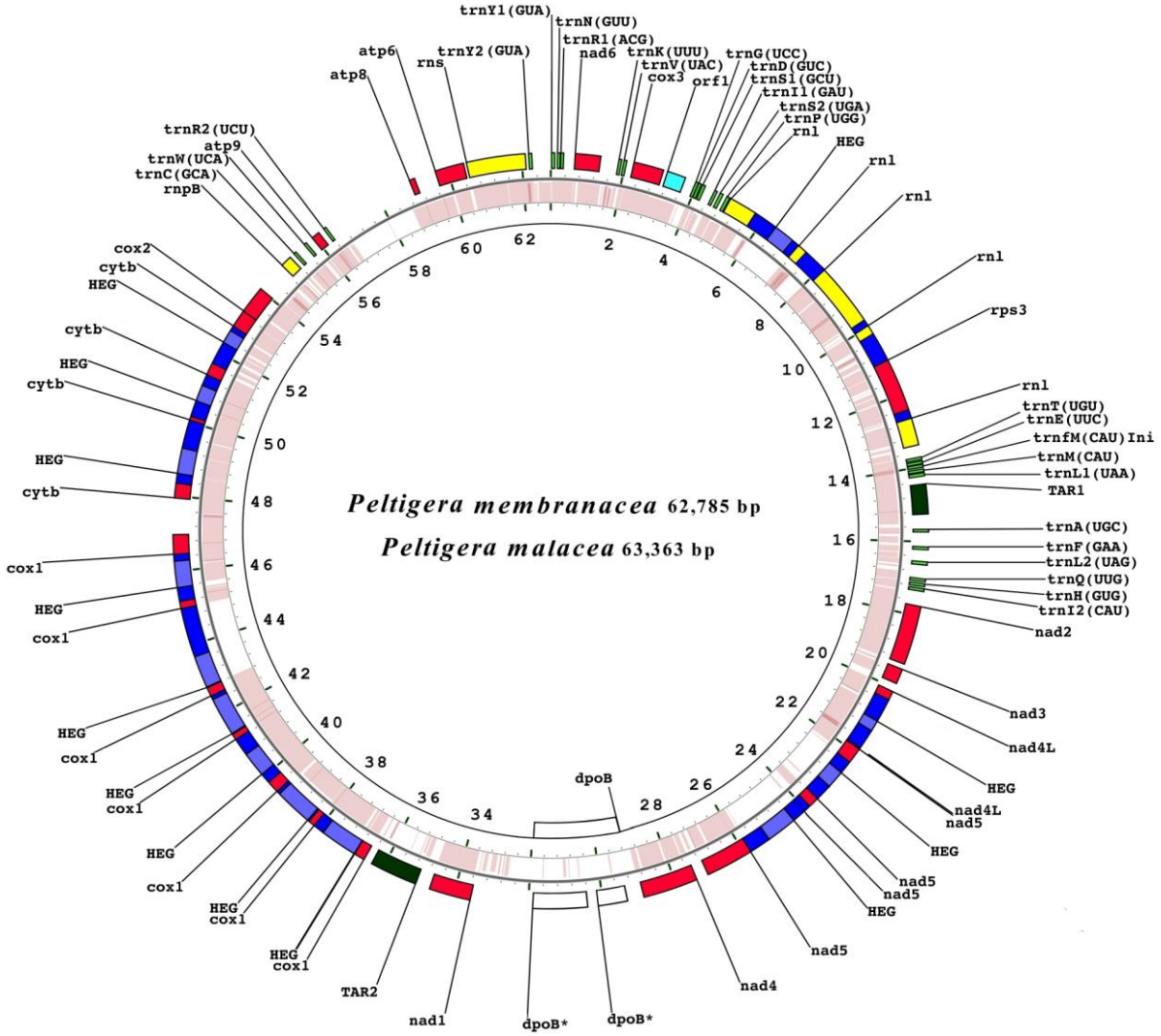


Figure 1. *Peltigera* mitochondrial genomes. Circular gene map representing the mt DNA of *P. membranacea*. All genes are transcribed clockwise (feature labels are affixed to the 5' end), *dpoB** is a pseudogene. Genes encoding proteins and ribosomal RNA are indicated by red and yellow boxes, respectively; tRNA genes, in green, are specified by standard one-letter amino acid code and anticodon. Introns and homing endonuclease genes (HEG) are shown in blue and light blue, and extragenic, transcriptionally active regions (TARs) and an unidentified ORF are in dark green and sky blue, respectively. The pink inner ring shows regions of identity with the genome of *P. malacea*; *dpoB* is only found in *P. malacea*. Numbers inside the map are in kb.

3.2. Codon usage

The mt genes of both *P. membranacea* and *P. malacea* contain 4480 codons (not counting *dpoB* and *rps3*). Codon usage is similar in both species and, consistent with the transmembrane localization of most mt proteins, codons for the nonpolar amino acids Leu, Phe, and Ile are the most abundant (Supplementary Table 1). As in other mt genomes, there is a strong bias for the use of A and U in synonymous codons: 84% of third positions are A or U, and 62% of the Leu codons are UUA. The genes *cytb, cox2* and *nad6* appear to use GUG, AUU and AUA, respectively, as initiation codons. Stipulation of an AUG codon in these cases would call for subsequent RNA editing, deletion of highly conserved amino acids, or overlap with preceding genes, resulting in a long N-terminal extension. These alternatives are not supported by transcript mapping (Fig. 2). GUG is well established as an alternative initiation codon in bacteria and the first position anticodon C in mt initiator tRNA appears to allow flexible pairing, permitting binding to AUA and AUU codons in addition to AUG (Bullerwell et al . 2003). All genes have a UAA stop codon, except for *cox2*, *cox3*, and *dpoB* which have UAG.

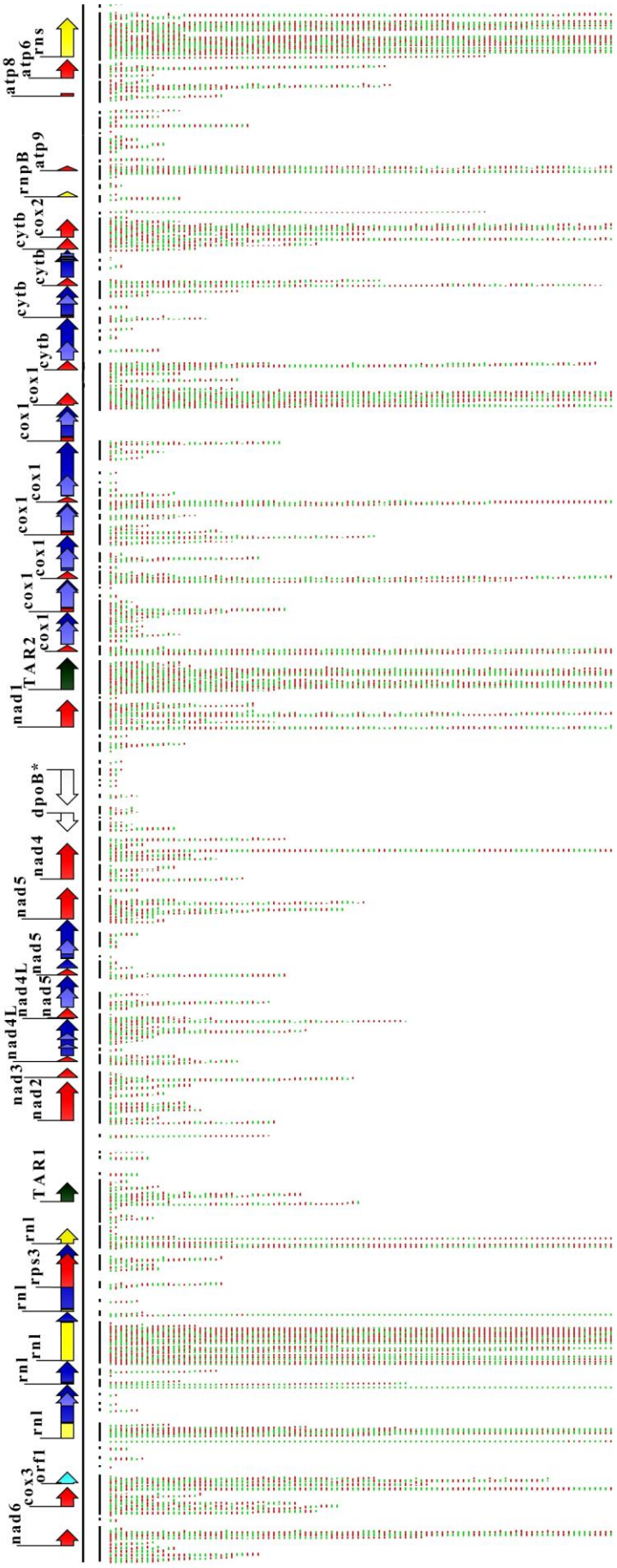


Figure 2. *P. membranacea* mitochondrial transcriptome. Mapped reads are shown below the annotated mt genome. Discontinuous line indicates RNA coverage.

