THÈSE DE DOCTORAT

présentée par :

Florent WALTZ

soutenue le : **6 Décembre 2018**

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline : **Sciences de la Vie et de la Santé**
Spécialité : **Aspects Moléculaires et Cellulaires de la Biologie**

Characterization of the mitochondrial translation apparatus of *Arabidopsis thaliana*

THÈSE dirigée par :
  M. GIEGE Philippe   Docteur, université de Strasbourg

RAPPORTEURS :
  Mme. LURIN Claire   Docteur, université de Paris-Saclay
  M. LIGHTOWLERS Robert   Professeur, université de Newcastle

AUTRES MEMBRES DU JURY :
  Mme. DUCHENE Anne-Marie   Professeur, université de Strasbourg   Examinatrice interne

  M. MIREAU Hakim   Docteur, université de Paris-Saclay   Membre invité
  M. HASHEM Yaser   Docteur, université de Bordeaux   Membre invité

Acknowledgments/Remerciements

First, I would like to thank the members of the Jury, Pr. Anne-Marie Duchêne, Dr. Claire Lurin and Pr. Robert Lightowlers for agreeing to evaluate my work. As invited members, I would also like to thank Dr. Yaser Hashem and Dr. Hakim Mireau.

Un grand merci à Philippe pour m'avoir accueilli au sein de son équipe. Merci à lui pour sa patience, son implication et ses excellents conseils tout au long de ma thèse qui m'ont permis d'acquérir de nombreuses connaissances scientifiques et techniques. Merci de m'avoir fait confiance en me permettant de travailler sur ce super projet, même quand ça ne marchait pas. Merci également pour la liberté que j'ai pu avoir sur le projet et de m'avoir donné la possibilité de participer à de nombreux congrès !

Merci à Anne-Marie, Laurence, Gag, Hélène, Thalia, José, André et Heike d'avoir bien voulu partager leurs connaissances au cours de nos séances de questions lors des entrainements de Master, et plus tard pendant ma thèse. J'ai aussi appris beaucoup grâce à eux. Merci particulièrement à Laurence pour sa confiance lors des RNA Salons.

Merci à Hakim et son équipe pour m'avoir accueilli pendant deux semaines à l'INRA de Versailles. Merci à Trung (Master Trung) pour sa patience durant mon apprentissage du ribosome-profiling.

Merci à Yaser pour le temps passé à l'IGBMC sur le microscope, malgré les galères, et sans qui la partie structurale du projet n'aurait pas été possible. Merci également à Jailson pour ses conseils pour les purifs de ribosomes.

Mes remerciements s'adressent bien évidemment aux membres de l'équipe, présents ou passés, qui ont fait de ces 3 ans une expérience enrichissante, merci pour leur bonne humeur, leur soutien et leurs précieux conseils. Anthony qui m'a appris la majorité des techniques de bases, Cédric qui est maintenant parti et Matthieu qui reprend la suite, Géraldine pour sa bienveillance et ses conseils, Ayoub pour son caractère jovial et Kamel qui a maintenant sa propre équipe. Je tiens aussi à remercier particulièrement Mathilde qui est arrivée au moment où j'ai débuté ma thèse, et avec qui j'ai toujours apprécié discuter et qui m'a été d'une grande aide avec les manips. Merci tout particulièrement à Pierre également pour les bons moments passés au labo ou en dehors, ainsi que pour son aide.

Je tiens également à remercier les stagiaires qui sont passés par le labo, notamment Herrade qui m'a suivie à la trace dans mon parcours académique et avec qui j'ai également pu passer d'excellents moments en dehors du travail (du coup aussi merci à Thibaud pour supporter Herrade au quotidien).

Egalement merci à mes camarades de Master, qu'ils aient continué à l'IBMP ou non, Nicolas, Deborah, Yannick, ainsi que les autres étudiants en thèse pour leur bonne humeur (presque) toujours au rendez-vous, Hélène, Marlène, Marco, Magdalène, Arnaud, Guillaume, Stéphanie, Marion, Thibaud, Adrien, etc …

Merci à Joern et mes collègues moniteurs avec qui j'ai pu durant ces trois dernières années découvrir l'enseignement, qui fut aussi une expérience à part entière.

Merci à Johana, Lauriane et Philippe de la plateforme protéomique pour avoir toujours été à l'écoute et avoir été aussi réactifs, même lorsque je leur demandais de passer des séries de 12 échantillons en 3 jours. Sans eux rien de ce que j'ai fait n'aurait été possible.

Merci à Todd et Stefano pour les conseils durant ma mi-thèse. Thank you Misha for your advices and for allowing me to use your gradient collector.

Merci à ma famille et particulièrement à mes parents qui ont su me montrer le chemin et éveiller ma passion pour les sciences. Merci pour m'avoir toujours soutenu depuis le début même si je sais qu'il est parfois difficile de saisir les tenants et les aboutissants de ce que je fais.

Merci à Caroline avec qui j'ai partagé ses trois dernières années. Les longues heures au 306 pendant le stage à discuter, les voyages autour du monde (jusqu'en Chine), les concerts, etc. Merci pour le soutien que tu m'as apporté dans les bons et les mauvais moments. J'ai pu partager / faire / voir / visiter / découvrir tellement de choses que je n'avais pas faites avant et je t'en suis reconnaissant. Je te fais entièrement confiance pour ta thèse, tu en es plus que capable, n'en doute pas. Merci moumou.

Je tiens aussi à remercier mes bons amis Nicolas et Nathan qui m'ont aussi soutenus à leur manière durant toute ma scolarité (c'est-à-dire pratiquement depuis la maternelle).

Plus généralement, merci à tout le personnel de l'IBMP, les responsables des différentes plateformes et les membres des différentes équipes que j'ai pu croiser au détour d'un couloir.

"Success is not final, failure is not fatal: it is the courage to continue that counts." - Winston Churchill

Abbreviations

aa : amino acid

ADP : Adenosine diphosphate

AOX : Alternative Oxydase

Asn : Asparagine

Asp : Aspartate

ATP : Adenosine triphosphate

AtPRORP : *Arabidopsis thaliana* PRORP

BN-PAGE : Blue Native PAGE

Ca^{2+} : Calcium

cDNA : complementary DNA

co-IP : Co-immuno-precipitation

CP : Central protuberance

cryo-EM : Cryogenic electron microscopy

C-ter : Carboxyl-terminus

DNA : Deoxyribonucleic acid

ETC : Electron Transport Chain

FAD : Flavin adenine dinucleotide

Fe/S : Iron-sulphur

FECA : First Eukaryotic Common Ancestor

GDP : Guanosine diphosphate

GFP : Green Fluorescent Protein

GTP : Guanosine triphosphate

IMM : Internal mitochondrial membrane

IMS : Inter-membrane space

kb : Kilobase

kDa : KiloDalton

kDNA : Kinetoplastid DNA

LB medium : Lysogeny broth medium

LC-MS/MS : Liquid chromatography–mass spectrometry

LECA : Last Eukaryotic Common Ancestor

LSU : Large Subunit

MAM : Mitochondrial associated membranes

MCS : Membrane contact sites

Met : Methionine

MICOS : Mitochondrial contact site and cristae organizing system

mRNA : messenger RNA

MS : Mass spectrometry

MS medium : Murashige and Skoog medium

mtDNA : mitochondrial DNA

MTS : mitochondrial targeting sequence

NAD : Nicotinamide adenine dinucleotide

nDM : n-Dodecyl β-D-maltoside

N-ter : amino-terminus

O$_2$: Dioxygen

OD : Optical density

OMM : Outer mitochondrial membrane

OXPHOS : Oxidative phosphorylation

PAGE : Polyacrylamide gel electrophoresis

PCD : Programmed cell death

PCR : Polymerase chain reaction

PfPRORP : *Plasmodium falciparum* PRORP

P$_i$: Inorganic phosphate

PPR : Pentatricopeptide repeat

PRORP : Protein Only RNase P

PTC : Peptidyl transferase center

PVDF : Polyvinylidene fluoride

RcPRORP : *Romanomermis culicivorax* PRORP

RNA : Ribonucleic acid

rPPR : ribosomal PPR

rRNA : ribosomal RNA

RT : Reverse transcription

S : Svedberg

SD sequence : Shine-Dalgarno sequence

SDS : Sodium dodecyl sulfate

Ser : Serine

ssRNA : Single stranded RNA

SSU : Small subunit

TCA cycle : Tricarboxcylic acid cycle

T-DNA : Transfer DNA

TIM : Translocase of the inner membrane

TOM : Translocase of the outer membrane

TPR : Tetratricopeptide repeat

tRNA : Transfert RNA

UTR : Untranslated Transcribed Region

VDAC : Voltage-dependent anion channels

WT : Wild-type

List of figures and tables

List of figures

List of tables

Table of content

INTRODUCTION

RESULTS

DISCUSSION

MATERIALS AND METHODS

RESUME EN FRANCAIS

Introduction

Introduction

This thesis is about translation in plant mitochondria. Before addressing the topics directly linked to my work, I wish to give some insights about the crucial role that the acquisition of mitochondria has played for the evolution of life on Earth. This is presented as a preamble to my thesis introduction.

Preamble: Mitochondria as the key to eukaryote origin and expansion

Life arose relatively early in the history of Earth. The first traces of life can be dated back to 3.8 billion years ago, a little more than 500 million years after Earth formation. Before the emergence of cells, a "pre-biotic" period saw the development of the molecular components required for life. A widespread theory proposes that during this pre-biotic time the first complex molecule that arose was RNA. In this hypothetical stage of Earth evolution, called "RNA world", self-replicating RNA molecules proliferated before the evolution of DNA and proteins (Bernhardt, 2012). This theory is supported by the fact that several core components of the cell, like the ribosome – the universal cellular machine that synthesizes proteins – are composed mainly of RNA, which may constitute relics of this ancient world.

During approximately 2 billion years, the Earth was populated by simple unicellular organisms, "complex" life only appeared much later. Morphologically complex living organisms – plants, animals, fungi – and single-celled "protists" (such as *Trypanosoma brucei* or *Plasmodium falciparum*), descend from one singular ancestor LECA – Last Eukaryotic Common Ancestor – estimated to have lived ~1.6–2 billion years ago. Therefore, they form in phylogenetics what we call a monophyletic group. This group, called Eukarya (literally "true nucleus"), or more commonly Eukaryotes, was named after their major morphological feature which is the presence of a membrane-enclosed nucleus containing the cell's genetic information. The other super-group of organisms, the first ones that have appeared on Earth, are called Prokaryotes (regrouping Bacteria and Archaea). They were named in opposition to the Eukaryotes, as they do not possess a nucleus (Sapp, 2005; Zimmer, 2009). Even though eukaryotes were originally classified according to the presence or not of a nucleus, many more cellular and molecular traits define if a given organism belongs to Eukaryotes. Among those characteristics, the nucleus should harbor nuclear pores – inside the nucleus, nuclear lamina is found along with linear chromosomes with telomeres, facilitating sexual reproduction. Complex regulatory mechanisms, including chromatin (histones …), an RNAi system and small non-coding RNAs, orchestrate gene expression at different levels.
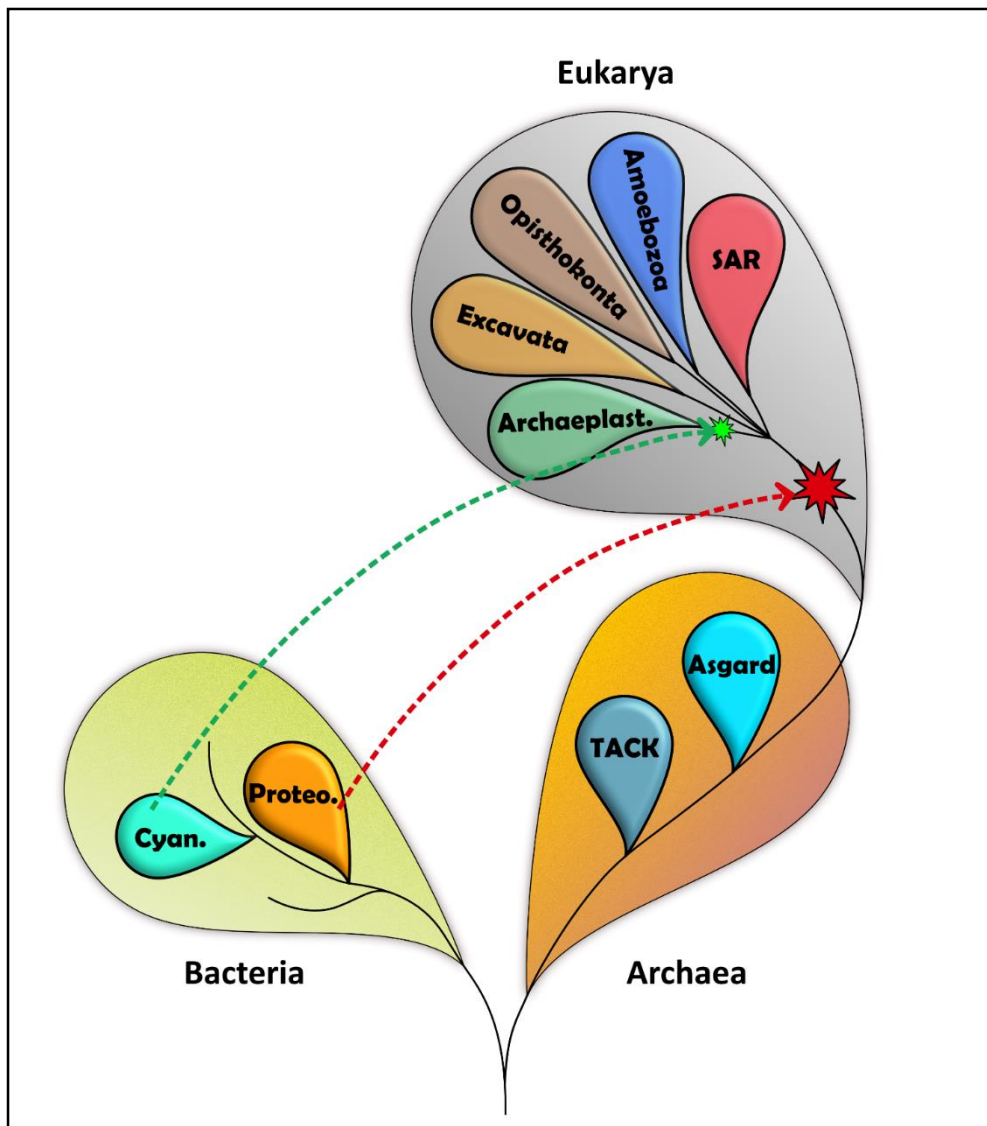
Figure 1: The two-domains tree of life
Schematic representation of the "tree of life". In the 70's all cellular life was divided into the three major evolutionary lines: Eukarya (or eukaryotes), Bacteria and Archaea.

But it was later discovered that eukaryotes emerged from Archaea, hence the two-domains tree of life here represented. In the early 2010's eukaryotes were found to branch within, or as sister to, the Thaumarchaeota, Aigarchaeota, Crenarchaeota and Korarchaeota (TACK) superphylum. But more recent phylogenomic analyses now suggest that eukaryotes originated from within the Asgard archaea phylum or that they represented a sister group to them (Eme et al., 2017).

The red and green arrows represent the two major endosymbiosis events, in red between an α-proteobacteria and the ancestor of eukaryotes and in green a cyanobacteria and the ancestors of Archaeplastida.

Transcription is uncoupled from translation and involves extensive RNA processing (including intron splicing, capping and polyadenylation). Translation itself is much more complex, putting into action eukaryote-specific ribosomes and numerous additional translation factors. Also, an elaborate protein regulation and recycling system composed of the proteasome and an ubiquitin signaling systems are always found. The cellular eukaryotic environment is highly compartmentalized by the presence of a sophisticated endomembrane systems composed of the endoplasmic reticulum, the Golgi apparatus, endosomes, lysosomes and peroxisomes. A complex actin-based and tubulin-based cytoskeleton and associated molecular motor proteins enabled intracellular trafficking, cell motility and a complex cell cycle, including meiosis, allowing sexual reproduction. They are also able to synthesize a wide range of eukaryote-specific lipids and phospholipids (eg. sterols and sphingolipids) (Eme et al., 2017). And of course, eukaryotes are host of aerobic or facultatively aerobic, mitochondria that descended from a once free-living alpha-proteobacterium (Andersson et al., 1998).

The origin of eukaryotes has always been, and is still, subject to strong debates. The pioneering work of **Carl Woese** and colleagues revealed that all cellular life could be divided into the three major evolutionary lines: Eukarya (or eukaryotes), Bacteria and Archaea. Life was therefore represented as a three-domains tree of life, each domain represented as a monophyletic group, Archaea and Eukarya sharing a unique common ancestor. But already in the early 80's molecular phylogenetic analyses showed that eukaryotes and archaea were sister groups. These results were reinforced by the multiple metagenomic approaches which allowed the sequencing and discovery of new archaeal organisms, hence permitting a better comprehension of Archaea. Most experts tend now to class Eukaryotes as a sister group of the archaeal clade Asgard, named after the realm of the gods in Scandinavian mythology, as several eukaryote-specific features never found before in any prokaryotes (specific ribosomal proteins, cytoskeleton component, ubiquitin system, trafficking machinery,…), were found in this group of organisms, favoring of a two-domains tree of life rather than the original three-domains one (Fig 1) (Zaremba-Niedzwiedzka et al., 2017).

Even though it now appears that Eukaryotes emerged from Archaea, they are by nature chimeric and symbiotic organisms. Chimeric first because when you look into eukaryote genomes, a large proportion of the genes can be traced back to prokaryote (eg. about 36% in yeast (Cotton and McInerney, 2010)). Among those prokaryotic-like genes, most of them can be traced back to bacteria (55–70%) whereas a smaller proportion is related to Archaea (20–35%). In term of
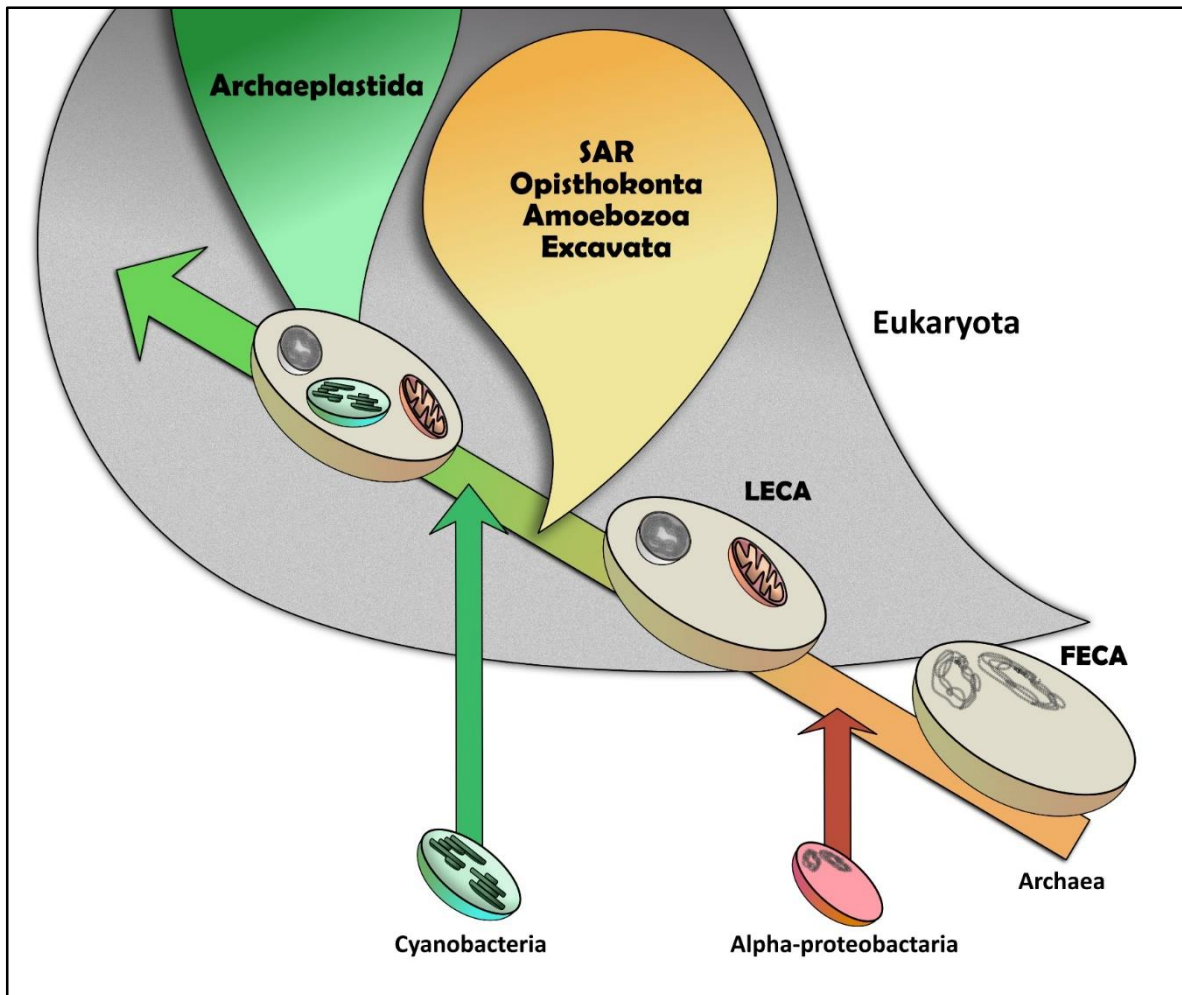
Figure 2: The endosymbiosis in eukaryote evolution
Schematic representation of the two major endosymbiosis events that led to the eukaryotes that we know today. The first major one happened during the event of eukaryogenesis (the series of events that drove the evolution of FECA to LECA). The α-proteobacteria, that later became the mitochondrion, was acquired by an "early eukaryote", but the exact course of events is not fully understood. Later the engulfment of a cyanobacteria by an eukaryote led to the apparition of the "green-phylum" Archaeplastida.

functions, genes with a bacterial ancestry are overwhelmingly linked to metabolic processes, whereas archaeal genes tend to be involved in information processing – DNA replication, transcription and translation machineries – essential functions that comfort the archaeal origin of eukaryotes. This huge proportion of bacterial genes reveals the chimeric nature of eukaryotes. In the 60's **Lynn Margulis** already proposed (Sagan, 1967) that eukaryotes were not the product of standard natural selection, but the result of a serial of endosymbiosis events, therefore called the "serial endosymbiosis theory". This concept was based in part on early hypotheses, for instance from the 19th century French botanist **Andreas Schimper** (born in Strasbourg in 1856). On a footnote to his 1883 work, he made the observation that the chloroplasts that are found in photosynthetic organisms shared many characteristics with cyanobacteria and thus proposed that the combination of two separate organisms may have given rise to modern photosynthetic organisms (Schimper, 1883). The Margulis theory stipulates that a number of bacteria cooperated together so closely that some cells got inside others, thus leading to the formation of the nucleus and all the complex compartmentalization of eukaryotes, predicting both the chimeric and symbiotic nature of eukaryotes. This theory was later proven wrong, but not entirely as the events of endosymbiosis took place twice during eukaryote evolution (Fig 2) (Lane, 2017). Once during early eukaryote evolution, before LECA, thus more than ~2 billion years ago, leading to what we now know as the mitochondria, which is conserved across (almost) all eukaryotes. Organisms that would have diverged before the acquisition of mitochondria (or amitochondriate eukaryotes) were investigated but never found. Such organisms, named Archezoa, were proposed by **Thomas Cavalier-Smith** (Cavalier-Smith, 1989). Organisms like *Giardia intestinalis* and *Trachipleistophora hominis* were thought to be "archezoan", however it was later proven that these organisms possess relics of mitochondria and had thus lost the latter secondarily, not being evolutionary intermediates. The kingdom Archezoa has therefore been abandoned (Poole and Penny, 2007). The second endosymbiosis event occurred later, about 1.5 billion years ago, with the acquisition of the chloroplast, and contributed to the apparition of the "green phylum" (or Viridiplantae which is made up of the green algae and the land plants) of eukaryotes (Dorrell and Howe, 2012). Those two components of eukaryotic cells, called organelles, are primarily energy producing compartments.

These compartments are both able to produce energy due to the presence in their membrane system of an electron transport chain (ETC). It relies on a series of protein complexes that transfer electrons from electron donors to electron acceptors via redox reactions, and couples this electron transfer with the transfer of protons across a membrane. The flow of protons across
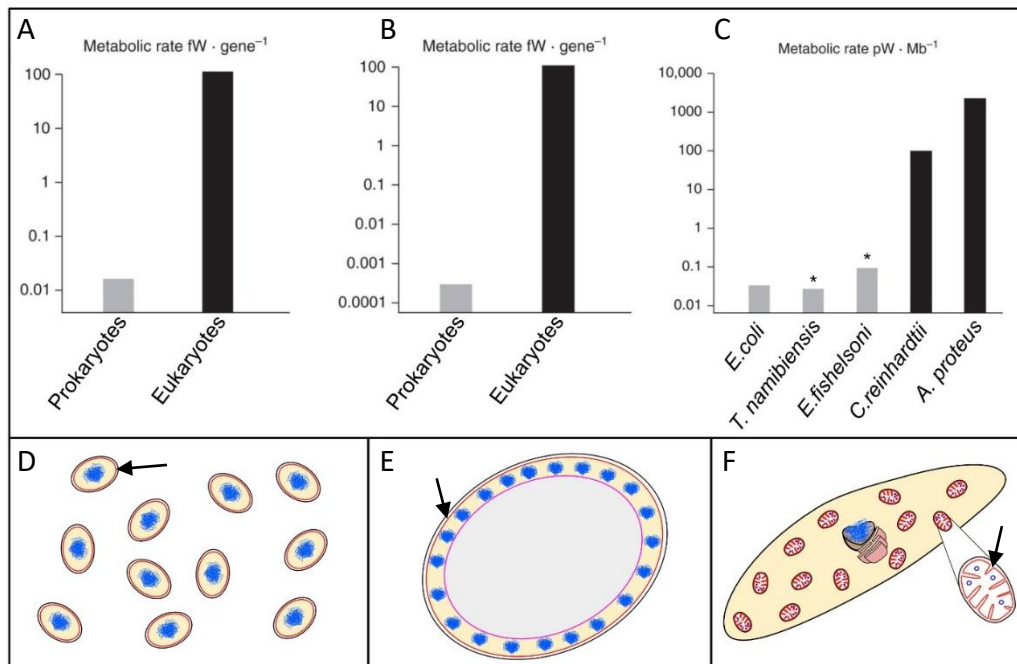
Figure 3: Energy availability in prokaryotes versus eukaryotes

In **A** the mean energy per gene in prokaryotes versus eukaryotes equalized for genome size are represented. The same is represented in **B** but equalized for genome size and cell volume. In **C** the power available per haploid genome (energy per gene X number of genes in one haploid genome) is represented for different prokaryotic and eukaryotic organisms. Note the log scale is each case, eukaryotes have 4-5 orders of magnitude more energy available. Both *T.namibiensis* and *E.fishelsoni* are giant bacteria, but even though they are able to reach such proportions their internal volume is metabolically quite inert. To sustain this giant life style they require multiple copies of their genomes, in the case of *E.fishelsoni* 200,000 copies of its 3.8 Mb genome, in order to produce enough energy. In the end, even if large bacterium do exist they are not able to sustain "true" cellular complexity, as the totality of their energetic resources have to be mobilized to sustain this life style.

D, **E** and **F** are the organisms discussed above. **D** Schematic representations of a medium sized prokaryote (*E.coli*), **E** a very large prokaryote with its inert vacuole in grey (*T.namibiensis*) and **F** a medium-sized eukaryote and its multiple mitochondria allowing large energy availability (*A.proteus*). Bioenergetic membranes across which chemiosmotic potential is generated are drawn in red and indicated with a black arrow and DNA is indicated in blue.

Derived from (Lane, 2011; Lane and Martin, 2010)

the membrane creates a electrochemical gradient that ultimately drives the ATP synthase and permits, every ten protons passing through it, the full rotation of the ATP synthase's head allowing 3 ADP molecules to be turned into 3 ATP molecules, which will be used in all metabolic processes (Abrahams et al., 1994). The oxidative phosphorylation, which is the name of this process in mitochondria, was first hypothesized in the early 60s by **Peter Mitchell**, called the "Chemiosmotic theory" (Mitchell, 1961), which won him the 1978 Chemistry Nobel Prize. In chloroplast, a somewhat similar phenomenon happens, the photophosphorylation, but this process is light dependent. It highly resembles oxidative phosphorylation, the only evolutive innovation compared to the latter is the addition of the chlorophyll pigments to use the energy of the sun to create a high-energy electron that splits water, releasing O_2 and protons. Electrons then move spontaneously from donor to acceptor through the electron transport chain.

The acquisition of mitochondria during the eukaryogenesis, and the fact that is was retained since then, in one form or another (eg. hydrogenosomes or mitosomes) – at the exception of a Monocercomonoid parasitic protist of that entirely lost mitochondria (Karnkowska et al., 2016) – prove the original symbiotic nature of eukaryotes. This cooperative way of functioning is for sure not as symbiotic as it was originally thought to be, with the original host more-or-less "enslaving" the original endosymbiont. Indeed mitochondria is now non-autonomous as the large majority of genes originally encoded in the endosymbiont were either lost or transferred to the nucleus of the host cell. From the archaeal ancestor of eukaryotes, FECA (First Eukaryotic Common Ancestor) to LECA, many other evolutive events happened, aside the acquisition of mitochondria. The exact nature of those events, and their chronology, is still debated. In the case of mitochondria, most competing scenarios can be roughly grouped into either mito-early, which considers the driving force of eukaryogenesis to be mitochondrial endosymbiosis into a simple host, or mito-late, which postulates that an already somewhat complex cell predated the original endosymbiont through phagocytosis. Along that, which of standard natural selection, horizontal gene transfer, or endosymbiosis/phagocytosis contributed the most to what LECA was, is still an unsolved question. Many eukaryote-specific traits for sure arose from standard natural selection. As predicted by the evolutionary theory, complex traits arise via a series of small steps (Darwin, 1859). Nevertheless, the acquisition of mitochondria gave a huge boost to the evolution by simply releasing the energetical constraint of these cells. Indeed, when comparing the energy available per gene between a prokaryote and a simple eukaryote, a eukaryotic nuclear gene governs nearly 5,000 times more energy flux than a prokaryotic gene (Fig 3) (Lane, 2011; Lane and Martin, 2010). Indeed, prokaryotes can generate energy only by pumping

protons across their membrane, thus being limited by their size. On the other hand, eukaryotes pack hundreds of mitochondria into a single cell allowing much more energy production. With that much power available, eukaryotes were able to develop a large number of new functions, many listed before, but most importantly, it permitted to release the structural constraint of cell size and genome size. Indeed eukaryotic cells are usually much larger (around 10-100 times), they have much larger genomes (around 40-600 times) and more genes (around 5 times), along with the cell structure itself being much more complex compared to prokaryotes. Overcoming this energetical barrier contributed much likely to the rapid radiation of eukaryote into the five super-groups of organisms that we know today (Adl et al., 2012). To date, the bacteria *Rickettsia prowazekii* is the more closely related organism to mitochondria in term of genetic structure (Andersson et al., 1998), but the original endosymbiont was most likely a part of a sister group of α-proteobacteria (Martijn et al., 2018). On the opposite, the jakobid *Reclinomonas americana*, and jakobids in general, possess the mitochondrial genome which most closely resembles the ancestral proto-mitochondrial genome, due to their high gene content (Lang et al., 1997).