3.3. tRNA genes

The *P. membranacea* and *P. malacea* mt genomes each encode 26 tRNAs that specify all 20 common amino acids. There are two distinct tRNAs for amino acids that have six codons (Arg, Leu, Ser), and two distinct tRNAs for Tyr in each species. Both *Peltigera* species also have three different tRNA genes with CAU anticodons, normally specifying Met: One encodes the elongator $tRNA^{Met}$, another the initiator $tRNA^{fMet}$, but the third encodes a $tRNA^{Ile}$. The *Peltigera* mt genome uses all three Ile codons (AUU, AUC and AUA), and a tRNA with a GAU anticodon that should pair with the Ile codons AUC and AUU by wobble pairing (Crick, 1966) was identified, but no tRNA gene with a UAU anticodon was detected for pairing to AUA (an unmodified version of this anticodon should also pair with the the Met codon AUG). The third tRNA gene with a CAU anticodon therefore appeared a likely candidate for the role of decoding AUA, as it conforms to the determinants of a $tRNA^{Ile}$ described in bacterial genomes (Silva et al . 2006), and presumably could be restricted to decoding AUA by a lysidine modification (Takemoto et al . 2009) as has been demonstrated in potato mitochondria (Weber et al . 1990). A similar situation seems to exist in other fungal mt genomes where there appears to be an excess of tRNA genes with CAU anticodons and none with UAU anticodons. Phylogenetic analysis places these four tRNA types on distinct branches (Fig. 3a). Consistent with this model, nuclear homologs of *tilS*, encoding tRNAIle lysidine synthetase necessary for the lysidine modification (Suzuki & Miyauchi, 2010), can be found in *P. membranacea* and *P. malacea,* as well as in representatives from major fungal groups (Fig. 3b; Table 1). Their deduced protein sequences have clear homology to bacterial TilS and conservation of key residues (Soma et al . 2003). In contrast to the mitochondrion, the cytoplasm of eukaryotes makes use of a $tRNA^{Ile}$ with inosine in the first position of the anticodon, allowing decoding of all three Ile codons.

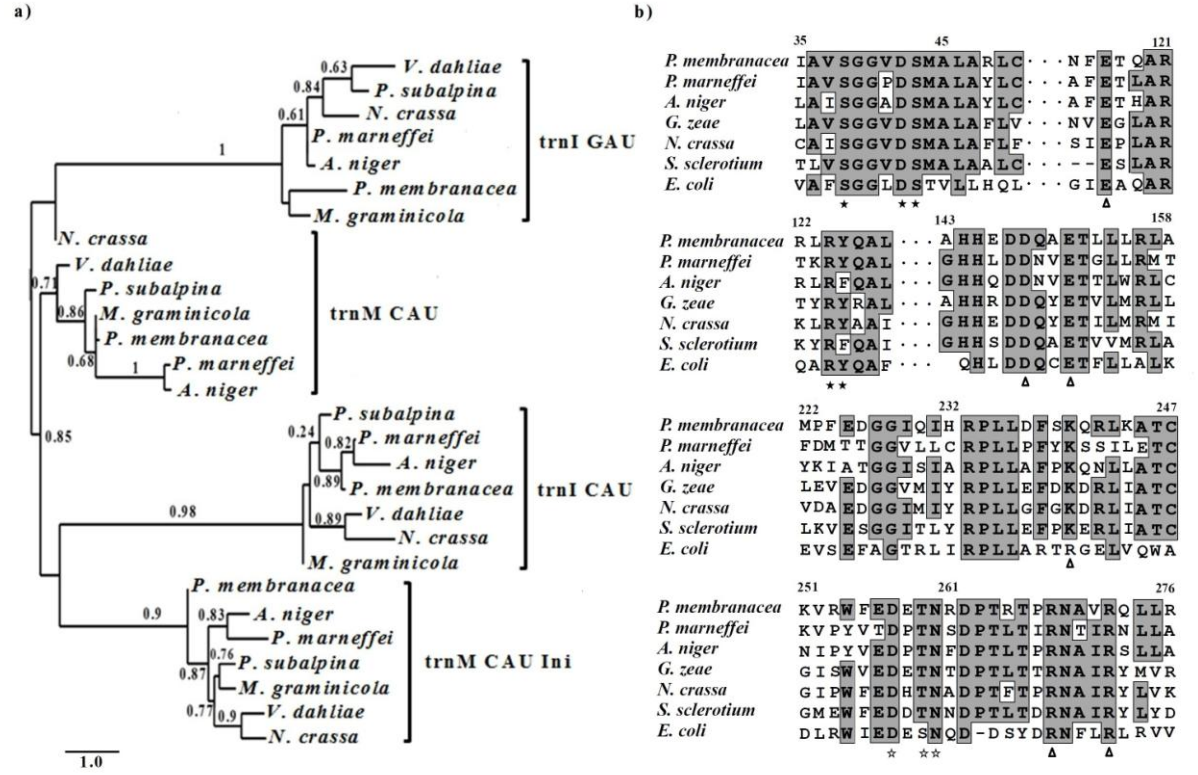
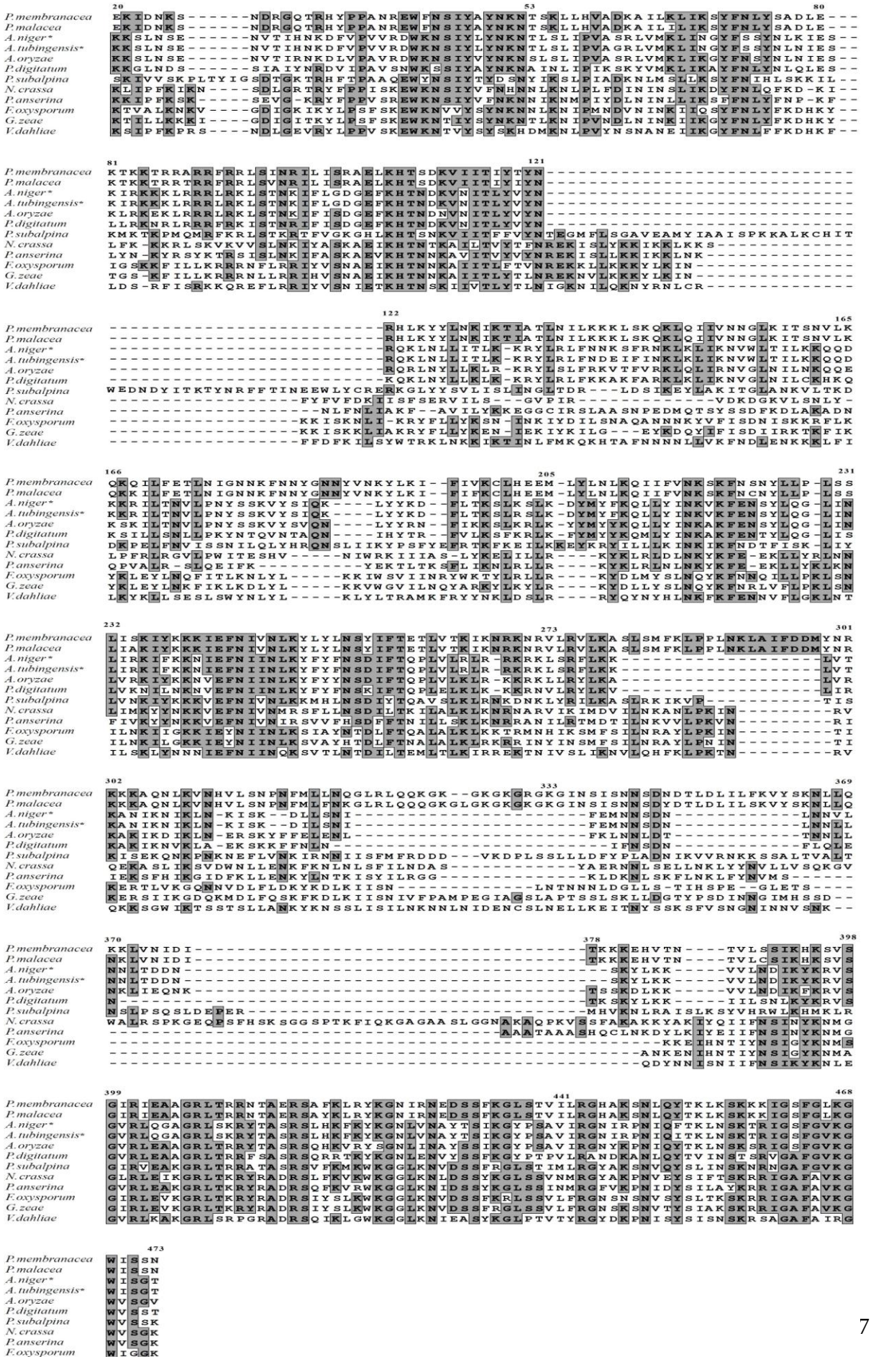


Figure 3. Fungal mt Met and Ile tRNAs and TilS. a) The phylogenetic tree of mt Met and Ile tRNAs was constructed using the maximum likelihood method implemented in the PhyML program (v3.0 aLRT) using the substitution model HKY85. Reliability of internal branching was assessed using the aLRT test and the likelihood ratios are shown above the branches (>0.90 indicates significant likelihood). These sequences marked as trnI CAU were originally annotated as trnM in GenBank. b) Selected regions from the N-terminal half of $tRNA^{Ile}$ lysidine synthetase containing recognition and catalytic functions were aligned. Conserved amino acids are highlighted and residues putatively involved in the recognition of ATP, tRNA, and L-lysine (Nakanishi et al 2005) are marked with filled stars, open stars and triangles respectively. For full names of fungi and accession numbers see Table 1. *E. coli* TilS accession: BAA77863.