Altogether, endosymbiosis is a process that had a profound impact on all extant eukaryotes. The acquisition of mitochondria marks a crucial step in eukaryotes emergence and is one of the clearest examples of an evolutionary transition. Beside the evolutive questions, mitochondria is also a crucial subject of study for human health as their dysfunction has been associated with an increasingly large number of inherited disorders and is implicated in common diseases, such as neurodegenerative disorders, cardiomyopathies, metabolic syndrome, cancer, and obesity (Nunnari and Suomalainen, 2012; De Silva et al., 2015). In plants, this organelle has also attracted considerable attention since they specify a widely expanded trait leading to an inability of plants to produce functional pollen, called "cytoplasmic male sterility". Plant mitochondria are thus of huge agronomical interest (Chen and Liu, 2014; Horn et al., 2014).

The compartments of gene expression in plants (Viridiplantae)

The term "plant" refers to a group of eukaryotic organisms possessing the following traits: multicellularity, presence of cell walls composed of cellulose and the ability to carry out photosynthesis with primary chloroplasts, using light, water and carbon dioxide to synthesize nutriments, making them autotrophic organisms. Viridiplantae (literally "green plants") encompass a group of eukaryotic organisms made up of the green algae, which are primarily aquatic, and the land plants (embryophytes), which emerged within them. Embryophytes include the vascular plants, such as ferns, conifers and flowering plants.

In plants, three distinct cell compartments carry out genetic expression: the nucleus, mitochondria and chloroplasts. In these compartments genetic expression is differently fulfilled, and is tightly orchestrated through inter-organellar crosstalk, mitochondria and chloroplasts being completely dependent of the nucleus. As described above, mitochondria and chloroplasts were acquired through two consecutive endosymbiosis events. The first event happened about 1.5–2 billion years ago, which resulted in the acquisition of mitochondria. It appears that the original bacterium which later became mitochondria was an α-proteobacterium (Gray et al., 1999) but these results are now being reconsidered (Martijn et al., 2018). A secondary endosymbiosis event occurred between an already fully-fledged eukaryote and a cyanobacteria over 1 billion years ago (Dorrell and Howe, 2012), which ultimately lead to the formation of plastids, and gave rise to the "green-phylum".

During evolution, most of the genes that were initially encoded in the original endosymbionts were either transferred to the nucleus of the host cell, or between endosymbionts (Hao et al., 2010), or lost (Brown, 2003). The genome size of around 4,500 genes in the original endosymbionts decreased to 3–67 genes in mitochondria and 23–200 genes in chloroplasts, as a result these are now semi-autonomous. For example, in *Arabidopsis thaliana*, 18% of the nuclear genes appear to be derived from cyanobacteria (Martin et al., 2002). Genetic expression in the nucleus and in the cytosol of plants is similar to what can be found in other eukaryotes. In chloroplasts and mitochondria, almost 2 billion years of evolution led to the apparition of specific gene expression mechanisms combining bacterial-like traits with novel features that evolved in the host cell. The most striking differences are observed between mitochondria from different big groups of eukaryotes. Indeed, as the evolutive radiation into the five big groups of eukaryotes (Adl et al., 2012) seems to have occurred quickly after the acquisition of mitochondria, each of them developed specific features.
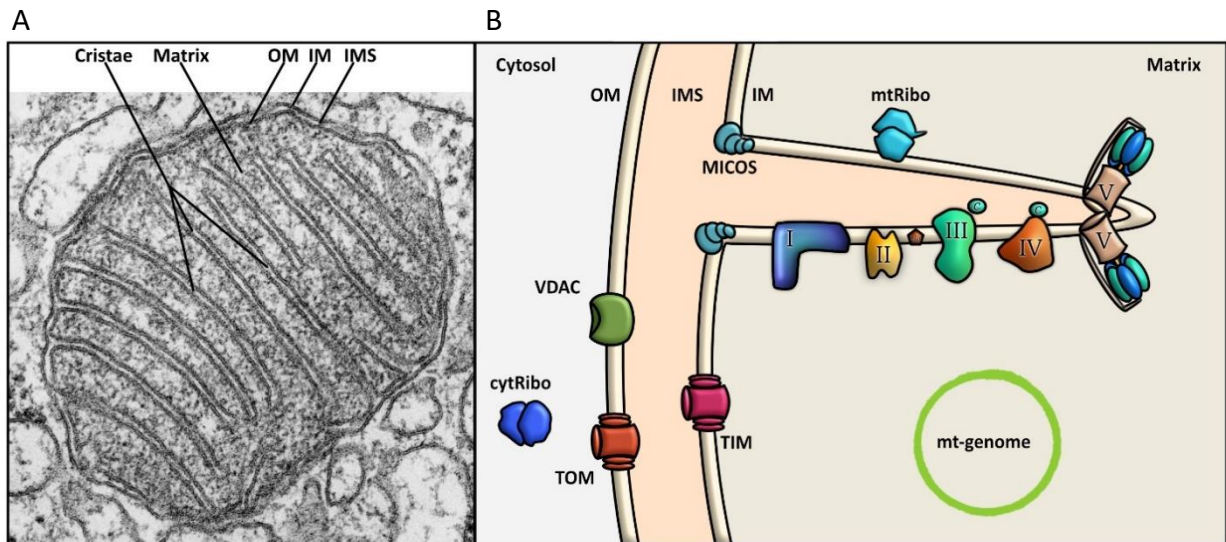
Figure 4: The overall mitochondria organization
A Electron micrograph of an animal mitochondria, the different main components are indicated. **B** Schematic representation of the different components of mitochondria (described in the main text).

Mitochondria in plants and other eukaryotes

Overall structure of mitochondria:

Mitochondria are composed of two different membranes: the outer and inner membranes. Those two membranes of distinct natures delimit two biochemically different compartments; the inter membrane space (IMS) and the matrix. The outer membrane is rather permeable thanks to the presence of β-barrel shaped porins, called VDAC for voltage-dependent ion channel (Hodge and Colombini, 1997; Mihara and Sato, 1985). Those VDAC proteins play a key role in regulating metabolic and energetic flux by allowing the transport of several metabolites such as ATP, ADP, pyruvate, malate, … Besides VDAC, the outer membrane also contains so called Translocase of the Outer Membrane, or TOM complexes (Ahting et al., 1999; Dekker et al., 1998). TOMs are involved in the entry of larger molecules in mitochondria, mainly proteins, which are recognized if a signalling sequence at their N-terminus, named MTS, is present which then actively triggers the import (Emanuelsson et al., 2007; Omura, 1998). With the relocation of the majority of the original endosymbiont's genes in the nucleus, most of the mitochondrial proteome has to be first translated by cytosolic ribosomes (often being associated to mitochondria when translating mitochondria-targeted proteins (Gold et al., 2017)) to be later imported in mitochondria, the import machineries are therefore of huge importance for mitochondrial biogenesis. Proteins first pass by the intermembrane space, IMS, where they will be sorted to either be inserted in the outer membrane, stay in the IMS, or go into the matrix or the inner membrane through the Translocase of the Inner Membrane, TIM (Fig 4) (Koehler et al., 1998; Sirrenberg et al., 1996).

The inner membrane encloses the protein-rich matrix and is the seat of the molecular machinery of chemiosmosis. It is rich in cardiolipin, an unusual phospholipid which constitutes roughly 20% of the IMM (internal mitochondrial membrane) and may contribute to the inner membrane impermeability (Hoch, 1992). Indeed, unlike the outer membrane, the inner membrane is freely permeable only to oxygen, carbon dioxide, and water. It is much less permeable to ions and small molecules than the outer membrane which requires special membrane transporters to enter or exit the matrix. This feature is essential for the functioning of the electron transport chain, where protons have to be pumped from the matrix to the IMS to be used by the ATP synthase. The surface of the inner membrane is much larger than that of the outer membrane. As a result, it has to be heavily folded, forming numerous invaginations called *cristae* (Griparic and van der bliek, 2001). By significantly increasing the total membrane surface area they

also increase the available working space. The *cristae* are connected to the inner boundary membrane via tubular structures termed crista junctions (Daems and Wisse, 1966). These internal structures can greatly vary between organisms and even tissues. Their structure seem to be governed by the mitochondrial contact site and cristae organizing system (MICOS) at the *cristae* junctions, but also by the organization of ATP synthases into dimers (Hahn et al., 2016), and of dimers into rows, which is a feature common to mitochondria of all species examined to date. In fungi, plants, and metazoans, the dimers are V-shaped and associate into rows along the highly curved ridges of lamellar cristae but in *Paramecium tetraurelia* it is U-shapped, forming tubular cristae (Mühleip et al., 2016).

Furthermore, mitochondria do not sit alone in the cell. Indeed, a great number of reports have recently shown that mitochondria are in close contact with the endoplasmic reticulum. These contact sites called MAMs (mitochondrial associated membranes), or more generally membrane contact sites (MCS), have mostly been studied in mammals and yeast (Vance, 2014). They contribute to the inter-organelle communication, to the modulation of mitochondrial morphology and function as well as to processes like lipid synthesis, apoptosis and Ca^{2+} homeostasis. Even if these contact sites are more discrete in plants, reports tend to show that they are also present (Mueller and Reski, 2015).

Metabolic functions supported by mitochondria:

Mitochondria are the power stations of eukaryotic cells and the sites of many important metabolic reactions. Among them, we can cite amino acid and nucleotide metabolism, lipid, quinone and steroid biosynthesis, and of course iron-sulfur (Fe/S) cluster biogenesis. As energy producers, their main role is to use reducing agents, derived from catabolic reactions like TCA cycle and β-oxidation of fatty acids, to fuel the oxidative phosphorylation. By channeling electrons through the respiratory chain complexes and creating a transmembrane electrochemical gradient, the ATP synthase is activated, which ultimately allow the conversion of ADP+P_i to ATP, the biochemical energy currency.

Hence, mitochondria are crucial to maintain the high ATP/ADP ratio that is required for the functioning of the many biochemical reactions taking place in eukaryotic cells. Additionally, the TCA cycle generates numerous metabolic intermediates that are utilized by various anabolic pathways. Therefore, mitochondria sustain several crucial metabolic reactions and produce the majority of the cell ATP, constituting biosynthetic and bioenergetic organelles.
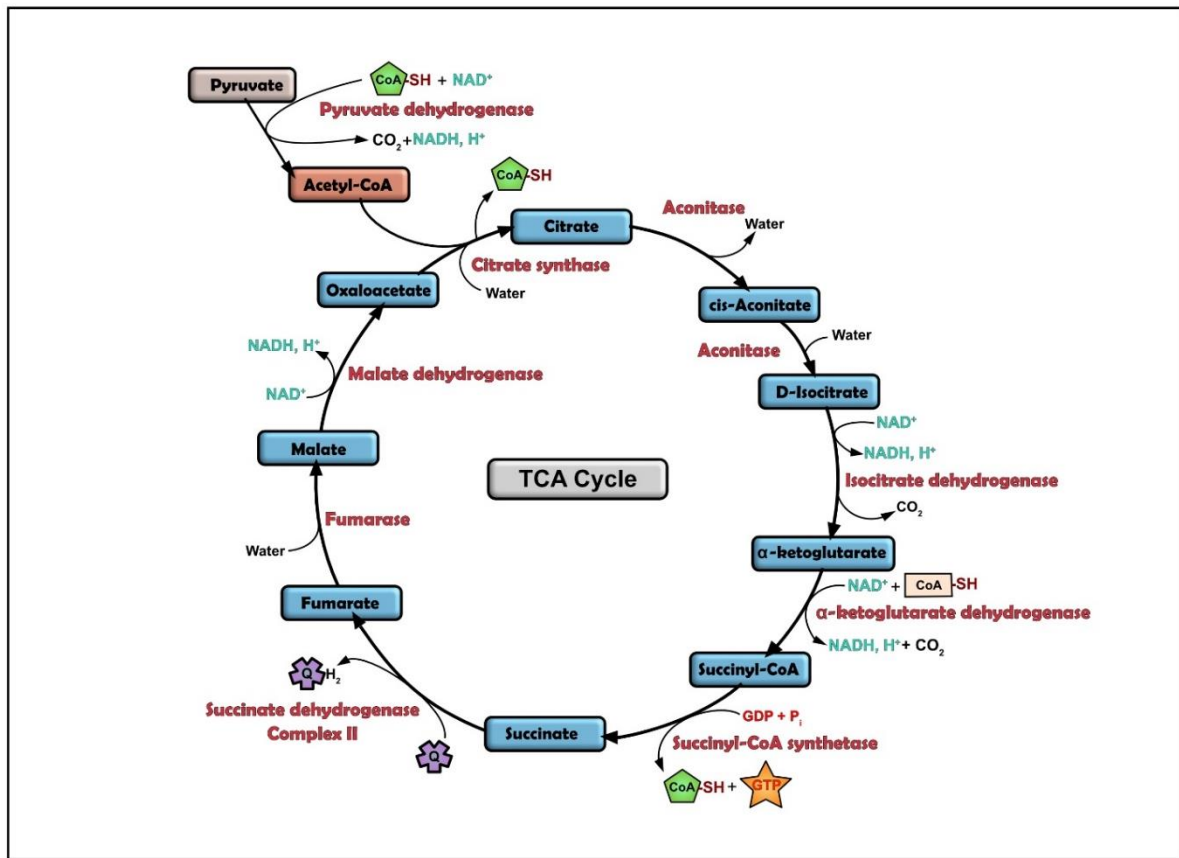
Figure 5: The TCA cycle

Schematic representation of the TCA cycle (TriCarboxcylic Acid cycle).

The cycle consists in a series of chemical reactions used to release stored energy through the oxidation of acetyl-CoA derived from carbohydrates, fats, and proteins. The acetyl-CoA and water are used and allow the reduction of NAD^+ to NADH, and produces carbon dioxide as a waste byproduct. The cycle also provides precursors of certain amino acids. The reducing agent NADH is fed into the oxidative phosphorylation (electron transport) pathway to produce usable chemical energy in the form of ATP. The TCA cycle and the oxidative phosphorylation are closely linked, as Complex II is part of the TCA cycle.

Aside from metabolic functions, mitochondria also play a crucial role in regulating cells life but also death. Indeed, in animals, mitochondria play a central role for a particular programmed cell-death (PCD) event, apoptosis. Calcium signaling and the release of Cytochrome *c* were found to be essential (Wang and Youle, 2009). As in animal cells, mitochondria also seem to be involved in PCD in plants. Upon induction of PCD, plant mitochondria aggregate and swell, a process known as the mitochondrial morphology transition, similar to what is observed during apoptosis (Van Aken and Van Breusegem, 2015). Mitochondria are also key players in regulation of Ca^{2+} due to their ability to accumulate rapidly and transiently this ion involved in cell signaling (Clapham, 2007).

Iron-sulfur cluster synthesis

Iron–sulfur (Fe/S) clusters belong to the most ancient protein cofactors in life, and fulfil many different functions, including electron transfer, redox sensing, enzyme catalysis and sulphur activation. In mitochondria, they are involved in various metabolisms, such as the TCA cycle (aconitase), the electron transfer chain (respiratory complexes I–III), fatty acid oxidation (ETF-ubiquinone oxidoreductase), and in lipoate and biotin biosynthesis (lipoate and biotin synthases) (Lill et al., 1999; Stehling and Lill, 2013). They also retain essential roles in other compartments of the cell, they are involved for example in DNA replication and repair in the nucleus (Stehling and Lill, 2013). Because of the simple nature of Fe/S clusters and their widespread distribution, they have been assigned essential roles in the evolution of life (Martin and Russell, 2003). In eukaryotes, the synthesis of Fe/S clusters and their insertion into apoproteins requires almost 30 proteins, the process being initiated in mitochondria. Interestingly, even in highly degenerated mitochondria, the synthesis of Fe/S clusters has been retained: it is found in Giardia mitosomes, Trichomonas hygrogenosomes or in the mitosome of the apicomplexan parasite *C. parvum* (van der Giezen, 2009). In Monocercomonoides, the only eukaryote described to date to be completely devoid of mitochondria or related organelles, a cytosolic sulfur mobilisation system, acquired through horizontal gene transfer, provides the Fe/S clusters required for protein synthesis. The regular mitochondrial Fe/S cluster synthesis pathway is considered to have been lost secondarily (Karnkowska et al., 2016).

Tricarboxylic acid (TCA) cycle – Krebs cycle

The TCA cycle is the driver of cellular respiration. It was characterized by **Hans Krebs** for which he received the Nobel Prize for Physiology and Medicine in 1953, and after whom the cycle is sometimes named (Krebs cycle) (Fig 5). Taking place in the mitochondrial matrix, it uses the two-
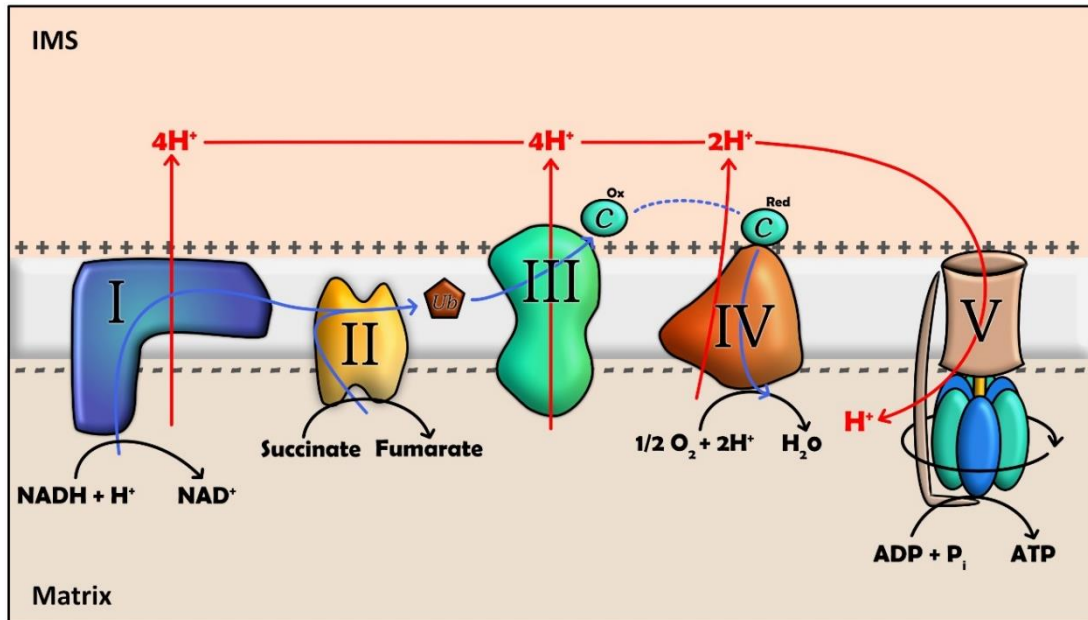
Figure 6: The mitochondrial ETC – the oxydative phosphorylation
Schematic representation of the respiratory chain. The different components, i.e. respiratory complexes I to IV and the ATP synthase (V) as well as the full process are described in the main text.

carbon organic compound acetyl-CoA, produced by the oxidation of pyruvate and originally derived from catabolism of sugars (glycolysis) , fats (lipids beta-oxydation), or proteins (Fernie et al., 2004). In a series of redox reactions, the TCA cycle allow the conversion of acetyl-CoA in the form of NADH, $FADH_2$ and ATP. In a single turn of the cycle, two carbons enter from acetyl-CoA, allowing the formation of three molecules of NADH and one molecule of $FADH_2$, and one molecule of ATP or GTP. The TCA cycle does not produce much ATP directly. However, the NADH and $FADH_2$ generated by this process can make a lot of ATP indirectly. These reduced electron carriers will feed the electron transport chain and, through oxidative phosphorylation, drive synthesis of ATP molecules produced in cellular respiration (Fernie et al., 2004). NADH and $FADH_2$ are also generated during lipids catabolism and glycolysis.

Oxydative phosphorylation and the respiratory chain

Oxidative phosphorylation (OXPHOS) is the main supply of energy in eukaryotic cells. This process is conceptually simple and mechanistically complex. It is supported by the respiratory chain complexes sitting in the IMM. NADH and/or $FADH_2$ act as electron donors that can be oxidized by the Complex I and the Complex II, which constitute the entry points of the respiratory chain. The electrons flow through the different complexes, thanks to a physical phenomenon called quantum electron tunneling, to reach the final electron acceptor O_2 (Hayashi and Stuchebrukhov, 2010). The electrons pass by four complexes, among which three are proton pumps. The whole process leads to the pumping of protons out of the mitochondrial matrix which result in an uneven distribution of protons across the IMM. This generates a pH gradient and a transmembrane electrical potential that creates a proton-motive force (PMF). The final phase of OXPHOS is carried out by Complex V, the ATP synthase, that is driven by the flow of protons back into the mitochondrial matrix (Fig 6).

As mentionned before, the respiratory chain is composed of five protein complexes. The first complex, Complex I or NADH:ubiquinone oxidoreductase constitutes the entry point of electrons through the oxidation of NADH into NAD^+, coupled with the proton translocation from the matrix to the IMS. It is the largest enzyme of the electron transport chain, composed of more than 40 subunits (49 in Arabidopsis (Peters et al., 2013)). The hydrophilic arm mediates the transfer of electrons, via Fe/S clusters to the bound ubiquinone. Reduction of the ubiquinone induces conformational changes in the membrane arm resulting in proton translocation across the membrane via four channels. Electrons can also enter the ETC through the Complex II or succinate ubiquinone oxidoreductase. The oxidization of succinate to fumarate by the Complex II allows the
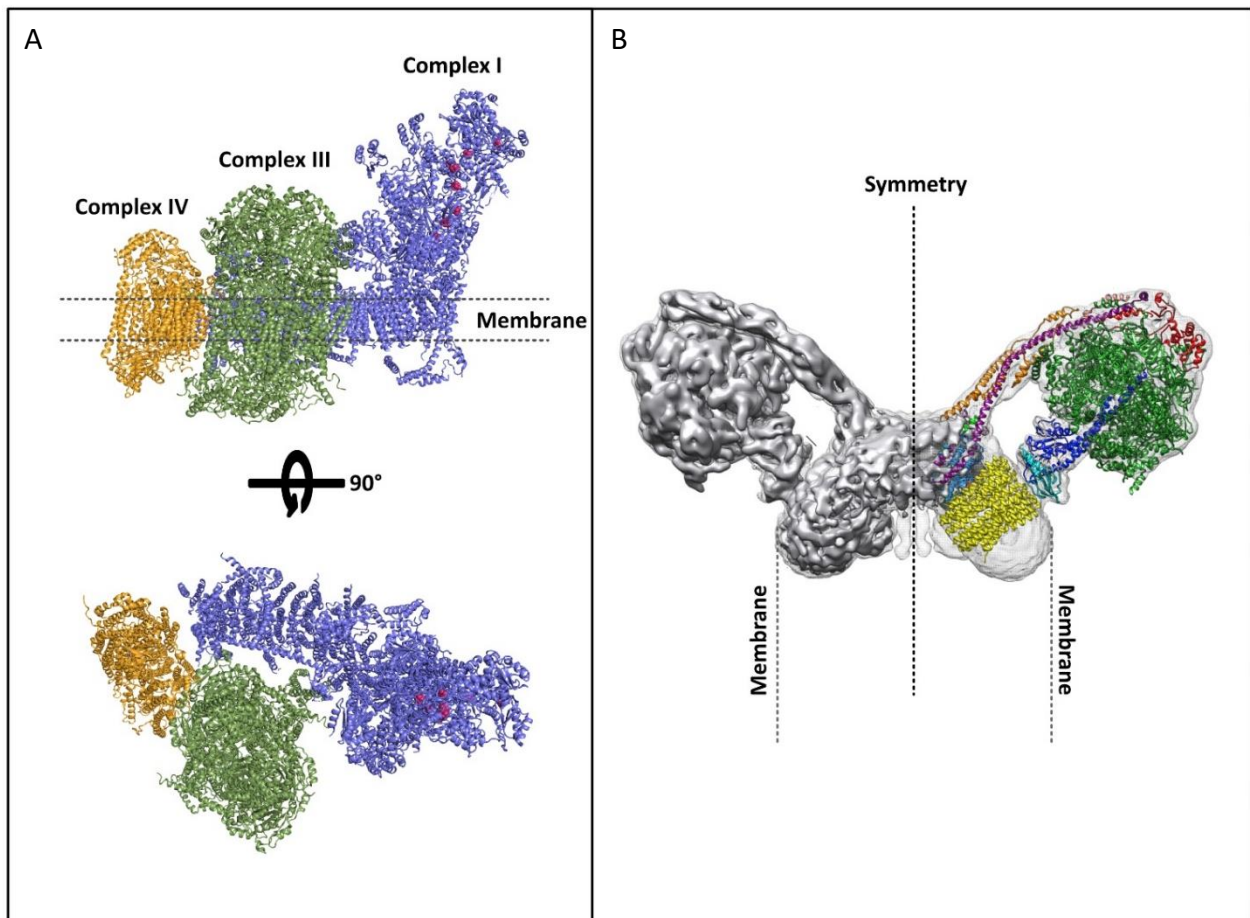
Figure 7: Supercomplexes organization
Atomic models of the **A** bovine respirasome and the **B** yeast dimeric ATPase.
A The respirasome (supercomplex I+III$_2$+IV$_1$) is the most prominent of the existing supercomplexes (except in plants). It contains all the components required to transfer electrons from NADH to the final acceptor, oxygen. Complex I is represented in blue, Complex III in green and Complex IV in yellow. (Sousa et al., 2016)
B The dimeric ATPase are located along the highly curved edges of the inner membrane cristae and the angle formed by these dimers has an effect on cristae structure. Those dimers shape the inner mitochondrial membrane and mediate cristae formation. Derived from (Hahn et al., 2016)

transfer of electrons through the $FADH_2$ intermediary to ubiquinone. The Complex II is the only respiratory chain complex that is not involved in the transfer of protons to the IMS. As a result, Complex I and II both contribute to the formation of ubiquinol, which is freely diffusible in the membrane, and will shuttle electrons to Complex III. The Complex III, or cytochrome *c* oxidoreductase, is usually imbedded in the membrane as a dimer (III_2). The electrons from ubiquinol are passed to the carrier cytochrome *c* via cytochromes b and *c*1. It can also oxidize ubiquinone to ubiquinol to pump two protons to the intermembrane space. Then, similar to ubiquinone, cytochrome *c* can travel between complex III and IV to transfer electrons. Cytochrome *c* transfers its electron to the last enzyme of the ETC, the Complex IV or cytochrome *c* oxidase. Once the electron is delivered, it is transferred to the final acceptor of the ETC, dioxygen, to convert it to two water molecules, leading to four protons being pumped to the intermembrane space during this process (Sousa et al., 2018).

By transporting electrons through the ETC, protons are pumped into the IMS, which lead to the establishment of an electrochemical gradient driving the proton-motive force (PMF). The ultimate step of the respiratory chain is for the protons to flow back to the matrix through the membrane-embedded ATP synthase. It is the ATP synthase, or Complex V, that converts the energy of spontaneous flow of protons into chemical energy of ATP bonds. Complex V is bipartite, composed of a water-soluble F_1-part connected to the membrane-embedded ring-like F_0-part by a central and a peripheral stalk. The protons pass first through the membrane-embedded rotor, the F_0 part of the F_1F_0 ATP synthase, inducing a rotation of the F_0 oligomeric ring that is transmitted into conformational changes in the three nucleotide-binding pockets of the F_1 head, thereby catalyzing ATP synthesis. As mentioned above, ATP synthases are organized in rows of dimers along crista edges, suggesting that the dimers are responsible for bending the IMM and dictating cristae morphology (Lau and Rubinstein, 2012; Stock, 1999).

Beside Complex V, the respiratory components are also able to associate into homo- or heterocomplexes called supercomplexes (Fig 7). They can be divided into three main groups: I + III_2, III_2 + IV_{1-2} and I + III_2 +IV_{1-4} (also called respirasomes, comprising different copy numbers of Complex IV). Complex II is the only enzyme of the respiratory chain which does not associate with the other respiratory complexes. The respirasomes are the most abundant in animal mitochondria (Gu et al., 2016), contrary to plants where the majority of Complex IV (>90%) is present in monomeric state, but respirasome were identified in potato mitochondria (Eubel et al., 2003). It was also shown that respirasomes can associate into respiratory dimer, or megacomplex. These

	H.sapiens	S.cere	A.thaliana	C.reinhardtii	T.brucei	P.falciparum	R.americana
RNAs	24	27	24		2		30
rRNAs	2	2	3	2	2	2	3
tRNAs	22	24	21	3	0	0	26
Proteins	14	8	33	8	18	3	69
Respiratory	13	7	19	7	14	3	26
r-proteins	0	1	7	0	2	0	27
other	1	0	7	1	2	0	16
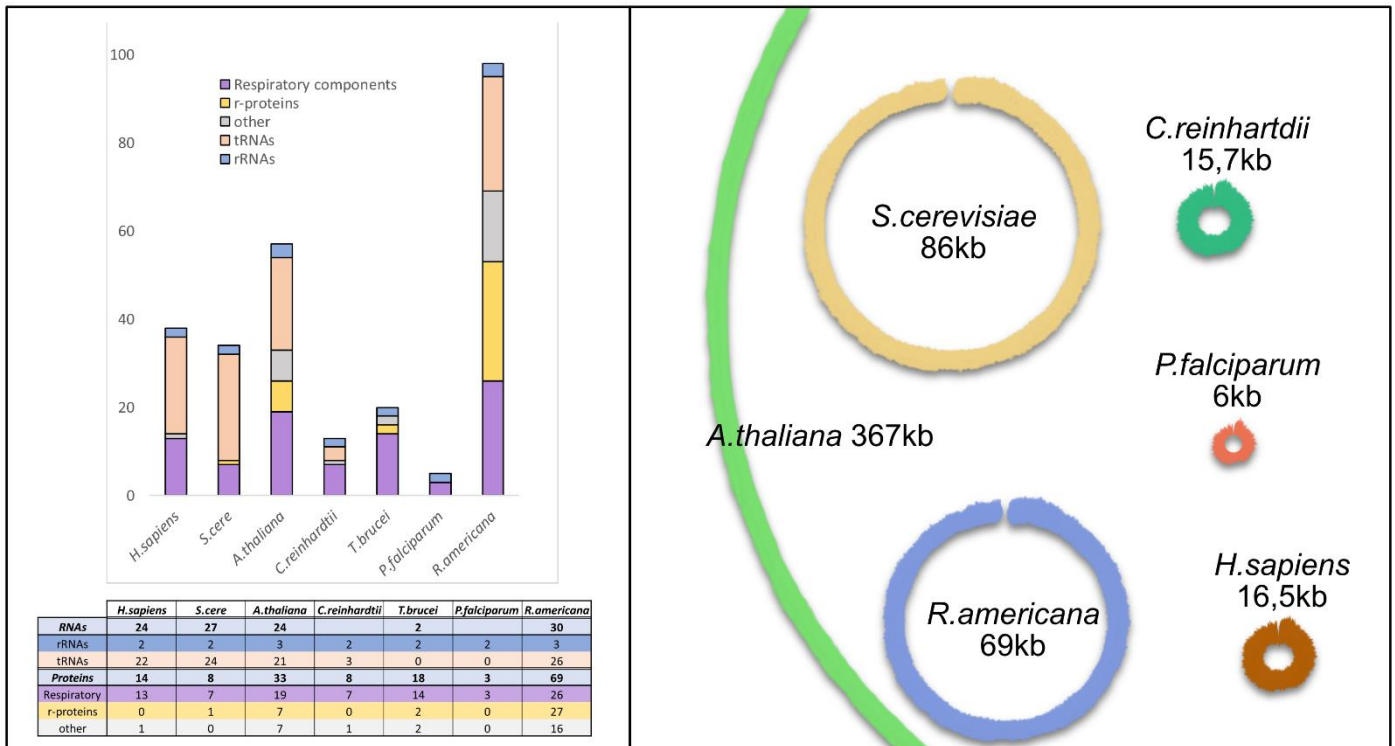
Figure 8: Mitochondrial genomes sizes and contents

A Classification of mitochondrial encoded genes from different eukaryotes. Genes encoded in mtDNA are universally involved in oxidative phosphorylation and mitochondrial translation. Especially in Jakobids, mitochondrial genomes encode additional genes involved in processes such as transcription or RNA maturation. rRNAs are always encoded in the mt-genomes, even if they are sometimes fragmented (e.g *C. reinhardtii* and *P.falciparum*), and a variable number of tRNA genes can be found. The different genes present are listed in Table 1.