3.4. Rps3

The *Peltigera* mt *rnl* genes contain a 2621 bp intron with an embedded ORF of 473 amino acids with conserved residues of the Rps3 protein (Supplementary Fig. 1). The group I intron splicing junctions (5'-CGCTAGGGAT/AACAGGCTAA-3') were verified by RNA sequencing and are identical to those previously reported among filamentous ascomycetes. Although it has not always been specifically annotated (e.g. *A. tubingensis*, *A. niger*, Table 1), the *rps3* gene appears to be common and strongly conserved in all major fungal groups (Bullerwell et al . 2000) including the *Pezizomycotina*, supporting the hypothesis that this arrangement is ancestral to the lineage. The placement of *rps3* within the *rnl* gene and their co-transcription has been suggested to enhance expression of *rps3* (Sethuraman et al . 2009), but as gauged by RNA sequencing, the transcript level of *rps3* in *P. membranacea* is lower than that of the other protein coding genes in the mitochondrion (Fig. 2).

Supplementary Figure 1(below). Rps3 protein alignment. Rps3 amino acid sequences from *P. membranacea, P. malacea* and 12 fungi from four classes were end-trimmed and aligned with ClustalX. Rps3 marked with an asterisk were missing in the published annotations. Conserved amino acids are highlighted. For full names and accession numbers see Table 1.


P.membranacea
P.malacea
A.niger*
A.tubingensis*
A.oryzae
P.digitatum
P.subalpina
N.crassa
P.anserina
F.oxysporum
G.zeae
V.dahliae

3.5. Respiratory chain genes *(atp, cyt, cox* and *nad)*

The majority of the proteins encoded by mt genes have a role in the respiratory chain. ATP synthase is a large protein complex that transforms the energy of the electromotive force into high energy bonds in ATP. Most of its subunits are encoded by nuclear genes, but several subunits are usually encoded by the mitochondria of fungi and other eukaryotes lineages, e.g. *atp9* is present in most fungal mitochondria but absent from *P. anserina* (Cummings et al . 1990). The *atp9* genes, as well as the atp6 and atp8 genes, are present in both *Peltigera* spp. RNA sequencing shows that the *atp9* gene is highly expressed in *P. membranacea* (Fig. 2), with a 74 base 5´UTR and 88 base 3´UTR that presumably contribute to stability and translational efficiency (Chen & Dieckmann, 1997).

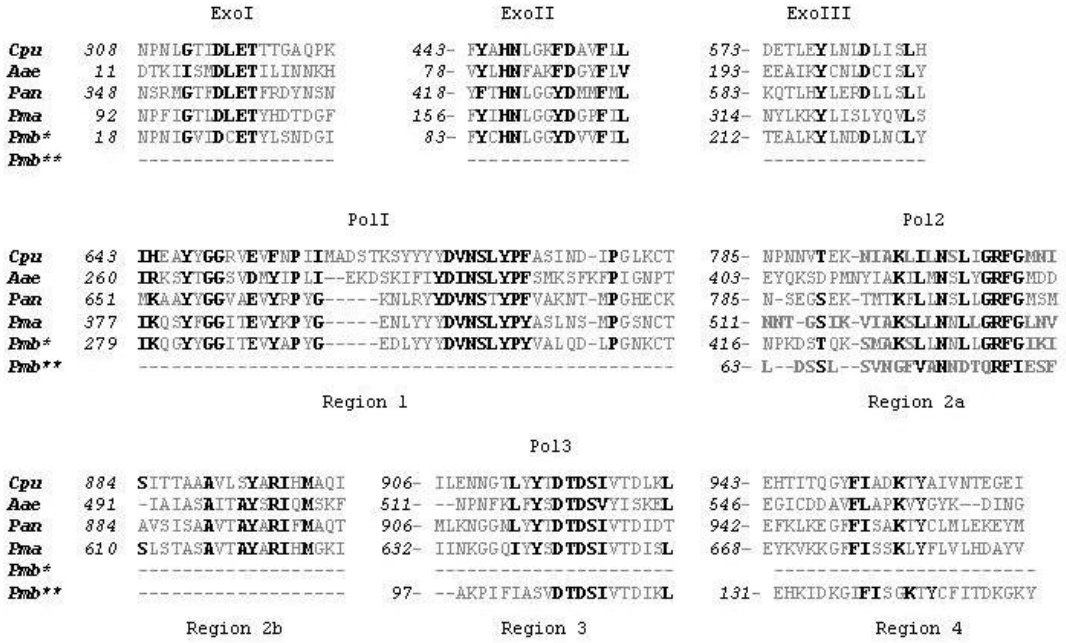
The *cox1, cox2* and *cox3* genes encode three subunits of cytochrome c oxidase ("complex IV"), the last enzyme in the respiratory electron transport chain of mitochondria, while *cytb* encodes a subunit of respiratory chain complex III (Yang et al . 2011). The most conserved genes, *cox1* and *cytb,* are often riddled with introns in ascomycetes (Paquin et al . 1997). In *Peltigera*, the *cox* genes and *cytb* together host ≥50% of all of the mt introns, with *cox1* bearing the greatest number: in *P. malacea* only 16% of this gene is coding. Fungal Cox1 proteins vary in length from 460 to 710 residues (Table 1), but are generally in the range of 525-550. The predicted Cox1 protein of *P. membranacea* is 631 residues. The larger than average size arises from a 3' extension that encodes the C-terminal 77 amino acid, and is consistent with transcriptomic data (Fig. 2). *P. malacea* similarly has an additional 73 amino acids. The extensions are 93% identical in amino acid sequence, indicating common ancestry, but neither sequence is conserved in other known Cox1 proteins. The extensions do not have as strong a codon usage bias as observed in other mt genes encoding proteins, possibly reflecting a recent addition. The longest extensions are found in *M. graminicola* (141 residues) and *P. brasiliensis* (156 residues) (Table1), but neither have similarity to each other or to *Peltigera*.

In contrast to the plethora of non-coding intervening sequences and Group I introns in the mtDNA that separate protein coding genes into many fragments, a gene fusion is proposed between *cytb* and *cox2.* No canonical stop codon is detected for *cytb,* no canonical start for *cox2,* and both coding sequences are in the same frame. These observations suggest the possibility of a fusion of the two genes, with the single stop codon at the end of *cox2* serving

both. A preponderance of fusion mRNAs (Fig. 2) among transcripts for this region supports this model in *P. membranacea*, although it is also consistent with the possibility of a single message with exceptionally long, untranslated extension, or a polycistronic message, as found in bacteria. Fusion proteins between c*ytb* and *cox2* have not been reported, but other fusion transcripts have been reported from the mt genomes of eukaryotic microbes. Analyses of proteins in *Acanthamoeba castellanii* revealed that a fusion transcript (*cox1* and *cox2*) gave rise to two products upon translation (Gawryluk & Gray, 2010; Lonergan & Gray, 1996), while the possibility of a new multidomain protein from a fusion transcript (*cytb* and *cox3*) in the dinoflagellate *Oxyrrhis marina* remains open (Slamovits et al . 2007).

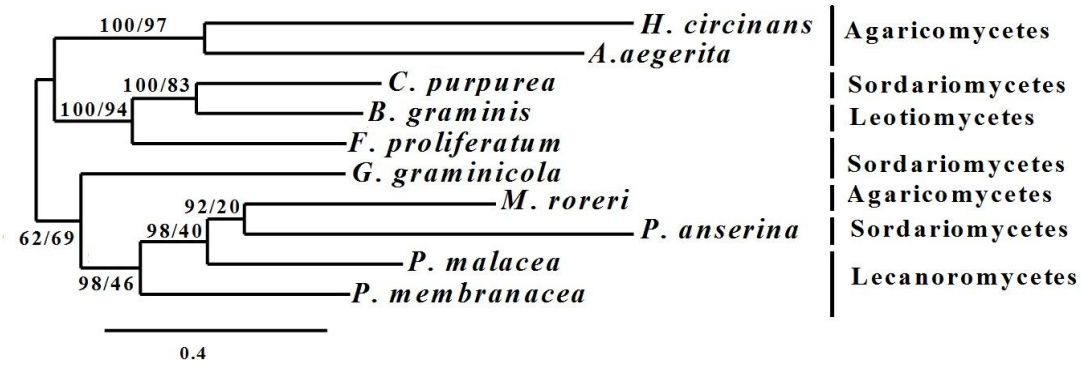
3.6. *dpoB* and *rnpB*

The *dpoB* gene, encoding a type B2 DNA polymerase (Dpo), is most often found in free linear plasmids that may populate fungal mitochondria (Griffiths, 1995), but integrated forms have occasionally been reported in mt genomes (Formighieri et al . 2008; Mouhamadou et al . 2004). Dpo includes an N-terminal domain of variable size that in some cases codes for the terminal protein for priming plasmid replication (Laday et al . 2008; Rohe et al . 1992), a 3'-5' exonuclease domain (with three conserved regions ExoI, ExoII and ExoIII) and a C-terminal polymerase domain (with a series of conserved “Pol” regions) (Blanco et al . 1991; Rohe et al . 1992). Integrated orthologs of *dpoB* were observed at the same locus, between *nad4* and *nad1*, in both *Peltigera* spp. In *P. malacea,* the 697 amino acid predicted Dpo contains all three conserved regions of the exonuclease domain and most of the polymerase domain (Pol1, Pol2a, Pol2b, Pol3, and Pol4 regions) (Supplementary Fig. 2).



Supplementary Figure 2. DpoB alignment. Conserved exonuclease domains marked as ExoI, ExoII, and ExoIII. Polymerase domains marked as Pol1, Pol2, and Pol3. Aae = *Agrocybe aegerita* (AAC33727), Cpu = *Claviceps purpurea* (P22373), Pan = *Podospora anserina* (S26947). Pma =*Peltigera malacea*, Pmb = *P. membranacea*. Pmb* shows the initial part of Dpo, Pmb** contains the second part. Conserved regions according to Rohe et al., 1992 or Blanco et al., 1991 shown above and below the alignment, respectively.