B The size of the different mitochondrial genomes exposed in **A** are represented. The genomes are represented as master-circles even though this might not represent the *in vivo* structure of genomes.

results were obtained from different organisms, including porcine, potato, yeast and bacteria, suggesting the existence of a higher order arrangement of respiratory chain elements across species (Bultema et al., 2009; Davies et al., 2011; Gu et al., 2016; Heinemeyer et al., 2007; Sousa et al., 2016, 2013). The organization and complexes found between organisms vary. For example, in *Saccharomyces cerevisiae*, Complex I is not found but is compensated by several peripheral membrane NADH dehydrogenases (Yamashita et al., 2007).

In plants and other eukaryotes, another component is sometimes found, the alternative oxidase or AOX (McDonald and Vanlerberghe, 2006). The AOX is involved in an alternative route for electrons passing through the electron transport chain. AOX is non-proton pumping and since it bypasses complexes III and IV, it dramatically reduces the energy yield of respiration. It is found throughout the plant kingdom but also in the other kingdoms: it is sporadically found in protists and fungi, and also in many animal phyla, although clearly absent from vertebrates and arthropods (Vanlerberghe, 2013). The expression of the AOX gene is influenced by stresses, but the benefit conferred by this activity remains uncertain. It may enhance an organisms' ability to resist these stresses, through reducing the level of oxidative stress (Maxwell et al., 1999).

Mitochondrial genomes:

Mitochondria possess their own genome, vestige of its free-living bacterium ancestor. This genome is localized in the mitochondrial matrix. Over the course of evolution, the majority of the original genome of the endosymbiont was transferred to the nucleus or lost. As a result, only few genes remain encoded in the mitochondrial genome (Gray, 2012). Moreover, the evolutive radiation of eukaryotes most likely occurred after the acquisition of mitochondria, consequently, even if gene transfer was already an ongoing process, great differences in term of structure, size, gene content and expression mechanisms can be observed between the different groups of eukaryotes.

Sizes and gene contents

The size of mitochondrial genome greatly varies between organisms (Fig 8 B). It can be highly reduced like in *Plasmodium falciparum*, which is the smallest known mitochondrial genome with only 6 kb in size, harboring only three protein-coding genes, highly fragmented rRNA genes and no tRNA gene (Feagin et al., 2012). It can also be extremely large, *Silene conica* being the largest with an 11 Mb multichromosomal mt-genome, exceeding the size of some bacterial and even some nuclear genomes (Sloan et al., 2012). Large genomes are also a common feature of the Cucurbitaceae family (Alverson et al., 2011), but for most plants, mt-genomes have about ~300-

	H.sapiens	S.cere	A.thaliana	C.reinhardtii	T.brucei	P.falciparum	R.americana
	16,5kb	86kb	367kb	15,7kb	N/A g	6kb	69kb
RNAs	24	27	24		2		30
SSU rRNA	+	+	+	Fragmented	+	Fragmented	+
LSU rRNA	+	+	+	Fragmented	+	Fragmented	+
5S rRNA			+				+
tRNAs	22	24	21	3			26
RNAse P		rpm1					rnpB
Proteins	14	8	33	8	18	3	69
Complex I - NADH dehydrogenase							
nad1	+		+	+	+		+
nad2	+		+	+	MURF2		+
nad3	+		+		+		+
nad4	+		+	+	+		+
nad4L	+		+		CR3 (putative)		+
nad5	+		+	+	+		+
nad6	+		+	+			+
nad7			+		+		+
nad8					+		+
nad9			+		+		+
nad10							+
nad11							+
Complex II - succinate-coenzyme Q reductase							
sdh2							+
sdh3							+
sdh4			+				+
Complex III - Coenzyme Q - cytochrome c reductase / Cytochrome b							
Cytb/cob	+	+	+	+	+	+	+
Complex IV - Cytochrome c oxidase							
cox1	+	+	+	+	+	+	+
cox2	+	+	+		+		+
cox3	+	+	+		+	+	+
cox11							+
Complex V - ATP synthase							
atp1			+				+
atp3							+
atp4			+				+
atp6	+	+	+		+		+
atp8	+	+	+				+
atp9		+	+				+
Ribosomal proteins							
rps1							+
rps2							+
rps3		+	+		MURF5 h		+
rps4			+				+
rps7			+				+
rps8							+
rps10							+
rps11							+
rps12			+		+		+
rps13							+
rps14							+
rps19							+
rpl1							+
rpl2			+				+
rpl5			+				+
rpl6							+
rpl10							+
rpl11							+
rpl14							+
rpl16			+				+
rpl18							+
rpl19							+
rpl20							+
rpl27							+
rpl31							+
rpl32							+
rpl34							+
Cytochrome c maturation							
ccmA							+
ccmB			+				+
ccmC			+				+
ccmF							+
ccmFC			+				
ccmFN1			+				
ccmFN2			+				
Other proteins							
			mttB (tatC like)				tatA c
							tatC c
							rpoA a
							rpoB a
							rpoC a
							rpoD a
							secY e
							ssrA (tmRNA)
							yejU b
							yejV b
							yejW b
							tufA (EF-Tu)
				rtl (RT-like)			
			matR f		MURF1 d		
					CR4 d		
	Humanin d						

500 kb. This variety of genome sizes in plants is a direct cause of their composition, large repeat sequences promoting recombinations (Gualberto and Newton, 2017). But larger genomes does not imply that more genes are present. In metazoan, the genome is relatively small and conserved, ranging between 15–17 kb and about 16 kb in human. The gene content is also quite stable, with 37 genes. In Arabidopsis, where the genome is 367 kb and contains 57 genes, only 20 genes more, for a genome 20 times larger than that of metazoan (Fig 8 A) (Unseld et al., 1997). The genes still encoded in the mitochondrial genome are rather conserved throughout eukaryotes, and can be classed in two groups for the protein-coding ones: "ribosomal protein" and "bioenergetics"; but additional proteins are sometimes encoded in mt-genomes (Table 1). The former are involved in ribosomal subunit synthesis and mainly occur in protist and plant mt-genomes; the later code for subunits of the respiratory chain complexes as well as cytochrome *c* maturation proteins. Genes coding for the mitochondrial rRNAs are also always found. They are among the few genes universally encoded by mtDNA across eukaryotes (Gray, 2012). tRNAs, required for the translation of the few proteins encoded in the mt-genome, are also found however the presence of a complete minimalist set of tRNA genes encoded by the mt-genome is more an exception than a general rule (Salinas-Giegé et al., 2015). In fungi, *Saccharomyces cerevisiae* encodes a complete set of tRNAs, but in trypanosomatides (e.g. *Trypanosoma brucei*, *Leishmania tarentolae*) and alveolates such as Plasmodium, the mt-genomes can be completely devoid of tRNA genes (Salinas-Giegé et al., 2015). Interestingly, the mt-genomes of angiosperms contain chloroplast-like genes, as a replacement for tRNA[His] and tRNA[Asn] that were lost in all investigated angiosperms (Fey et al., 1997).

Organization and structures

The mt-genomes organization also greatly differs between eukaryotes. In metazoan, the genomes are quite reduced, with minimal to no intergenic regions, and no intron sequences in vertebrate mtDNA. All protein-coding genes and rRNAs are flanked by tRNAs (excepted for the COIII - ATP6 junction). This allows the production of large polycistrons which will be processed by pre-tRNA maturation enzymes, RNase P and Z, releasing individual processed mRNAs, rRNAs and tRNAs (Ojala et al., 1980). In plants, the genes are separated by large non-coding regions that are not conserved across species, contributing to the extensive size of plant mt-genomes. Those non-coding sequences are the result of horizontal gene transfer, most likely derived from chloroplastic, nuclear, or viral DNA (Gualberto and Newton, 2017). Several genes contain introns, mainly of group II, that are self-catalytic ribozymes (Cech, 1986). In Arabidopsis, 23 group II introns are present. Interestingly, Arabidopsis and all angiosperms possess the matR gene, encoded within

intron 4 of nad1. This gene encodes for the MatR maturase, a protein binding to several group II introns *in vivo*, but its putative roles in splicing are yet to be determined (Brown et al., 2014b; Unseld et al., 1997).

Mitochondrial genomes are usually represented as singular circular molecules. However, this is not always the case. In *Cucumis sativus*, the mitochondrial genome assembles into three circular chromosomes of different lengths (Alverson et al., 2011). In kinetoplastid protists such as *Trypanosoma,* the mitochondrial genome is composed of ~50 so called maxicircles and thousands of minicircles. Maxi- and mini-circles form a packed network of circular DNA that constitute the kinetoplast (called kDNA). This arrangement contributes to the specific posttranscriptional processing modification where uridines are inserted into, or deleted from, messenger RNA precursors (Aphasizhev and Aphasizheva, 2014). But mt-genomes are not always arranged into circular molecules, linear mt-genomes also exist. It has been described in the green-algae Chlamydomonas (Smith et al., 2010), Plasmodium, some fungi, and several cnidarian animals (Nosek and Tomáška, 2003). An unusual situation has been identified in a single-celled protist relative of animals, *Amoebidium parasiticum*, whose large mtDNA (>200 kbp) consists of several hundred linear chromosomes that share elaborate terminal-specific sequence patterns (Burger et al., 2003). In the case of linear mt-genomes, they are protected by specialized end-structures, such as covalently closed single-stranded DNA termini or protective proteins, and they also tend to have telomere-like repeats (Burger et al., 2003; Nosek and Tomáška, 2003).

Why is the genome retained?

The question remains open on why mitochondria (and chloroplast) retained a genome. It would certainly be more advantageous in term of energy if the cell would not need to have a second complete gene expression machinery in mitochondria, with a complete different DNA replication system, RNA polymerase and ribosomes, all this required to express a very small set of genes. Several models have been proposed to understand why mitochondria actually retained a genome.

One of the hypothesis is that the impossibility to transfer the remaining genes in the nucleus relies on the differences between the nuclear genetic code and the one used in mitochondria. Indeed, in animals, yeast and several protists (D.N.J. de Grey, 2005), the codon usage is different. For example, UGA is a 'stop' codon in the universal code and in plant mitochondria, but it codes for tryptophan in animals and fungi mitochondria. This is not the case in plants.

It was also proposed that the mitochondrial gene transfer is still an ongoing process and that we are only witnessing an intermediary phase of eukaryote-mitochondria evolution. Eventually all genes should/could be transferred to the nucleus – researchers are even trying, through genetic engineering, to artificially transfer mitochondrial genes to the nucleus, not without difficulties. For example, a nuclear encoded copy of cytochrome *b* fused to a mitochondrial targeting signal is not imported in mitochondria as it forms protein aggregates (Claros et al., 1995).

This lead to the hydrophobicity hypothesis which stipulates that the product of the few remaining genes are too hydrophobic to be synthesized in the cytoplasm and then imported into mitochondria (von Heijne, 1986). This is supported by the observation that the two large ribosomal RNAs, as well as the highly hydrophobic proteins cytochrome *b* and Cox1 are the ones universally still encoded in mt-genomes.

An alternative explanation is that the retention of these small but functional genomes persist because organellar gene expression must be under direct redox control (Allen, 2003). This hypothesis is termed CORR (co-location for redox regulation), and explains that genes in mitochondria and chloroplast must be under the direct regulatory control of the redox state of their gene products to allow the fine tuning of mitochondrial gene expression in response to metabolic changes. Such a mechanism has been described in yeast: Mss51, a translational activator of Cox1, is able to bind heme B. It could therefore sense oxygen levels to modulate Cox1 synthesis and its subsequent assembly into Complex IV, the major oxygen-consuming mitochondrial enzyme (Soto et al., 2012).

Mitochondrial gene expression:

Mitochondrial gene expression is a complex – patchy – mechanism, completely dependent on nuclear-encoded factors. Indeed, only a few genes involved in mitochondrial gene expression are encoded in the mt-genome, these being most frequently ribosomal RNA and ribosomal proteins required for the final step of gene expression. For their expression, mitochondrial genes must be transcribed, their RNAs then undergo a number of post-transcriptional maturations and they are translated. Thus the majority of the factors involved in these processes are encoded in the nucleus, expressed in the cytosol and imported into the mitochondria.

tRNAs, which are crucial for translation are also, in some organisms, imported from the cytosol. In human and yeast for example, tRNA import is not required as a full set of tRNAs is

present, but it does occur nonetheless (Salinas-Giegé et al., 2015). In contrast, Trypanosoma and Plasmodium mt-genomes have no tRNA genes, therefore they all have to be imported (Hancock and Hajduk, 1990; Salinas-Giegé et al., 2015). In plants, tRNAs have to be imported from the cytosol to ensure translation, as mt-genomes are incomplete. In Arabidopsis, 6 tRNAS must be imported (Salinas-Giegé et al., 2015; Salinas et al., 2008).

Due to the early evolutive radiation of eukaryote and the bacterial origin of mitochondria, gene expression in mitochondria combines bacterial-like features, inherited from the original endosymbiont, with eukaryote traits coming from the host cell (Adl et al., 2012; Gray, 2012). Moreover specific gene expression features evolved independently in the different groups of eukaryotes. As a result, gene expression mechanisms in mitochondria are unique and diverse between species.

Transcription

Gene expression starts with transcription to synthesize RNA. To fulfil mitochondrial transcription, most eukaryotes possess a nuclear-encoded phage-type RNA polymerase (mtRNAP or NEP for nuclear-encoded polymerase) which replaced the ancestral bacterial-type RNA polymerase (Liere and Börner, 2011). In contrast, in Jakobids, primitive protists of which *R.americana* belongs, the mitochondrial genome still encode a bacterial-type RNA polymerase (Lang et al., 1997).

In plants, several promoters are necessary for the transcription of the genome. In Arabidopsis mt-genome, among the 57 genes, five transcripts are produced in poly-cistronic form, three transcripts are trans-spliced and the others are monocistronic (Forner et al., 2007). Two nuclear genes code for two RNA polymerases, RpoTm is targeted to mitochondria and RpoTmp is targeted to both mitochondria and chloroplast, another one, RpoTp, is targeted only to chloroplast (Liere and Börner, 2011). RpoTm is involved in the transcription of the majority of the mt-genome, thus the role of RpoTmp is still poorly understood, but *in vitro* and *in vivo* studies showed that a subset of mitochondrial genes depend on RpoTmp for their expression (Kühn et al., 2009). In human, only two polycistronic transcripts are synthetized. The transcription starts from two promoters, a heavy strand promoter and a light strand promoter (Pearce et al., 2017), which will generate the two pre-transcripts that will later be processed. In yeast, several transcription start sites are found producing polycistronic precursor molecules encoding two or more coding sequences (Christianson and Rabinowitz, 1983).