The translation ends shortly after the Pol4 region, much like that of the integrated Aa-*polB* gene of *Agrocybe aegerita* (Bois et al . 1999), without a clear Pol5 region that may be more typical of other classes of Dpo proteins. In *P. membranacea, dpoB* is represented as a pseudogene (*dpoB**) on the opposite strand. The main 1461 nt section of *dpoB** is disrupted by three stop codons. Its hypothetical translation (reading through the stops) includes the three Exo motifs, Pol1 and Pol2a regions. No relict of the Pol2b region is evident in the adjacent sequence, but an ORF immediately downstream on the same strand contains another fragment of *dpoB**, with a 262-residue predicted translation that includes Pol3 and Pol4 regions. There are no clear footprints (e.g. repeats) in the regions surrounding the *dpoB* genes to indicate recent plasmid integration events, as have been deduced in other fungi (Formighieri et al . 2008). Analyses of geographically well-separated samples of *P. malacea* and *P. membranacea* by PCR mapping indicate that the rearrangements in this region of the mt DNA are species-specific (unpublished), even if their origins are unclear. Interestingly, the closest known relative of the *Peltigera dpoB* gene resides in a basidiomycete, *Moniliophthora roreri*, (55% identity to GenBank accession YP004347432, and Supplementary Fig. 3). This suggests an incidence of horizontal gene transfer in a common ancestor, followed by rearrangement in one lineage during or after divergence of the two *Peltigera* spp.



Supplementary Figure 3. Mitochondrial DpoB phylogeny. Amino acids of the clearly conserved initial part of the *P. membranacea dpoB* pseudogene and the *P. malacea dpoB* gene were aligned to corresponding sequences from eight fungi. Branch support value, in percent, indicates Bayesian posterior probability, followed by maximum likelihood bootstrap value if different. Branches are drawn to scale, with the bar indicating 0.4 substitutions per site. *Agrocybe aegerita* (AAC33727), *Hebeloma circinans* (CAA72280), *Moniliophthora roreri* (YP_004347432), *Glomerella graminicola* (EFQ24845), *Blumeria graminis f. sp. hordei* (AY189817.1), *Claviceps purpurea* (P22373), *Fusarium proliferatum (*YP_001718360), *Podospora anserina* (S26947).

Mitochondrially-encoded *rnpB* genes are irregularly distributed among fungi (Seif et al . 2003) and not always annotated (e.g. *M. graminicola and A. nidulans*, Table 1). They have been reported in a few filamentous and yeast-form ascomycetes, zygomycetes (Seif et al . 2005) and most recently, possibly in the mycorrhizal symbiont, *Glomus intraradices* (Lee & Young, 2009). The *rnpB* genes in *Peltigera* are AT-rich, similar in size (75.2% and 75.8% AT, and 245 and 249 nucleotides, respectively), and carry sequences corresponding to conserved regions CR1 and CRV (Chen & Pace, 1997). They are flanked downstream by a tRNA gene as in other fungi, but are flanked upstream not by tRNA genes, as is typical, but by *cox2* (Fig. 1), an arrangement described previously only in *Schizosaccharomyces fibuligera* (Seif et al . 2003).

3.7. Group I introns and homing endonucleases

Group I introns are mobile genetic elements expressing self-splicing RNAs (ribozymes) varying in size from 142 bp to >3000 bp. These introns interrupt a wide range of organelle and nuclear rDNA genes in fungi, algae and many other unicellular eukaryotes. Group I introns have five main subgroups and several subdivisions based on their RNA secondary structure and can direct their own insertion at specific targets (4-6 nt) (Haugen et al . 2005; Lang et al . 2007). The mt introns are useful markers for population and evolution studies as they target conserved sequences (Mullineux et al . 2010) but are less subject to strong selection and therefore rapidly accumulate mutations (degenerate) or may be eliminated by precise deletion.

Homing endonucleases similarly direct enzyme-mediated insertion of a DNA sequence into a target site of 15-45 nt (Haugen et al . 2004). Genes encoding homing endonucleases (HEGs) of two types, LAGLIDADG and GIY-YIG, are often strongly associated with Group I introns (Haugen et al . 2005). It is still unclear how they co-evolve, but it has been argued that integration of HEGs into introns promotes the propagation of both the Group I introns and the HEGs (Haugen et al . 2005; Lambowitz et al . 1999).

In *Peltigera* species, as in other fungi, the majority of mt Group I introns carry HEGs, most often for the LAGLIDADG type enzymes. Of the 17 Group I introns in *P. membranacea*, nine have LAGLIDADG sequences (two degenerate), four have GIY-YIG sequences and four, including the intron in *rnl* that carries *rps3* (discussed above), lack HEGs. Similarly, of

the 20 Group I introns in *P. malacea*, 12 have LAGLIDADG sequences, five have GIY-YIG (two of each type degenerate), and three, including the one carrying *rps3*, lack HEGs. In *P. membranacea*, RNAs corresponding to many of the HEGs can clearly be detected, but other intronic RNAs are generally found at lower levels (Fig. 2).

The *cox1* gene is the most common target for insertion by Group I introns but the number of integrated elements varies widely, (*e.g.* from none in *V. dahliae* to 14 in *P. anserina* among the ascomycetes, Table 1). Because insertion of the LAGLIDADG HEGs into this gene has been observed in diverse taxonomic groups, the positions of insertion can be used to examine the evolutionary dynamics of intron mobility (Férandon et al . 2010). There are 6 and 9 introns with HEGs in the *cox1* genes of *P. membranacea* and *P. malacea*, respectively, at a total of ten insertion sites (Fig. 4) of which all but one, "AJ" (designated in this study), are considered widely distributed (Férandon et al . 2010).

Presence of a *Peltigera* intron at a site also occupied by LAGLIDADG HEG orthologs in other ascomycetes and basidiomycetes suggests that the intron-containing state at that position is likely ancestral, e.g. orthologous HEGs from both *Peltigera* spp. and other fungi occur at site "K", suggesting it is ancestral among dikaryomycota (Supplementary Fig. 4).

Thus, absence of introns at sites "G", "Y" and "AE" in *P. membranacea* is interpreted as a loss from this species, rather than as an independent gain in *P. malacea*; a loss from *P. malacea* is likewise inferred at site "AD". Insertion at the site "AJ", after the S554 codon of the *cox1* gene in *P. malacea* has not been reported before and is not found in *P. membranacea*. In this case, rather than a loss, the intron is predicted to be a recent gain, possibly of a particularly active HEG-intron combination. Putative orthologs of the HEG at "AJ" in *P. malacea* occur not only at a variety of other locations within *cox1*, e.g. intron 7 (39% identity, 60% similarity) at position "U" in *G. zeae* (YP_001249330) or *orf 441* in intron 5 (43% identity, 63% similarity) at "AK", another novel position in *P. anserina* (YP_004376358.1) (see also Fig. 4), but also within other mt genes, e.g. *P. anserina nad3* intronic *orf403* (NP_074914.1) and *M. roreri nad5* intron 1 (YP_004376358.1). The overall impression remains, however, that intron loss is more common in the *P. membranacea* lineage.