In human and yeast transcription factors are required to fulfil transcription, contrary to bacteriophage T7 RNA polymerase that do not need any additional factors. In human, mtTFA and mtTFB2 as well as mtTEF play a critical role in mitochondrial transcription (Litonin et al., 2010; Minczuk et al., 2011). In yeast, mtTFA enhance transcription initiation but is not necessary (Liere et al., 2011). In Arabidopsis and other higher plants, the nuclear genomes encode homologues of fungal and animal mtTFB, but the involvement of these proteins in mitochondrial transcription could not be established (Kühn et al., 2009). It has been proposed that pentatricopeptide repeat (PPR) proteins might be involved in transcription initiation. For instance, mammalian mtRNAP contains PPR motifs and yeast mtRNAP associates with protein complexes including PPR proteins (Lightowlers and Chrzanowska-Lightowlers, 2013; Shadel, 2004).

Maturation and degradation

The regulation of mitochondrial gene expression is mostly performed at the post-transcriptional stage. Transcripts are first expressed as precursors and undergo the following steps of maturation: splicing, RNA editing, 5' and 3' maturation of transcript ends ... These processes are usually related to the prokaryotic processes, but several are also entirely specific to mitochondria, and most of these maturation mechanisms are performed by eukaryote specific proteins such as PPR proteins.

Definition of transcript ends

In the cytosol, to enhance their stability, mRNAs are capped at the 5' extremity and polyadenylated in 3' (Schaefke et al., 2018). This is not the case in organelles, where no cap is added in 5' of the mRNAs. Furthermore, polyadenylation in the cytosol stabilizes mRNAs, whereas in mitochondria the polyadenylation of coding and non-coding RNAs directs them towards degradation by the polynucleotide phosphorylase (PNPase) (Gagliardi and Leaver, 1999; Holec et al., 2006).

In plant mitochondria, 5' and 3' ends of mRNAs, as well as of tRNAs and rRNAs, go through several maturation steps to become functional. It is known that the maturation of 5' and 3' extremities of organelle RNAs involves several distinct ribonucleases. For the 3' extremities definition it was shown that PPR proteins (MTSF1 in the case of nad4 transcript) could bind the pre-messenger and potentially block the progression of the mitochondrial 3'-5' exonucleases (Haïli et al., 2013) such as the PNPase and RNR1 (Perrin et al., 2004). Concerning the 5' maturation of
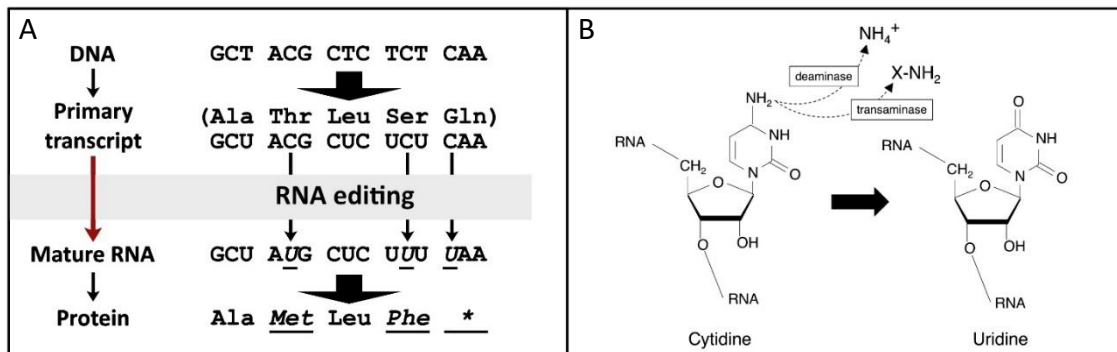
Figure 9: RNA editing in plant mitochondria
A RNA editing in flowering plants mitochondria is a post-transcriptional process which changes specific C-residues in the primary transcript, to U-residues in the mature mRNA. The amino acid sequence encoded by the fully edited, mature mRNA is different from the protein sequence encoded by the genomic DNA and the primary transcript. 441 editing sites are found in the mitochondrial mRNA and tRNA population of *A.thaliana.* RNA editing restores codons conserved throughout evolution.
B The biochemical reaction of RNA editing is a deamination. The amino group from the cytidine may be removed by a cytidine deaminase-like enzyme or may be transferred by a transaminase to a receptor molecule.
Derived from (Takenaka et al., 2008)

mRNAs, little is known, but an increasing number of reports also tend to point that PPR proteins may act as sequence-specific RNA-binding transcript "delimiters" (Hauler et al., 2013; Stoll et al., 2014). The enzymatic activity being performed by a yet unknown ribonuclease, even though two potential ribonucleases, MNU1 and MNU2, where proposed to be involved in this process (Stoll and Binder, 2016). In the case of the plant mitochondria tRNAs, they are first transcribed as pre-tRNAs which are then processed in order to become functional. First the 5' leader sequence is cleaved by a PRORP (Protein Only RNase P) enzyme (Gobert et al., 2010). The 3' trailer is cleaved by an RNase Z (Canino et al., 2009). Then multiple specific bases are modified, which allow the stabilization of the 3D structure and finally, the CCA triplet is added in 3' of tRNAs (Salinas-Giegé et al., 2015).

In the unicellular photosynthetic organism *Chlamydomonas reinhardtii,* the mitochondrial mRNAs are different from what is usually found in Viridiplantae. Indeed, similar to metazoan, mRNAS do not possess 5' UTR and directly start at the AUG initiation codon. It was recently discovered that a portion of mRNAs harbor C-rich 3' tails in addition of the A/U rich tails that are added post-transcriptionally at the 3'-extremity of all mRNAs (Cahoon and Qureshi, 2018; Salinas-Giegé et al., 2017).

In metazoan the transcripts extremities are defined by endonucleolytic cleavage. Indeed, the mt-genome is organized such as rRNAs, tRNAs and mRNAs are all immediately contiguous without non coding sequences, tRNAs "punctuating" the genome (Ojala et al., 1980). Therefore, the primary transcripts are processed by cleavage of the 5' and 3' termini of mitochondrial tRNAs, releasing rRNAs and mRNAs. This processing, is mediated by RNase P and RNase Z endonucleases (Bernt et al., 2013; Holzmann et al., 2008). Additionally, for the majority of human mitochondrial mRNAs (except nad6) the mitochondrial poly(A) polymerase (mtPAP) restore a UAA stop codon at the 3' extremities of the transcripts (Ojala et al., 1981; Tomecki et al., 2004).

RNA editing

RNA editing is a process where discrete changes are performed at specific sites within an RNA molecule after it has been transcribed. The editing events may include the insertion, deletion, and base substitution of nucleotides within the transcript, usually to restore a proper sequence for protein synthesis (Fig 9). In plant mitochondria the major type of RNA editing is the conversion of C-to-U that takes place at hundreds of sites (e.g. Giegé 1999) and more rarely U-to-C (Gualberto et al., 1989; Gutmann et al., 2012a). The full composition of the editosome complex have yet to be established, but several proteins have been identified. It has been proposed that PPR proteins
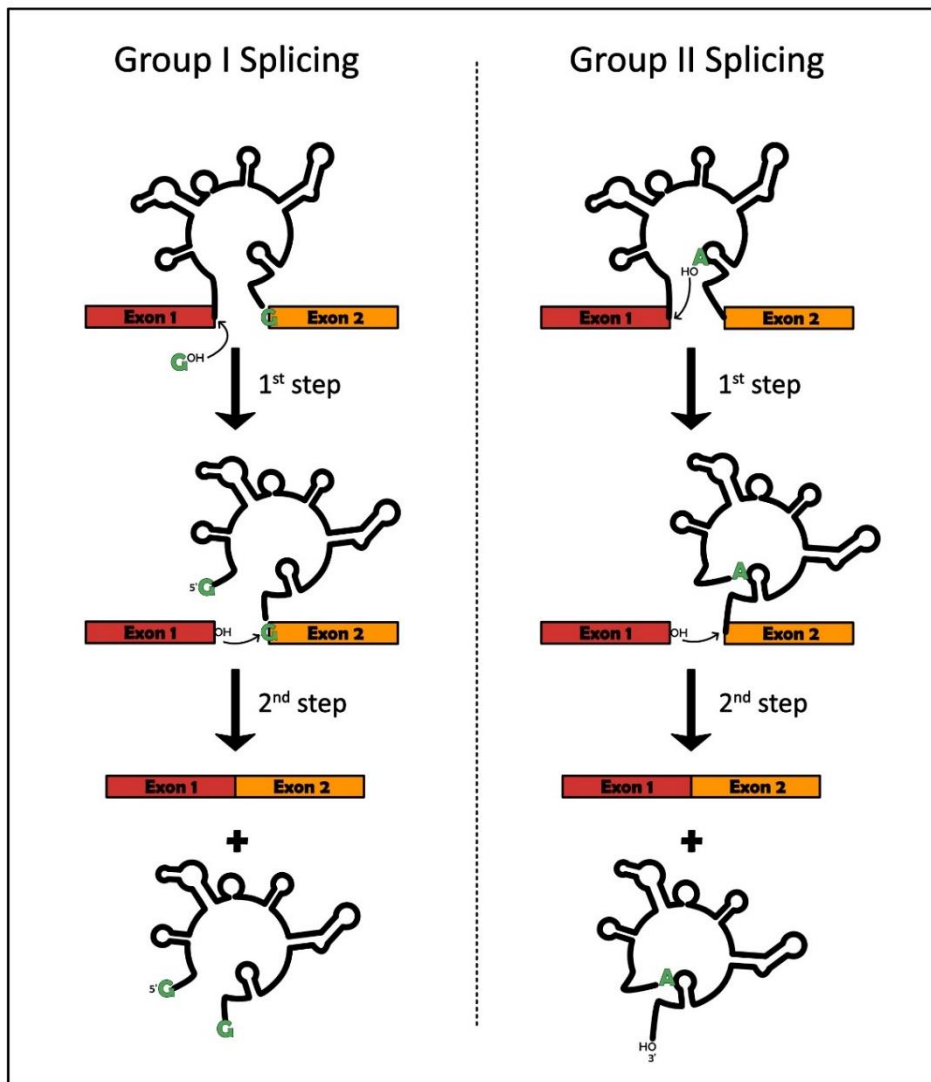
Figure 10: Mitochondrial introns
Schematic representation of Group I and group II introns splicing. Both groups are large self-splicing ribozymes. Both self-splicing introns perform two consecutive trans-esterification reactions in the process of exon ligation. The first step of splicing in a group I intron involves nucleophilic attack at the 5'-splice site by the 3'-OH of an exogenous guanosine cofactor. This reaction adds the guanosine onto the 5'-end of the intron and releases the 5'-exon. In the second step, the 5'-exon attacks the 3'-exon boundary, characterized by a G residue, which releases the intron and ligates the 5' and 3'-exons. The result of these reactions is that the flanking exons are ligated and the intron is released as a linear molecule with an uncoded G at the 5'-end.
Unlike the group I introns, group II introns utilize an internal nucleotide for the first step of splicing. It is the 2'-OH of a highly conserved bulged A nucleotide of the intron that attacks the 5'-splice site. This results in the release of the 5'-exon and formation of a lariat structure whose 5'-end of the intron is covalently attached to the 2'-OH of the bulged A. The second step is similar to the Group I introns, but there is no G at the 3'-intron junction.

might specify the editing site, as the PLS-class PPR proteins do possess the additional E domain that has been shown to be essential for editing. They can also possess a DYW domain, which exhibits some similarity to cytidine deaminase domains found in other systems. Therefore, it was proposed to be the catalytic factor for plant editing. However, no deaminase activity has been detected through expression of recombinant DYW-type PPR editing factors. Other factors have also been identified, like the RIP/MORF editing factors, where mutagenesis of a single gene encoding a RIP/MORF affects dozens or even hundreds of editing sites (Bentolila et al., 2012). More recently ORRM and OZ protein families have also been shown to be involved in RNA editing (Sun et al., 2016).

In *T.brucei* mitochondria another type of editing is observed, where Us are inserted into, or deleted from, messenger RNA precursors. This editing process contribute to restore accurate open reading frame by correcting frameshifts, introducing start and stop codons, and often adding most of the coding sequence. For this process, the editing positions and the number of Us deleted or inserted are specified by guide RNAs hybridizing to the pre-mRNAs (Aphasizhev and Aphasizheva, 2014).

Splicing

Depending on the organism, mitochondrial transcripts may contain introns. The two types of introns found in mitochondria are of group I and group II (Fig 10). Contrary to what is found in the nucleus, mitochondrial introns are auto-catalytic ribozymes both being able to self-splice *in vitro* (Cech, 1986) but require proteins for efficient splicing *in vivo*, to stabilize the catalytically active RNA structure. The presence and prevalence of each type of introns is variable between eukaryotes. In animals for example, mitochondrial introns were only described in basal metazoans like sponges or cnidarians (Huchon et al., 2015). In humans, no intron is found. More generally group II introns are less frequent than group I introns, except in land plants, where they are prevalent.

Indeed in plant mt-genomes, all introns are group II except for a single group I intron present in the cox1 gene of some plants, most likely acquired through horizontal gene transfer (Cho et al., 1998). In Arabidopsis on top of the classical introns, three trans-introns are found in nad1, nad2 and nad5 mRNAs. Compared to normal splicing, trans-splicing generates a single RNA transcript from multiple separate pre-mRNAs (Malek and Knoop, 1998). PPRs proteins have also been found to be involved in splicing events in plants (Brown et al., 2014b).
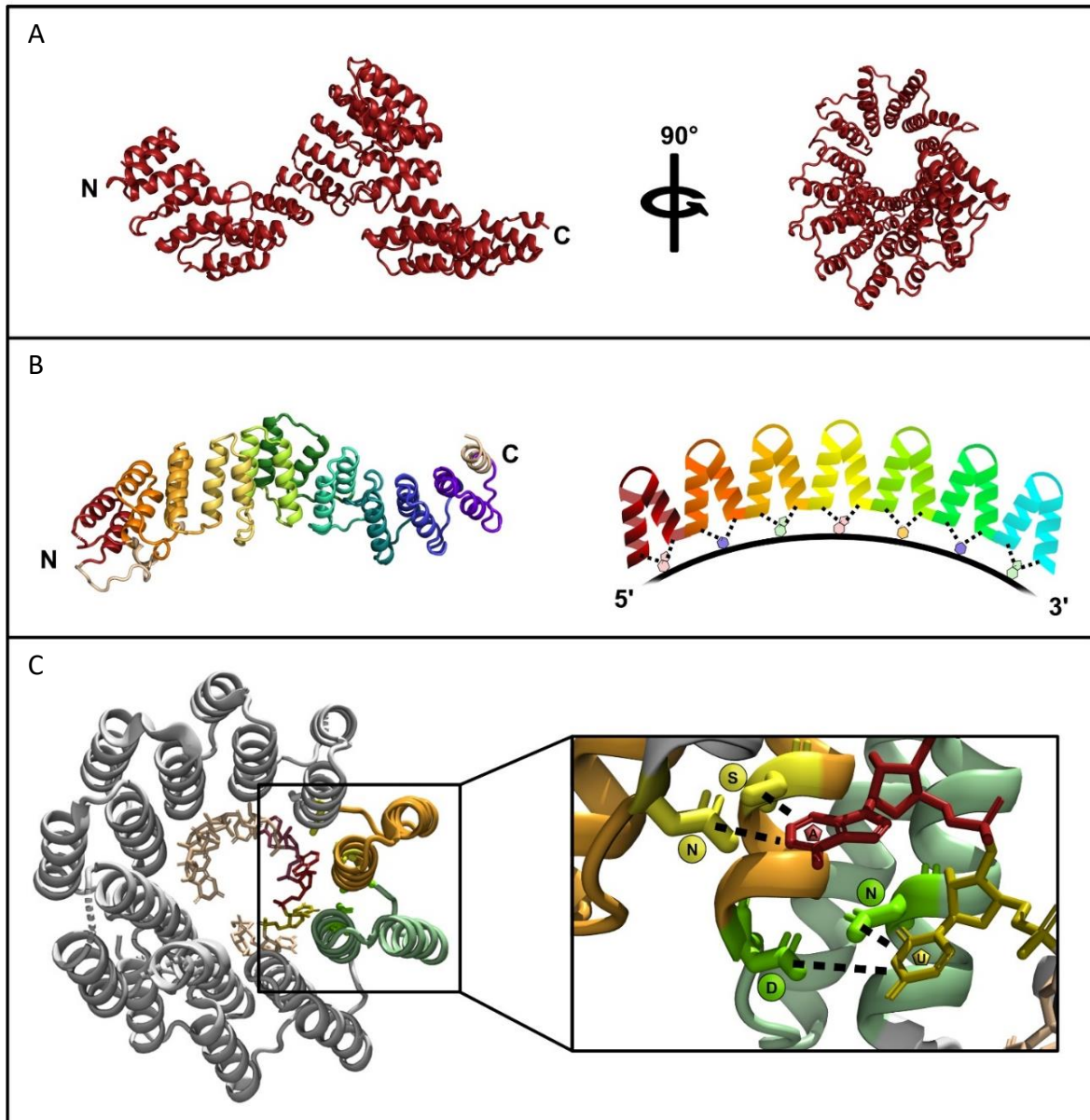
Figure 11: PPR proteins structure and binding mechanism
A Overall structure of a P-class PPR protein. Each PPR motif folds into a pair of antiparallel alpha helices, which interact to produce a helix-turn-helix motif. The series of PPR motifs form a super-helix ribbon-like sheet, with a central groove that allows the protein to bind RNA.
B Each PPR motif is represented in a different color and is involved in the recognition of a specific ribonucleic base.
C The base recognition is mediated by two specific amino acids at position 5 and 35 of each motif. Here an Asn and a Asp mediate the recognition of a U-residue for the green motif and a Ser and an Asn mediate the recognition of a A-residue for the yellow motif. 5I9D (Shen et al., 2016)

PPR proteins

As described previously, a large majority, if not all, of the mitochondrial gene expression mechanisms involve PPR proteins, especially in plants. PPR proteins, for "pentatricopeptide repeat proteins", were discovered in 2000 when the sequencing of *A.thaliana* genome was achieved (Aubourg et al., 2000; Small and Peeters, 2000), revealing more than 450 protein-coding genes for PPR proteins, the vast majority of them are predicted to be targeted to mitochondria or chloroplast (Colcombet et al., 2013). PPR proteins are composed of multiple PPR motifs, each one being 35 amino acids long (hence the name pentatricopeptide), occurring as tandem arrays (Fig 11). Most PPR proteins have sequence-specific RNA-binding activity, which seems to be specified by a so-called PPR-code, where each repeat recognizes one specific ribonucleotide (Barkan and Small, 2014; Barkan et al., 2012). The PPR proteins are involved in RNA binding, which was largely confirmed by *in vitro* and *in vivo* experiments, including gel shift, UV cross-linking and affinity assays (Prikryl et al., 2011; Shikanai and Okuda, 2011). Based on the distribution of PPR proteins in extant organisms, it seems likely that they arose from the TPR (tetratricopeptide repeat) protein family, a class of protein-protein interaction motifs (Small and Peeters, 2000).

Classes of PPRs

There is mainly two types of PPR containing proteins. The P-class PPR proteins, for Pure PPR, are only composed of canonical PPR repeats which are 35 aa long. Variants of the classical P motif also exist. Indeed, these motifs are the PPR-like S (Short) and PPR-like L (Long) which are respectively 31 and 35-36 amino acids (Cheng et al., 2016; Lurin et al., 2004). PPRs composed of P, L and S motifs constitute the PLS-class PPR proteins where triplets of P, L, and S motifs are repeated (Fig 12).

Additionally these two groups can be divided into several subgroups. For the PLS-class proteins, they almost always harbor at their C-terminal extremity additional domains denoted E or DYW which are specific to this class of proteins (Lurin et al., 2004). These domains were shown to be implicated in RNA editing in plant organelles (Okuda et al., 2007, 2009) and display a conserved signature similar to the catalytic-site of known cytidine deaminases (Salone et al., 2007), which is a the catalytic basis for C-to-U editing. The E domain seems to be essential for editing, whereas the terminal DYW domain is often facultative, and may be recruited/act in *trans* (Boussardon et al., 2012). Concerning P-class PPRs, they can also possess additional C-terminal domains, but completely unrelated to E or DYW domains. For example, the eukaryote specific PRORP proteins (for Protein Only RNase P) contain a stretch of PPR repeats involved in tRNA
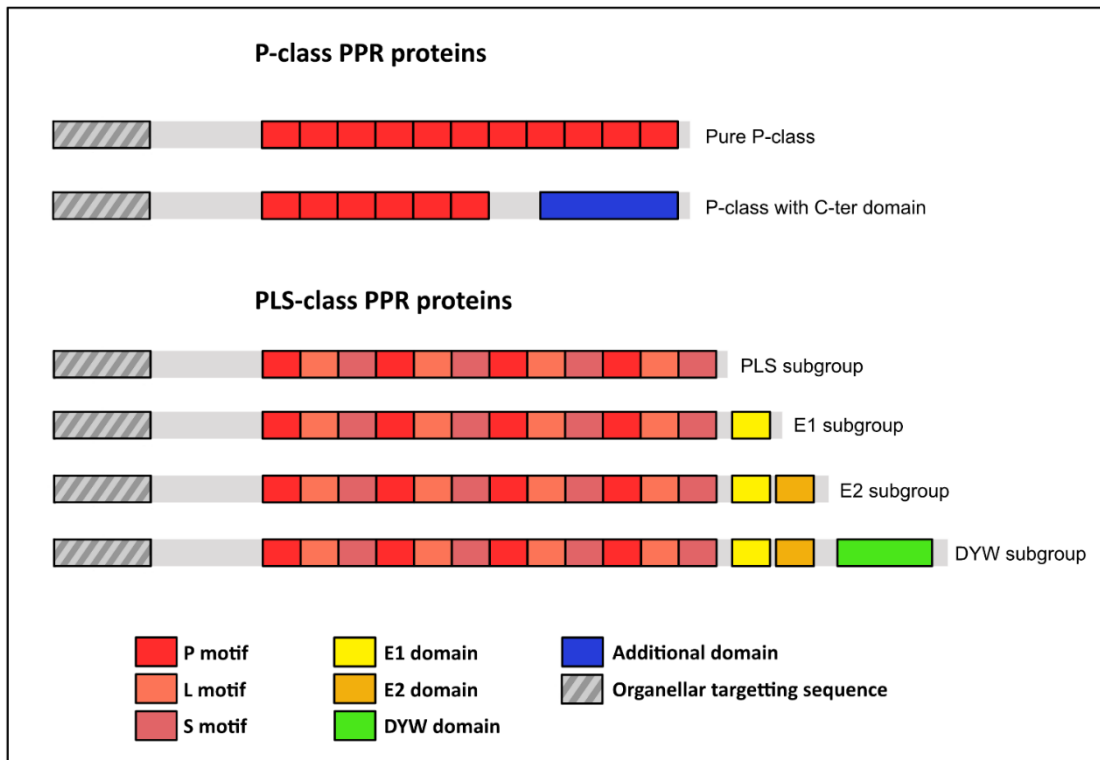
Figure 12: The different classes of PPR proteins

PPR proteins are composed of successions of PPR motifs and the number of motifs in each protein can vary from 2 to 35. The PPR proteins can be subdivided into two major groups. The P-class is composed of PPR proteins harboring only P (for pure) PPR motifs. Some of these proteins have additional non-PPR domains, e.g. PRORP proteins (Gobert et al., 2010). The PLS-class is composed of triplets of P, L (35 to 36 amino acids) and S (31 amino acids) motifs.

The PLS group can be further divided into subgroups. The two domains E1 and E2 are PPR-like motifs each composed of 34 aa which may contribute to RNA-binding and to base recognition. If E1 only is present, the protein belongs to the E1 subgroup and if E1 and E2 are present the protein belong to the E2 subgroup. In addition to the E1 and E2 domains, many PLS-class proteins possess a DYW domain which is proposed to be the catalytic domain implicated in the deamination of the target base. Some E2 subgroup proteins also have truncated DYW motifs which are termed E+.

Derived from (Cheng et al., 2016)

binding, coupled with a NYN catalytic domain for the removal of the pre-tRNA 5' leader sequence (Gobert et al., 2010; Schelcher et al., 2016). The human mitochondrial RNA polymerase POLRMT (Lightowlers and Chrzanowska-Lightowlers, 2013; Ringel et al., 2011) as well as the yeast mitochondrial RNA polymerase Rpo41 (Kruszewski and Golik, 2016) also harbor PPR motifs, but their molecular functions remain unclear.

Structure and mode of action

The amino acid sequence of each repeat is highly degenerate, which is not the case for their 3D structure. Indeed, each PPR repeat folds into a pair of antiparallel alpha helices, which interact to produce a helix-turn-helix motif. The series of PPR motifs form a super-helical ribbon-like sheet, with a central groove that allows the protein to bind RNA (Schmitz-Linneweber and Small, 2008). This typical structure is shared by proteins of the alpha-solenoid superfamily, like TPR repeats, armadillo repeats or HEAT repeat (Fournier et al., 2013). One particularity of PPRs proteins is that they are modular proteins, similarly to the DNA-binding transcription activator–like (TAL) motifs and the RNA-binding PUF, or the HAT, OPR, mTERF, … in which the one-repeat:one-nucleotide binding is dictated by specific amino acids is each repeat (Filipovska and Rackham, 2012; Hammani and Giegé, 2014).

For a long time the helix-turn-helix structure of the PPR repeat was only a hypothesis. The insoluble nature of PPR proteins impeded their expression, and the study of their structure and mechanism of RNA recognition (Manna, 2015). It was only in 2011, with the crystal structure of the mammalian mitochondrial RNA polymerase, that the first observation of the PPR fold was achieved experimentally (Ringel et al., 2011). In 2012, the crystal structure of Arabidopsis PRORP1 confirmed the role of the PPR repeats in pre-tRNA binding (Howard et al., 2012). The high resolution of *Zea mays* PPR10 and *Brachypodium distachyon* THA8 in both free and RNA-bound states were the first to give structural insights into the mode of ssRNA recognition by PPRs (Ke et al., 2013; Yin et al., 2013).

PPRs bind to specific ssRNA target in a one-repeat:one-nucleotide mode of action (Fig 11 B). This recognition is mediated by a so-called "PPR code" in which two amino acids in each repeat specify the nature of the nucleotide to bind. Positions 5 and 35 (in the latest nomenclature) seem to be the ones playing the most important role in nucleotide recognition (Cheng et al., 2016). For example, threonine at position 5 with asparagine at position 35 will recognize adenine, whereas an asparagine and aspartate at these respective positions would specify uracil (Fig 11 C)(Barkan et al., 2012). However the code is degenerate, with several combinations of amino acids specifying
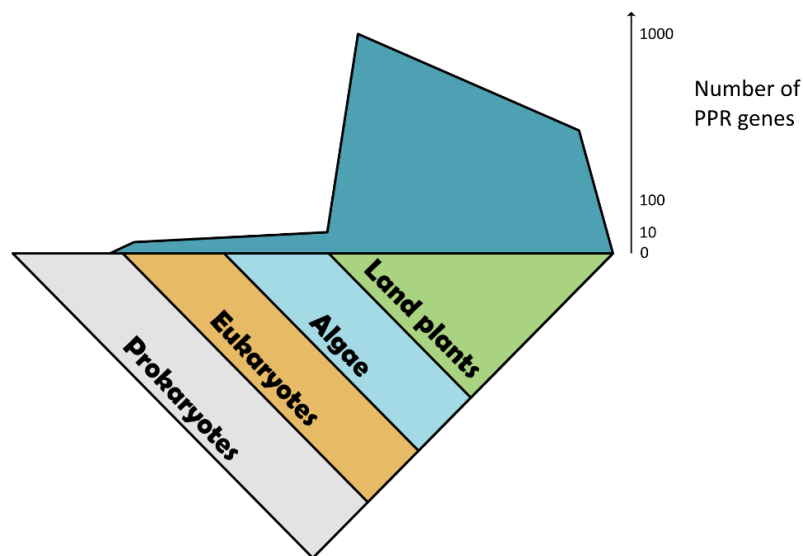
Figure 13: PPR repartition across eukaryotes
While the genes encoding PPR proteins are absent in prokaryotes (although there are some exceptions of plant pathogens who acquired PPR genes by HGT) and in small numbers in eukaryotes in general, PPR genes are present in great numbers in land plants. See Table 2. (Cheng et al., 2016; Lurin et al., 2004)

Organism	Number of PPR genes
R.solanacearum	1
H.sapiens	7
S.cerevisiae	15
C.reinhardtii	14
K.flaccidum	64
P.patens	133
S.moellendorffii	1568
A.thaliana	496
O.sativa	475
V.vinifera	534

Table 2: Number of PPR genes in different organisms
Ralstonia solanacearum is one of the few prokaryotes that encodes a PPR protein. It was likely acquired through horizontal gene transfer. Most of the eukaryotes encode between 5 to 15 PPR proteins as represented for Human and yeast here. The expansion of PPR genes in the green phylum correlates with the colonization of land. Interestingly *Klebsormidium flaccidum*, which represents a transition species from aquatic algae to land plants, encodes 64 P-class PPRs, more than most other algae. *Selaginella moellendorffii*, a very basal land plant, encode a huge number of PPR proteins, mainly PLS-class PPR, involved in RNA editing. In gymnospermae and above, the numbers of PPR proteins range from 400 to 600 in regular diploid organisms (Cheng et al., 2016; Herbert et al., 2013; Jalal et al., 2015; Lightowlers and Chrzanowska-Lightowlers, 2013).

the same nucleotide, and the same combination of amino acids sometimes being compatible with more than one nucleotide (Barkan and Small, 2014). Moreover additional positions may influence the binding mechanism (Yagi et al., 2013). The modular and specific mode of nucleotide recognition by PPR proteins could be of great use to engineer new tools to target RNA of interest, thus multiple studies were conducted to test the PPR code and create artificial PPRs to target new RNAs (Coquille et al., 2014; Miranda et al., 2018; Shen et al., 2016).

Repartition across eukaryotes

PPR proteins are eukaryotic specific at the exception of a few bacteria that contain PPR proteins. These bacteria are all pathogens or symbionts of eukaryotes which most likely acquired PPR-coding genes via horizontal transfer (Cazalet et al., 2010). This widespread repartition of PPRs across eukaryotes set the acquisition time of PPRs very early in eukaryotic evolution, most likely linked with the acquisition of mitochondria (Fig 13).