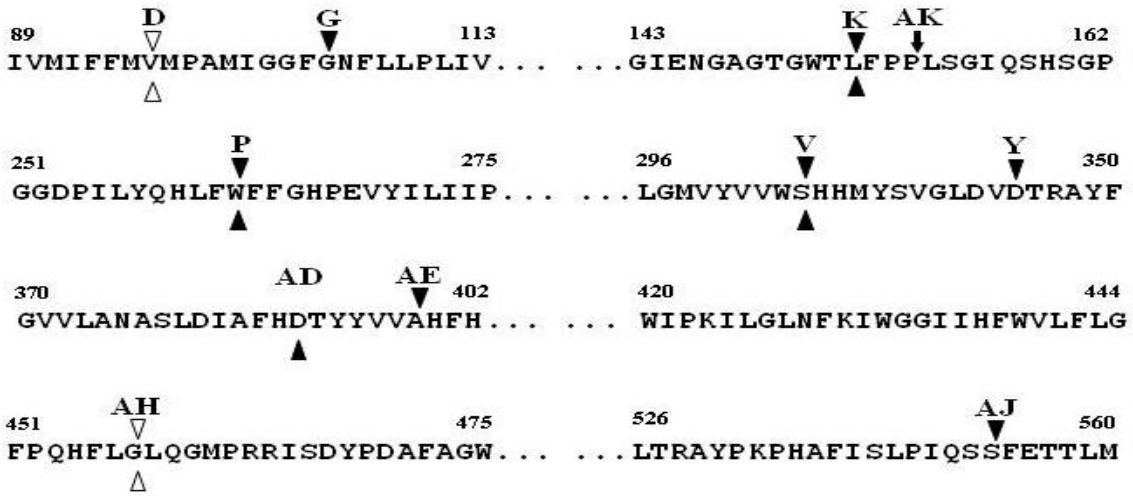
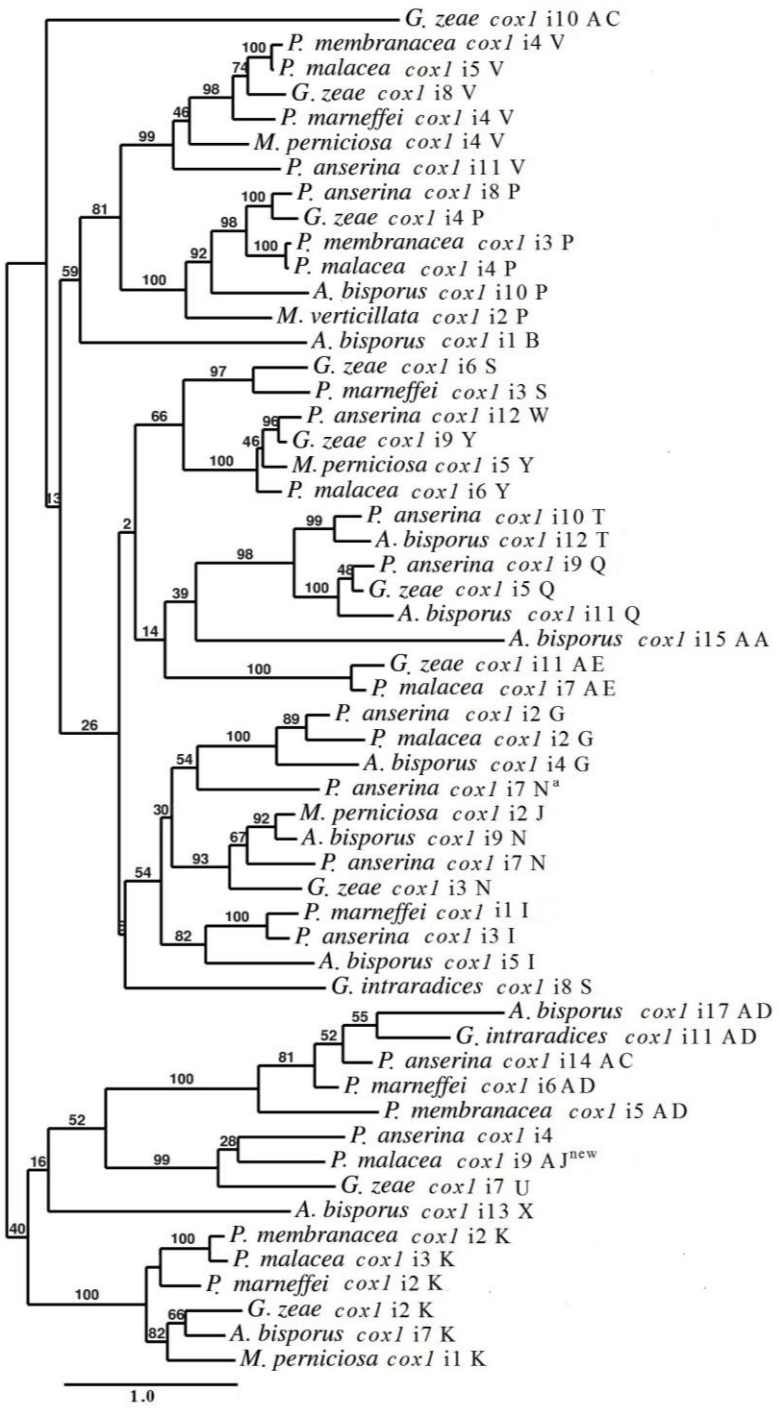


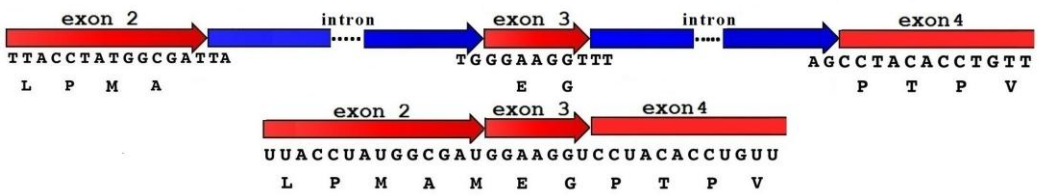
Figure 4. Group I introns in *cox1*. Insertion sites for introns carrying a LAGLIDADG HEG (black triangles) or a GIY-YIG HEG (white triangle) in *P. malacea* and *P. membranacea* (above and below the sequence, respectively) are shown relative to the derived Cox1 amino acid sequence of *P. membranacea*. Reference amino acid positions are numbered above each tract, and insertion sites are labeled as in Férandon et al. 2010. Insertion site AJ is novel. Insertion site AK (black arrow) is novel and not found in the *Peltigera* mt genomes, but the equivalent position in *P. anserina* is used by a presumptive ortholog of the intron at AJ in *P. malacea*.



Supplementary Figure 4. Phylogenetic analysis of LAGLIDADG homing endonucleases in *cox1*. LAGLIDADG homing endonuclease protein sequences were aligned using ClustalX and a maximum likelihood tree derived using PhyML 3.0 with 100 bootstrap replicates. Branches are drawn to scale, with the bar indicating 1.0 amino acid substitution per site. Labels show species and gene, number shows intron position in the gene. Insertion position class is indicated alphabetically as described in Ferandon et al 2010. N[a] indicates a second endonuclease gene at the same insertion point. Except for *Agaricus bisporus* (EU314927), full names and accession numbers are in Table 1.

3.8. Short exons

The largest protein-coding gene in *P. membranacea*, *nad5,* is interrupted by three introns that fragment the gene into four parts. The second and third introns are so close that an exon of only seven nucleotides is predicted (Supplementary Fig. 5). Small exons are not uncommon; in *Peltigera*, exons of 28 bp and 33 bp are also observed in *nad4L* and *cox1* respectively. Nonetheless, small exons, such as the 3 bp exons in *cox1* in *P. anserina* (Cummings et al . 1989) or *Agaricus bisporus* (Férandon et al . 2010), should be annotated with caution and ideally be supported by transcriptomic data, as alternative splicing options may also be possible. The 7 bp exon in *P. membranacea nad5*, in a highly conserved region and with the predicted splicing verified by RNA-seq data, shows that conservation of only a few nucleotides is necessary for splicing, as has been inferred from structural analysis (Adams et al . 2004).



Supplementary Figure 5. ***P. membranacea nad5*** **7 nt exon.** Nucleotide sequences flanking exon-intron junctions of introns 2 and 3 with corresponding RNA sequence and conserved amino acid sequence below.

3.9. Non-coding regions: Repeats and TARs

Intergenic regions or non-coding regions of the mtDNA vary greatly among species, but may play important roles in gene and genome function. The intergenic regions amount to 30% and 27% in *P. membranacea* and *P. malacea* respectively and are rich in simple direct, inverted and tandem repeats which can mediate genome evolution, *e.g.* by errors in recombination producing duplications, inversions and deletions (Alexander et al . 2010; Stone et al . 2010).

In *P. membranacea* and *P. malacea* there are 5 and 4 pairs of direct, perfect or near-perfect repeats from 76 bp to 205 bp (Supplementary Table 2). Two of these pairs are at equivalent positions in the two mt genomes. The first pair is ~124 bp in both species, and is in an *rnl* intron, as well as 1250 bp downstream in the highly conserved *rnl* sequence. The partners of the other pair, 82 bp in *P. membranacea* and 129 bp in *P. malacea*, are in the intergenic region between *nad6* and *trnK*, and 10 kb away in the intergenic region between *cox2* and *rnpB*. Although the repeat pairs (other than in *rnl*) have >80% identity within each species, there is no significant homology between the species. Thus there appears to be selection for maintaining these repeats, but not their sequence *per se*. Repeats can be found in many other ascomycete mt genomes but they do not appear to be a generally conserved feature (unpublished observations).

It has recently been shown that fungal mitochondria produce a variety of RNA molecules derived from intergenic regions, and that the concentration of these RNAs is similar to that of the protein coding RNAs (Bruno et al . 2010). Two transcriptionally active regions (TARs) were detected by RNA-seq analysis (Fig. 2). Each produces prominent RNAs that are predicted to form energetically stable structures similar to those in prokaryotes (Weinberg et al . 2009). The 800 nt TAR1 RNA (ΔG= -193 kcal/mole) region appears highly conserved (91% identity) between the two *Peltigera* spp., but the 1284 nt TAR2 RNA (Δ= -392 kcal/mole) region (see Appendix TAR figure), with a triple ~50 nt tandem repeat embedded, does not have a close homolog in *P. malacea*.

3.10. Phylogeny

The two *Peltigera* mt genomes are identical in the number and order of genes, except for the inversion of *dpoB*, and differ little in their coding sequences (Fig. 1). The arrangement of mt genes in the two *Peltigera* spp. was compared to that of representatives with the same basic gene set from four fungal phyla (Fig. 5). Approximately the same number of rearrangements separates the *Peltigera* mt genomes from those of other *Pezizomycotina* classes, with more differences observed as the species are further apart in the phylogenetic tree, as observed in other studies (Pantou et al . 2008). Two features common to the *Pezizomycotina* mt genomes, *rps3* embedded in an intron within *rnl* and a one base overlap of the *nad4L* and *nad5* reading frames are also found in other phyla and may therefore be ancestral. A phylogenetic analysis of 14 concatenated mt protein sequences from the two *Peltigera* spp. and 12 representative fungal species (Table 1) using maximum likelihood, maximum parsimony and Bayesian methods returned the same well-supported tree topology with robust bootstrap and posterior probability values (Fig. 6a). The same tree topology was obtained using individual proteins Atp6, Nad5 and Cox1 (data not shown). The genus *Peltigera* is a coherent group according to analysis of nuclear ribosomal RNA genes (Miadlikowska & Lutzoni, 2004) and the two representative mt genomes described here show this group separating deeply from those of other published sequences. Using the estimate of approximately 200 myr for the separation of *Eurotiomycetes* and *Sordariomycetes* (Taylor & Berbee, 2006), and an even rate of amino acid substitution, the two *Peltigera* species may have diverged roughly 10 myrs ago.

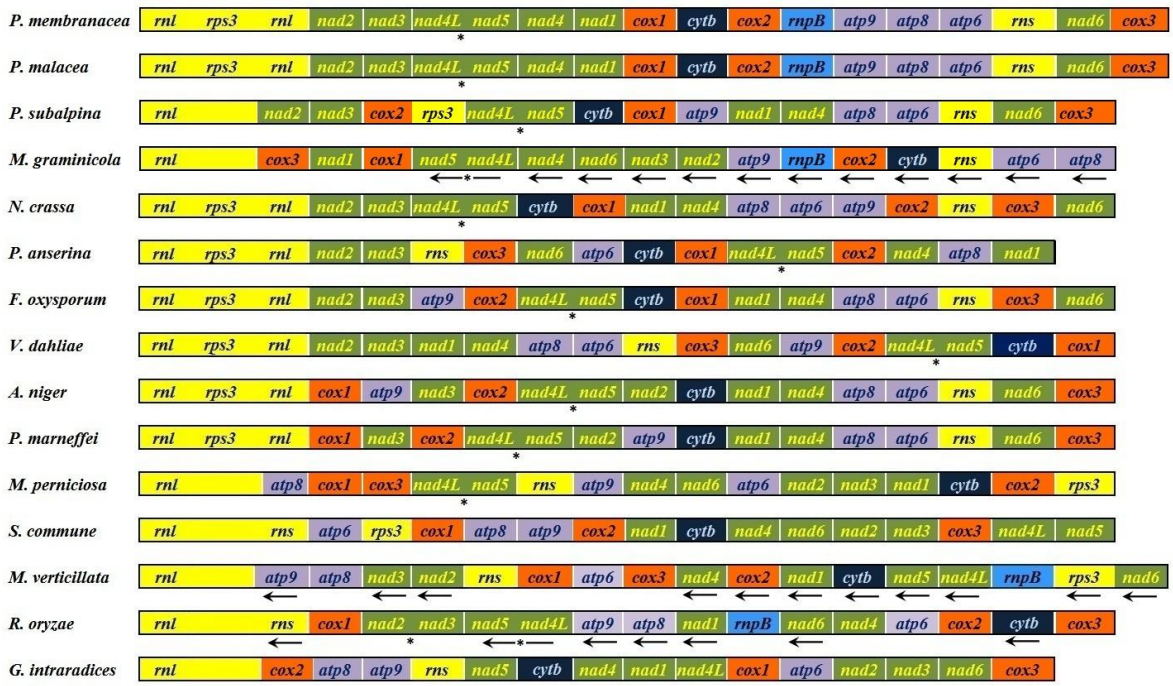


Figure 5. Mitochondrial gene order in *Peltigera* spp. and other fungi. Schematic maps show ribosomal RNA, *rnpB* and protein coding genes. The *rps3* gene is found either separate or within an intron in *rnl*. All genes are transcribed left to right except as indicated by arrows. One base overlaps in *nad4L-nad5* and *nad2-nad3* are indicated with an asterisk. For full names and accession numbers see Table 1.

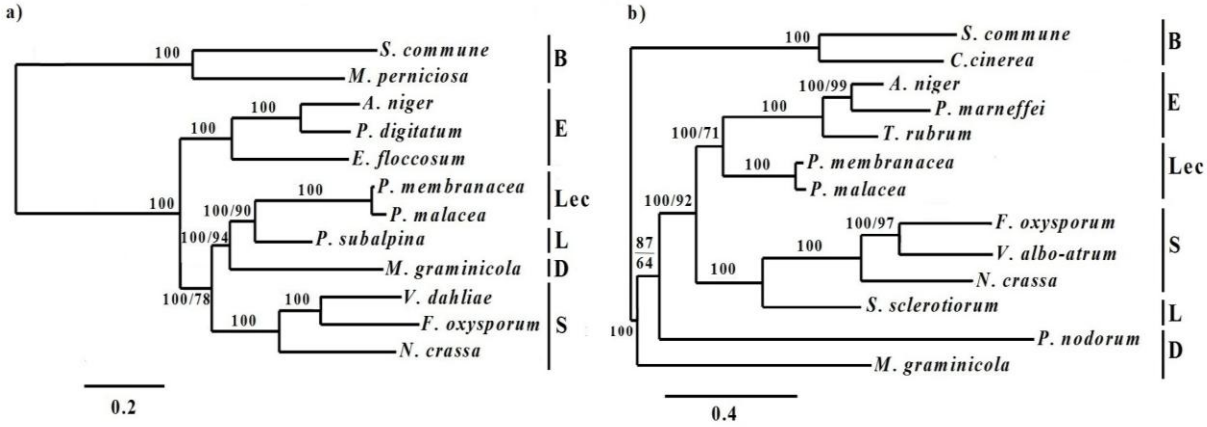
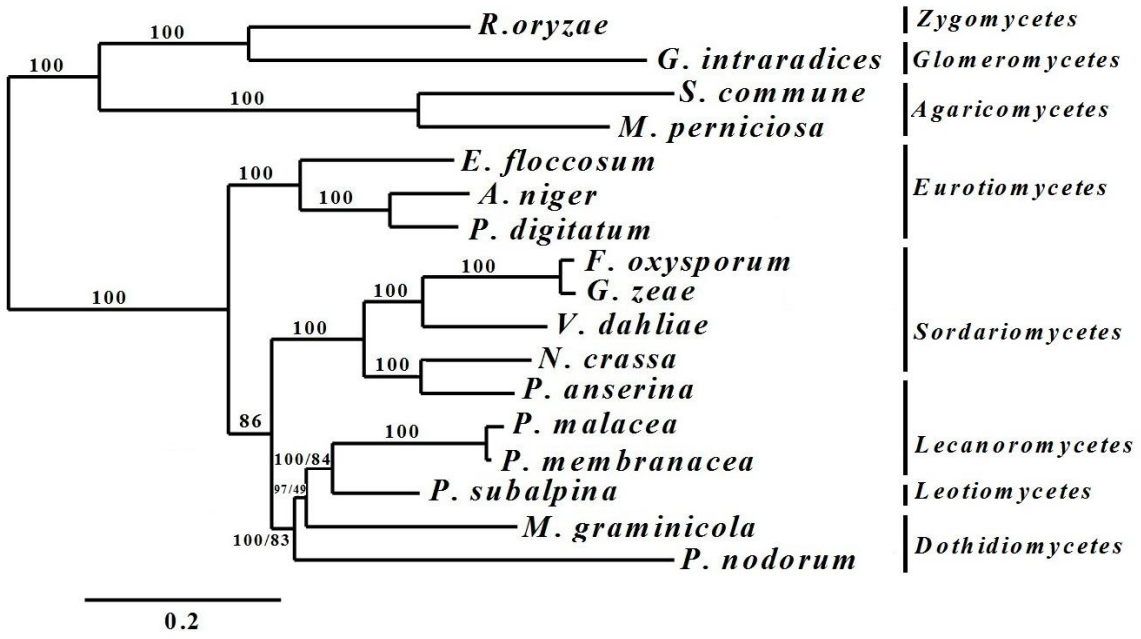


Figure 6. Phylogenetic trees based on mitochondrion encoded proteins and Rpb1, a nuclear encoded protein. a) Fourteen proteins encoded by the mitochondria of *P. membranacea* and *P. malacea* were concatenated and aligned to corresponding sequences from representatives of other classes of fungi (Table 1). Branch support value, in percent, indicates Bayesian posterior probability, followed by maximum likelihood bootstrap value if different. Branches are drawn to scale, with the bar indicating 0.2 substitutions per site. b) The complete protein sequences of Rpb1 from the species in a) or their closest relatives available in Genbank, plus *P. nodorum* (Table 1), were aligned and used for building a maximum likelihood tree. Branch support values as in a) with the bar indicating 0.4 substitutions per site. B = *Basidiomycotina*; E = *Eurotiomycetes*; Lec = *Lecanoromycetes*; L = *Leotiomycetes*; D = *Dothidiomycetes*; S = *Sordariomycetes*. For full names and accession numbers see Table 1.

The analysis based on 14 mt proteins, as well as another analysis based on the 12 mt proteins common to a larger set of taxa (Supplementary Fig. 6), show a clear divergence of *Lecanoromycetes* from *Eurotiomycetes*, in contrast to studies based on other nuclear and mt markers in which *Lecanoromycetes* appear as a sister group to *Eurotiomycetes* (James et al . 2006; Miadlikowska & Lutzoni, 2004; Nelsen et al . 2009). Furthermore, *Lecanoromycetes* appears closer to *Leotiomycetes* and *Dothidiomycetes* than to *Sordariomycetes*. The mt protein results partly agree with a more extensive phylogeny of *Pezizomycotina* made with a set of five nuclear genes (Spatafora et al . 2006), but differ in the placement of the single representatives from *Dothidiomycetes* and *Leotiomycetes*. When the full sequence of the nuclear encoded RNA polymerase large subunit (Rpb1) is used for analysis, a well-supported tree topology (Fig. 6b) similar to that previously found with nuclear encoded proteins is generated (James et al . 2006). This incongruence, found using maximum likelihood, maximum parsimony and Bayesian methods, can either reflect biological processes such as horizontal gene transfer and incomplete lineage sorting or tree-building errors (Galtier & Daubin 2008; Wiens et al . 2010), and may be resolved in the future with more rigorous analyses making use of rapidly accumulating genome resources.



Supplementary Figure 6. Multigene phylogeny with 17 species. Twelve proteins (except *atp8* and *atp9*) encoded by the mitochondria of *P. membranacea* and *P. malacea* were concatenated and aligned to corresponding sequences from 15 representatives of major fungal clades (Table 1). Branch support value, in percent, indicates Bayesian posterior probability, followed by maximum likelihood bootstrap value if different. Branches are drawn to scale, with the bar indicating 0.2 substitutions per site.

4. Conclusion

Metagenomic and next generation sequencing platforms are increasingly practical tools for a variety of detailed investigations of microorganisms in natural ecosystems, including lichen symbionts that are generally recalcitrant to pure culture *in vitro* (Bates et al . 2011; Magain et al . 2010; Schneider et al . 2011). The mt genomes of the mycobionts of *P. membranacea* and *P. malacea*, selectively assembled after respective whole thallus DNA sequencing and annotated with support of transcriptomic data from thalli of *P. membranacea*, were similar to those of related non-symbiotic fungi, and differed from each other primarily in their population of Group I introns and the condition of a plasmid-derived *dpoB* gene. These results suggest that there are few options for specific adaptations to a symbiotic state of existence at the level of organellar genome structure in terms of the basic suite of genes, but that the mt genomes of lichenized fungi are nonetheless dynamic in other respects. The mt genome of *P. membranacea* seems to have been more effective at eliminating invasive mobile elements: intron loss is more common in this lineage, and the integrated *dpoB* sequence has been eroded. The *Peltigera* HEG at "AJ" as well as the DpoB sequences show considerable similarity to equivalent elements in the basidiomycete, *M. roreri*, raising questions about horizontal movement of genes between ascomycete and basidiomycete mitochondria that potentially complicate phylogenetic analysis (Brown, 2003; Marcet-Houben & Gabaldón, 2010). Concatenated whole mt genome protein sequences were used to establish a well-supported phylogeny within the *Pezizomycotina*, but a contrasting phylogeny was also constructed from a conserved nuclear protein. It will be interesting to determine, with the growing number of fungal mt genomes available through various sequencing projects, whether this disparity reflects fundamental differences in the evolution of mt and nuclear protein coding genes.

Acknowledgements This research was funded by the Icelandic Research Fund. The authors are grateful to Dr. B.F. Lang for reviewing the *rnpB* data and for valuable comments. We also thank deCODE Genetics and Droplaug N. Magnúsdóttir, Guðbjörg Þ. Örlygsdóttir, Steinunn Snorradóttir and Ólafur Þ. Magnússon for technical support.

REFERENCES

1. Adams PL, Stahley MR, Gill ML, Kosek AB, Wang J, Strobel SA, 2004. Crystal structure of a group I intron splicing intermediate. *RNA* **10:** 1867-1887.

2. Alexander RP, Fang G, Rozowsky J, Snyder M, Gerstein MB, 2010. Annotating non-coding regions of the genome. *Nature Reviews Genetics* **11:** 559-571.

3. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ, 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215:** 403-410.

4. Anderson JB, Kohn LM, 1998. Genotyping, gene genealogies and genomics bring fungal population genetics above ground. *Trends in Ecology and Evolution* **13:** 444-449.

5. Bates ST, Cropsey GWG, Caporaso JG, Knight R, Fierer N, 2011. Bacterial communities associated with the lichen symbiosis. *Applied and Environmental Microbiology* **77:** 1309-1314.

6. Blanco L, Bernad A, Blasco MA, Salas M, 1991. A general structure for DNA-dependent DNA-polymerases. *Gene* **100:** 27-38.

7. Bois F, Barroso G, Gonzalez P, Labarere J, 1999. Molecular cloning, sequence and expression of Aa-polB, a mitochondrial gene encoding a family B DNA polymerase from the edible basidiomycete *Agrocybe aegerita*. *Molecular and General Genetics* **261:** 508-513.

8. Brown, 2003. Ancient horizontal gene transfer. *Nature Reviews Genetics* **4:** 121.

9. Bruno VM, Wang Z, Marjani SL, Euskirchen GM, Martin J, Sherlock G, Snyder M, 2010. Comprehensive annotation of the transcriptome of the human fungal pathogen *Candida albicans* using RNA-seq. *Genome Research* **20:** 1451-1458.

10. Bullerwell CE, Burger G, Lang BF, 2000. A novel motif for identifying Rps3 homologs in fungal mitochondrial genomes. *Trends in Biochemical Sciences* **25:** 363-365.

11. Bullerwell CE, Leigh J, Seif E, Longcore JE, Lang BF, 2003. Evolution of the fungi and their mitochondrial genomes. In: K A Dilip, and G K George (eds), *Applied Mycology and Biotechnology*, Elsevier, Amsterdam, pp.133-159.

12. Chen JL, Pace NR, 1997. Identification of the universally conserved core of ribonuclease P RNA. *RNA* **3:** 557-560.

13. Chen W, Dieckmann C, 1997. Genetic evidence for interaction between Cbp1 and specific nucleotides in the 5' untranslated region of mitochondrial cytochrome b mRNA in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology* **17:** 6203-6211.

14. Chevreux TP, Wetter T, Suhai S, 1999. Genome sequence assembly using trace signals and additional sequence information. In: Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB) (Germany) pp. 45-56.

15. Crick FH, 1966. Codon-anticodon pairing: the wobble hypothesis. *Journal of Molecular Biology* **19:** 548-555.

16. Cummings DJ, McNally KL, Domenico JM, Matsuura ET, 1990. The complete DNA sequence of the mitochondrial genome of *Podospora anserina*. *Current Genetics* **17:** 375-402.

17. Cummings DJ, Michel F, McNall KL, 1989. DNA sequence analysis of the 24.5 kilobase pair cytochrome oxidase subunit I mitochondrial gene from *Podospora anserina*: a gene with sixteen introns. *Current Genetics* **16:** 381-406.

18. Felsenstein J, 1985. Confidence limits on phylogenies - an approach using the bootstrap. *Evolution* **39:** 783-791.

19. Férandon C, Moukha S, Callac P, Benedetto J-P, Castroviejo M, Barroso G, 2010. The *Agaricus bisporus* cox1 gene: the longest mitochondrial gene and the largest reservoir of mitochondrial group I introns. *PLoS One* **5:** e14048.

20. Formighieri EF, Tiburcio RA, Armas ED, Medrano FJ, Shimo H, Carels N, Goes-Neto A, Cotomacci C, Carazzolle MF, Sardinha-Pinto N, Thomazella DP, Rincones J,

Digiampietri L, Carraro DM, Azeredo-Espin AM, Reis SF, Deckmann AC, Gramacho K, Goncalves MS, Moura Neto JP, Barbosa LV, Meinhardt LW, Cascardo JC, Pereira GA, 2008. The mitochondrial genome of the phytopathogenic basidiomycete *Moniliophthora perniciosa* is 109 kb in size and contains a stable integrated plasmid. *Mycological Research* **112:** 1136-1152.

21. Galtier N, Daubin V, 2008. Dealing with incongruence in phylogenomic analyses. *Philosophical Transactions: Biological Sciences* **363:** 4023-4030.

22. Galtier N, Nabholz B, Glemin S, Hurst GDD, 2009. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular Ecology* **18:** 4541-4550.

23. Gargas A, DePriest P, Grube M, Tehler A, 1995. Multiple origins of lichen symbioses in fungi suggested by SSU rDNA phylogeny. *Science* **268:** 1492-1495.

24. Gawryluk RM, Gray MW, 2010. An ancient fission of mitochondrial Cox1. *Molecular Biology and Evolution* **27:** 7-10.

25. Grant JR, Stothard P, 2008. The CGView Server: a comparative genomics tool for circular genomes. *Nucleic Acids Research* **36:** W181-184.

26. Griffiths, 1995. Natural plasmids of filamentous fungi. *Microbiology and Molecular Biology Reviews* **59:** 673-685.

27. Guindon, 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* **52:** 696-704.

28. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O, 2010. New algorithms and methods to estimate maximum likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology* **59:** 307-321.

29. Haugen P, Reeb V, Lutzoni F, Bhattacharya D, 2004. The evolution of homing endonuclease genes and group I introns in nuclear rDNA. *Molecular Biology and Evolution* **21:** 129-140.

30. Haugen P, Simon DM, Bhattacharya D, 2005. The natural history of group I introns. *Trends in Genetics* **21:** 111-119.

31. Hebert PDN, Cywinska A, Ball SL, deWaard JR, 2003. Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London Series B: Biological Sciences* **270:** 313-321.

32. Hofacker IL, Fontana W, Stadler PF, Bonhoeffer LS, Tacker M, Schuster P, 1994. Fast folding and comparison of RNA secondary structures. *Monatshefte für Chemie / Chemical Monthly* **125:** 167-188.

33. Huelsenbeck JP, Ronquist F, 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17:** 754-755.

34. James TY, Kauff F, Schoch CL, Matheny PB, Hofstetter V, Cox CJ, Celio G, Gueidan C, Fraker E, Miadlikowska J, Lumbsch HT, Rauhut A, Reeb V, Arnold AE, Amtoft A, Stajich JE, Hosaka K, Sung G-H, Johnson D, O/'Rourke B, Crockett M, Binder M, Curtis JM, Slot JC, Wang Z, Wilson AW, Schuszler A, Longcore JE, O/'Donnell K, Mozley-Standridge S, Porter D, Letcher PM, Powell MJ, Taylor JW, White MM, Griffith GW, Davies DR, Humber RA, Morton JB, Sugiyama J, Rossman AY, Rogers JD, Pfister DH, Hewitt D, Hansen K, Hambleton S, Shoemaker RA, Kohlmeyer J, Volkmann-Kohlmeyer B, Spotts RA, Serdani M, Crous PW, Hughes KW, Matsuura K, Langer E, Langer G, Untereiner WA, Lucking R, Budel B, Geiser DM, Aptroot A, Diederich P, Schmitt I, Schultz M, Yahr R, Hibbett DS, Lutzoni F, McLaughlin DJ, Spatafora JW, Vilgalys R, 2006. Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature* **443:** 818-822.

35. Keane T, Creevey C, Pentony M, Naughton T, McInerney J, 2006. Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evolutionary Biology* **6:** 29.

36. Laday M, Stubnya V, Hamari Z, Hornok L, 2008. Characterization of a new mitochondrial plasmid from *Fusarium proliferatum*. *Plasmid* **59:** 127-133.

37. Lambowitz AM, Caprara MG, Zimmerly S, Perlman PS, 1999. Group I and group II ribozymes as RNPs: clues to the past and guides to the future. In: R.F. Gesteland, T.R.

Cech, and J.F. Atkins (eds), *The RNA World*, 2nd edn. Cold Spring Harbor Laboratory Press, New York, pp.451-485.

38. Lang BF, Laforest MJ, Burger G, 2007. Mitochondrial introns: a critical view. *Trends in Genetics* **23:** 119-125.

39. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG, 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* **23:** 2947-2948.

40. Lee J, Young JP, 2009. The mitochondrial genome sequence of the arbuscular mycorrhizal fungus *Glomus intraradices* isolate 494 and implications for the phylogenetic placement of *Glomus*. *New Phytologist* **183:** 200-211.

41. Lonergan KM, Gray MW, 1996. Expression of a continuous open reading frame encoding subunits 1 and 2 of cytochrome c oxidase in the mitochondrial DNA of *Acanthamoeba castellanii*. *Journal of Molecular Biology* **257:** 1019-1030.

42. Lutzoni F, Pagel M, Reeb V, 2001. Major fungal lineages are derived from lichen symbiotic ancestors. *Nature* **411:** 937-940.

43. Magain N, Forrest LL, Sérusiaux E, Goffinet B, 2010. Microsatellite primers in the *Peltigera dolichorhiza* complex (lichenized ascomycete, *Peltigerales*). *American Journal of Botany* **97:** e102-e104.

44. Marcet-Houben M, Gabaldón T, 2010. Acquisition of prokaryotic genes by fungal genomes. *Trends in Genetics* **26:** 5-8.

45. McCaskill JS, 1990. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers* **29:** 1105-1119.

46. Miadlikowska J, Lutzoni F, 2004. Phylogenetic classification of peltigeralean fungi (*Peltigerales, Ascomycota*) based on ribosomal RNA small and large subunits. *American Journal of Botany* **91:** 449-464.

47. Mouhamadou B, Barroso G, Labarere J, 2004. Molecular evolution of a mitochondrial *polB* gene, encoding a family B DNA polymerase, towards the elimination from *Agrocybe* mitochondrial genomes. *Molecular Genetics and Genomics* **272:** 257-263.

48. Mullineux ST, Costa M, Bassi GS, Michel F, Hausner G, 2010. A group II intron encodes a functional LAGLIDADG homing endonuclease and self-splices under moderate temperature and ionic conditions. *RNA* **16:** 1818-1831.

49. Nakanishi K, Fukai S, Ikeuchi Y, Soma A, Sekine Y, Suzuki T, Nureki O, 2005. Structural basis for lysidine formation by ATP pyrophosphatase accompanied by a lysine-specific loop and a tRNA-recognition domain. *Proceedings of the National Academy of Sciences of the United States of America* **102:** 7487-7492.

50. Nash TH, 2008. Lichen Biology 2nd edn (Cambridge University Press).

51. Nelsen MP, Lücking R, Grube M, Mbatchou JS, Muggia L, Plata ER, Lumbsch HT, 2009. Unravelling the phylogenetic relationships of lichenised fungi in *Dothideomyceta*. *Studies in Mycology* **64:** 135-144-S134.

52. Pantou MP, Kouvelis VN, Typas MA, 2008. The complete mitochondrial genome of *Fusarium oxysporum*: Insights into fungal mitochondrial evolution. *Gene* **419:** 7-15.

53. Paquin B, Laforest M-J, Forget L, Roewer I, Wang Z, Longcore J, Lang BF, 1997. The fungal mitochondrial genome project: evolution of fungal mitochondrial genomes and their gene expression. *Current Genetics* **31:** 380-395.

54. Rohe M, Schrunder J, Tudzynski P, Meinhardt F, 1992. Phylogenetic relationships of linear, protein-primed replicating genomes. *Current Genetics* **21:** 173-176.

55. Schneider T, Schmid E, de Castro JV, Cardinale M, Eberl L, Grube M, Berg G, Riedel K, 2011. Structure and function of the symbiosis partners of the lung lichen (*Lobaria pulmonaria* L. Hoffm.) analyzed by metaproteomics. *Proteomics* **11:** 2752-2756.

56. Seif E, Leigh J, Liu Y, Roewer I, Forget L, Lang BF, 2005. Comparative mitochondrial genomics in zygomycetes: bacteria-like RNase P RNAs, mobile

elements and a close source of the group I intron invasion in angiosperms. *Nucleic Acids Research* **33:** 734-744.

57. Seif ER, Forget L, Martin NC, Lang BF, 2003. Mitochondrial RNase P RNAs in ascomycete fungi: Lineage-specific variations in RNA secondary structure. *RNA* **9:** 1073-1083.

58. Sethuraman J, Majer A, Friedrich NC, Edgell DR, Hausner G, 2009. Genes within genes: multiple LAGLIDADG homing endonucleases target the ribosomal protein S3 gene encoded within an rnl group I intron of *Ophiostoma* and related taxa. *Molecular Biology and Evolution* **26:** 2299-2315.

59. Silva FJ, Belda E, Talens SE, 2006. Differential annotation of tRNA genes with anticodon CAT in bacterial genomes. *Nucleic Acids Research* **34:** 6015-6022.

60. Sinnemann SJ, Andresson OS, Brown DW, Miao VP, 2000. Cloning and heterologous expression of *Solorina crocea pyrG*. *Current Genetics* **37:** 333-338.

61. Slamovits CH, Saidarriaga JF, Larocque A, Keeling PJ, 2007. The 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.

62. Soma A, Ikeuchi Y, Kanemasa S, Kobayashi K, Ogasawara N, Ote T, Kato J-i, Watanabe K, Sekine Y, Suzuki T, 2003. An RNA-modifying enzyme that governs both the codon and amino acid specificities of isoleucine tRNA. *Molecular Cell* **12:** 689-698.

63. Spatafora JW, Sung G-H, Johnson D, Hesse C, O'Rourke B, Serdani M, Spotts R, Lutzoni F, Hofstetter V, Miadlikowska J, Reeb V, Gueidan C, Fraker E, Lumbsch T, Lucking R, Schmitt I, Hosaka K, Aptroot A, Roux C, Miller AN, Geiser DM, Hafellner J, Hestmark G, Arnold AE, Budel B, Rauhut A, Hewitt D, Untereiner WA, Cole MS, Scheidegger C, Schultz M, Sipman H, Schoch CL, 2006. A five-gene phylogeny of *Pezizomycotina*. *Mycologia* **98:** 1018-1028.

64. Stone CL, Buitrago MLP, Boore JL, Frederick RD, 2010. Analysis of the complete mitochondrial genome sequences of the soybean rust pathogens *Phakopsora pachyrhizi* and *P. meibomiae*. *Mycologia* **102:** 887-897.

65. Suzuki T, Miyauchi K, 2010. Discovery and characterization of tRNAIle lysidine synthetase (TilS). *FEBS letters* **584:** 272-277.

66. Swofford, 1993. PAUP : A computer program for phylogenetic inference using maximum parsimony. *Journal of General Physiology* **102:** A9.

67. Takemoto C, Spremulli LL, Benkowski LA, Ueda T, Yokogawa T, Watanabe K, 2009. Unconventional decoding of the AUA codon as methionine by mitochondrial tRNAMet with the anticodon f5CAU as revealed with a mitochondrial in vitro translation system. *Nucleic Acids Research* **37:** 1616-1627.

68. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S, 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* **28:** 2731-2739.

69. Taylor JW, Berbee ML, 2006. Dating divergences in the Fungal Tree of Life: review and new analyses. *Mycologia* **98:** 838-849.

70. Thompson JD, Higgins DG, Gibson TJ, 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22:** 4673-4680.

71. Weber F, Dietrich A, Weil JH, Marechal-Drouard L, 1990. A potato mitochondrial isoleucine tRNA is coded for by a mitochondrial gene possessing a methionine anticodon. *Nucleic Acids Research* **18:** 5027-5030.

72. Weinberg Z, Perreault J, Meyer MM, Breaker RR, 2009. Exceptional structured noncoding RNAs revealed by bacterial metagenome analysis. *Nature* **462:** 656-659.

73. Werth S, 2010. Population genetics of lichen-forming fungi - a review. *Lichenologist* **42:** 499-519.

74. Wiens JJ, Kuczynski CA, Stephens PR, 2010. Discordant mitochondrial and nuclear gene phylogenies in emydid turtles: implications for speciation and conservation. *Biological Journal of the Linnean Society* **99:** 445-461.

75. Yang M, Ge Y, Wu J, Xiao J, Yu J, 2011. Coevolution study of mitochondria respiratory chain proteins: Toward the understanding of protein-protein interaction. *Journal of Genetics and Genomics* **38:** 201-207.

76. Zuker M, Stiegler P, 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Research* **9:** 133-148.