Most eukaryotes contain a small number of PPR-coding genes, but in the land-plant lineage the family has greatly expanded (O'Toole et al., 2008). Indeed, 7 PPR proteins are found in human (Lightowlers and Chrzanowska-Lightowlers, 2013), about 28 in *T.brucei* (Pusnik et al., 2007)*,* 15 in *S.cerevisiae* (Herbert et al., 2013a) (Table 2). In the green phylum, land colonization seems to correlate with the explosion of PPR-gene numbers. Indeed most aquatic green algae have about the same number of PPR genes as the rest of eukaryotes, contrary to land plants that have 100 or more PPR genes. There are around 400–600 PPR genes in most angiosperm genomes, *A.thaliana* having more than 450 PPR genes (Cheng et al., 2016; O'Toole et al., 2008). In more basal plants, the numbers vary greatly with more than 1500 PPR genes in *Selaginella moellendorfii*, which correlates with its high number of RNA editing sites (Banks et al., 2011).
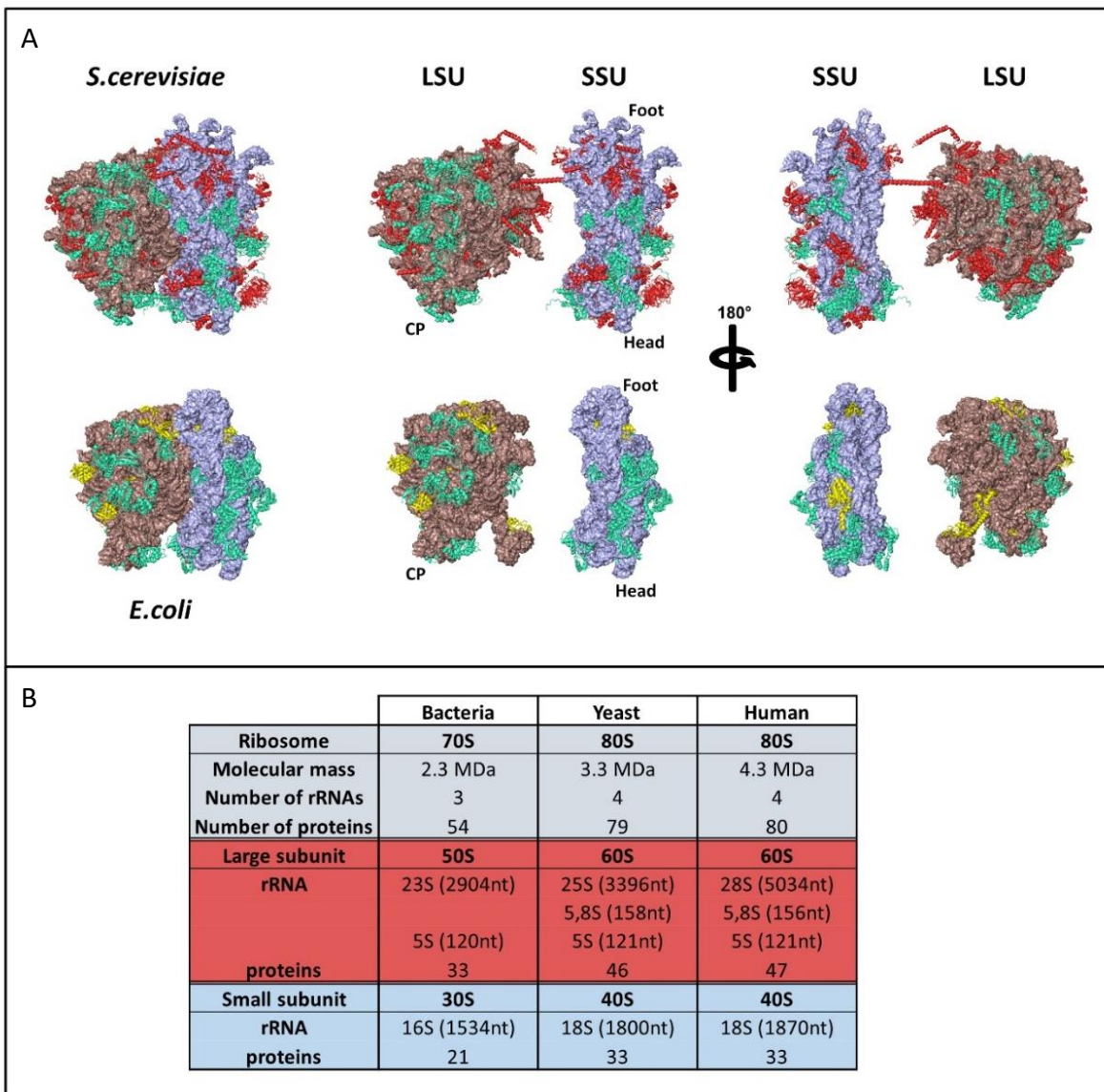
Figure 14: Eukaryotic and prokaryotic ribosomes structure and composition
A Structural comparison of the *E.coli* ribosome and the *S.cerevisiae* ribosome. LSU represent the large ribosomal subunit, SSU the small ribosomal subunit and CP the LSU central protuberance. The characteristics of these different ribosomes and of the human one are described in **B**. The eukaryotic ribosomes are larger compared to the prokaryotic ones. This is due to an enlargement of the ribosomal RNA and to the acquisition of several eukaryote-specific r-proteins. rRNA of the large subunit are shown in dark red, rRNA of the small subunit are shown in light blue. Universal r-proteins are shown in cyan, bacteria specific are shown in yellow, eukaryotic and archeal specific r-proteins are shown in red (Melnikov et al., 2012).
S.cerevisiae 4V88 (Ben-Shem et al., 2011), *E.coli* 4YBB (Noeske et al., 2015)

Protein synthesis

Generalities

Protein synthesis, or translation, is the final process whereby the intermediary genetic information (mRNA) is read to produce proteins. Translation is carried out by ribosomes, ubiquitous molecular machines that read mRNAs and catalyze the assembly of amino acids to form a polypeptide chain. The genetic code, by which triplets of nucleotides (codons) specify an amino acid is mediated by the tRNAs. Ribosomes therefore facilitates the recognition of the tRNAs anti-codons with the codons on the mRNAs and hold the peptidyl transferase activity to link the amino acids brought along with the tRNAs (Lafontaine and Tollervey, 2001).

Ribosomes are complex molecular machines found within all living cells. They are organized in two major functional part: the small ribosomal subunit (SSU), which reads the mRNA, and the large subunit (LSU), which contains the ribosomal catalytic site termed "peptidyl transferase center" (PTC), which joins the amino acids to form a polypeptide chain (Melnikov et al., 2012). Each of these subunits are huge ribonucleoprotein complexes, composed of one or more ribosomal RNA (rRNA) and a variety of ribosomal proteins (r-protein). They were first observed in the 50s by **George Emil Palade**, using electron microscopy, and were characterized at the time as "dense granules" (Palade, 1955). This discovery won him, jointly with Albert Claude and Christian de Duve, the 1974 Nobel Prize in Physiology or Medicine.

Even if ribosomes are ubiquitous to all living organisms, their structures and compositions, along with the factors that contribute to translation, are strikingly different (Fig 14). Ribosomes are particularly divergent between prokaryotes and eukaryotes, but also between bacteria and archaea (Eme et al., 2017). In the case of the SSU, prokaryotic SSU contains the 16S rRNA and 21 r-proteins (in *E.coli*), and more proteins are found in eukaryotes-related archaea. The eukaryotic SSU contains the 18S rRNA and 33 r-proteins (Melnikov et al., 2012). This number can vary between species, for example an additional kinetoplastid-specific r-protein was recently identified in *T.brucei* (Brito Querido et al., 2017). In the case of the LSU, prokaryotes have two rRNAs, the 5S and 23S, and 33 r-proteins (in *E.coli*), similarly to the SSU more proteins are found in archaeal organisms (Eme et al., 2017). The eukaryotic large subunit contains three rRNAs the 5S, 5.8S and 25S/28S (ranging from 3396 to 5034nt) along with 46-47 r-proteins (46 in yeast and 47 in animals). Therefore, compared to prokaryotes, eukaryotic ribosomes acquired on average 25 r-proteins and possess an additional rRNA in the LSU. Several r-proteins and rRNAs domains are conserved across life, forming a conserved functional-core (Melnikov et al., 2012). Aside mito- and chloro-
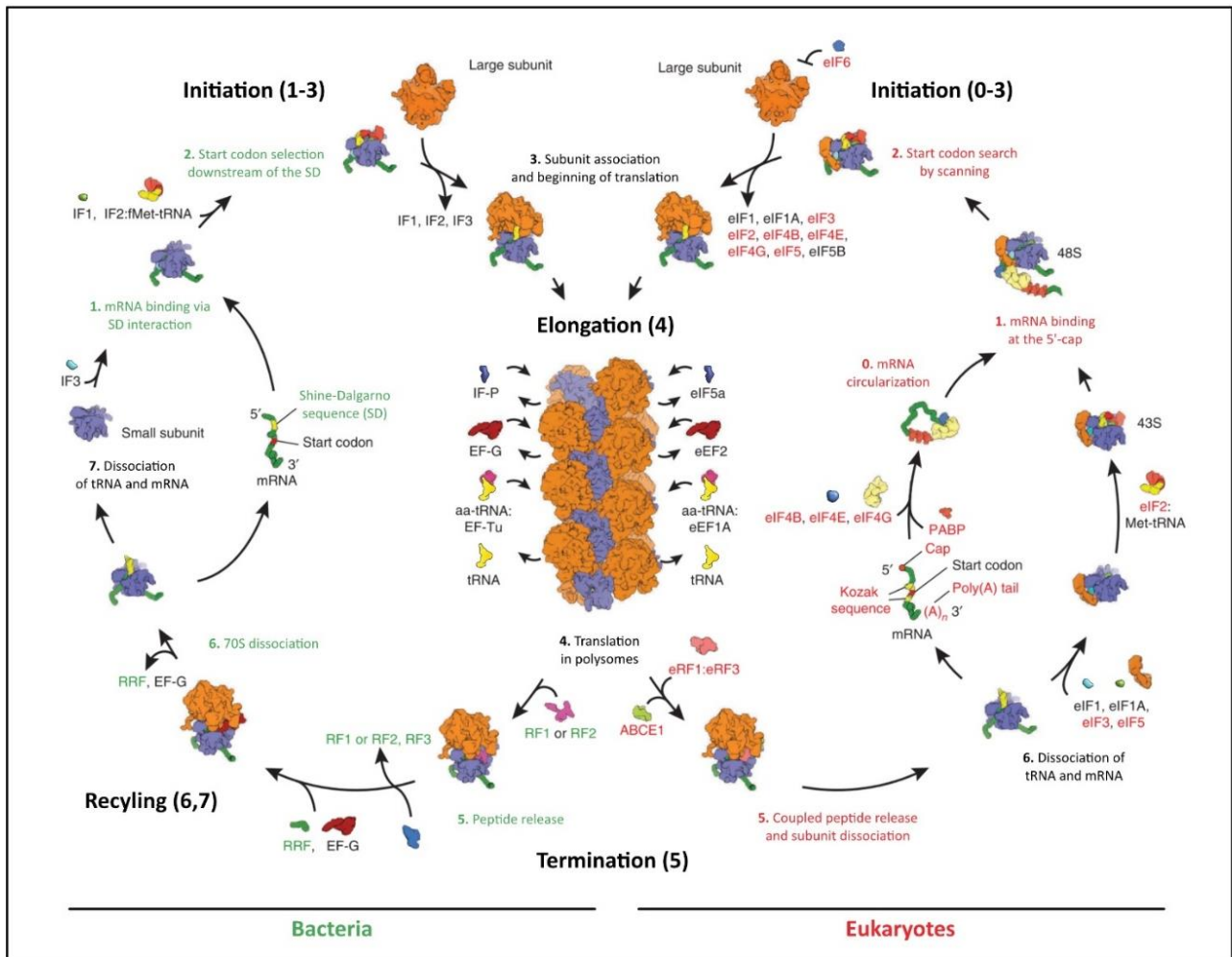
Figure 15: The processus of translation in eukaryotes and prokaryotes
Schematic representation of the different steps of translation – initiation, elongation of the polypeptide chain, termination and recycling of the ribosomes – with the different factors involved, compared between eukaryotes and prokaryotes.
Derived from (Melnikov et al., 2012)

ribosomes, 15 SSU r-proteins and 18 LSU r-proteins are considered universal between eukaryotes and prokaryotes (Ban et al., 2014).

To understand the molecular mechanism of translation, great deal of effort were deployed to resolve the ribosome structure at high resolution. The earliest high-resolution structures were resolved by X-ray crystallography, which allowed to study the prokaryotic ribosomes. The work of **Venkatraman Ramakrishnan**, **Thomas A. Steitz** and **Ada E. Yonath** won them the Nobel Prize in Chemistry in 2009. Crystallography of ribosomes being particularly challenging, electron microscopy was also used, first giving only low resolution results (Dube et al., 1998; Verschoor et al., 1998), but with the evolution of cryo-electron microscopy higher resolution were achieved, along with the possibility to study the ribosomes in a more "natural" context as compared to crystallography (Ben-Shem et al., 2011; Hashem et al., 2013; Khatter et al., 2015; Kühlbrandt, 2014; Weisser et al., 2017).

The mechanism of translation

The translation mechanism can be decomposed in three main steps: Initiation, Elongation and Termination (Fig 15). During initiation, the small subunit of the ribosome, together with an initiator tRNA^{Met} will bind to the 5' of the mRNA where the initiation codon is localized (Dyson et al., 1993; Kolitz and Lorsch, 2010). Several factors, called initiation factors, are involved in this process. Once the small subunit is correctly positioned on the initiation codon, the large subunit is recruited and the elongation starts. The factors involved and the processes by which the ribosomes are recruited to the mRNAs differ between prokaryotes and eukaryotes (Myasnikov et al., 2009). For example, in prokaryotes the SSU binds to the mRNA thanks to a consensus sequence called the Shine-Dalgarno sequence, whereas in eukaryotes the recruitment of mRNA involves the 5' cap structure (Merrick, 2004; Yonath, 2009; Yusupova et al., 2006).

At the beginning of the elongation, the initiator tRNA is placed in the P site of the LSU, and the A site is ready to receive the first aminoacyl-tRNA. The A site and the P site, are two of the three active sites of the ribosome, along with the E site. The **A** site is the point of entry for the new **a**minoacyl tRNA. The **P** site is where the **p**eptidyl tRNA is formed in the ribosome. And the **E** site is the **e**xit site of the uncharged tRNA after it transferred its amino acid to the growing peptide chain. During elongation, each additional amino acid is added to the nascent polypeptide chain in a three-step microcycle. First the correct aminoacyl-tRNA is positioned in the A site of the ribosome, then the peptide bond is formed from the P to the A site tRNA. Once the peptide bond is formed, the ribosome move onward by exactly one codon. This shift allows the uncharged tRNA to drift

out via the E site. It also exposes a new codon in the A site, and the whole cycle repeats. This whole process rely on elongation factors, which are also different between eukaryotes and prokaryotes.

Translation ends with termination. This step is triggered when a stop codon (UAA, UAG, or UGA) enters the A site. These termination codons are not recognized by any tRNAs. Instead, they are recognized by proteins called release factors which activate the hydrolysis of the ester bond on the peptidyl-tRNA to release the newly synthesized protein from the ribosome. To finish, the two subunits of the ribosome dissociate from the mRNA and from each other, to be reused in another round of translation.

In eukaryotes, additional ribosomes are found. They are localized in organelles, mitochondria and chloroplast, to perform the translation of the mRNAs still encoded by the organellar genomes. While chloroplast ribosome strongly resemble bacterial ribosomes (Boerema et al., 2018), mitochondrial ribosomes, or mitoribosomes, that are the subject of this thesis, diverged significantly from their bacterial counterparts during eukaryote evolution as described in details hereafter.

The mitochondrial ribosomes:

Mitoribosomes are evolutionarily derived from the ancestral α-proteobacterial ribosomes, inherited from the original endosymbiont. However, early biochemical studies as well as recent high-resolution structures showed that they have strongly diverged from their bacterial counterpart in terms of composition, function, and structure (Bieri et al., 2018; Brown et al., 2017; Desai et al., 2017; Greber et al., 2015). Moreover, not only have mitochondrial ribosomes dramatically diverged from the bacterial ones, but the evolutionary drift has also produced very different mitoribosomes across eukaryotic lineages. Indeed, even if certain mito-specific components were acquired early during the evolution of eukaryotes and are most likely shared by all mitoribosomes, a portion of the mitoribosomes components are specific to certain groups of eukaryotes. For instance, certain r-proteins are specific to yeast, others to animals and others to plants (Bonen and Calixte, 2006; Goldschmidt-Reisin et al., 1998; Kitakawa et al., 1997; Koc et al., 2000). Overall, the organization and regulation of protein synthesis in mitochondria rely on a highly degenerate prokaryotic scaffold animated by a large number of host-derived co-evolved factors.
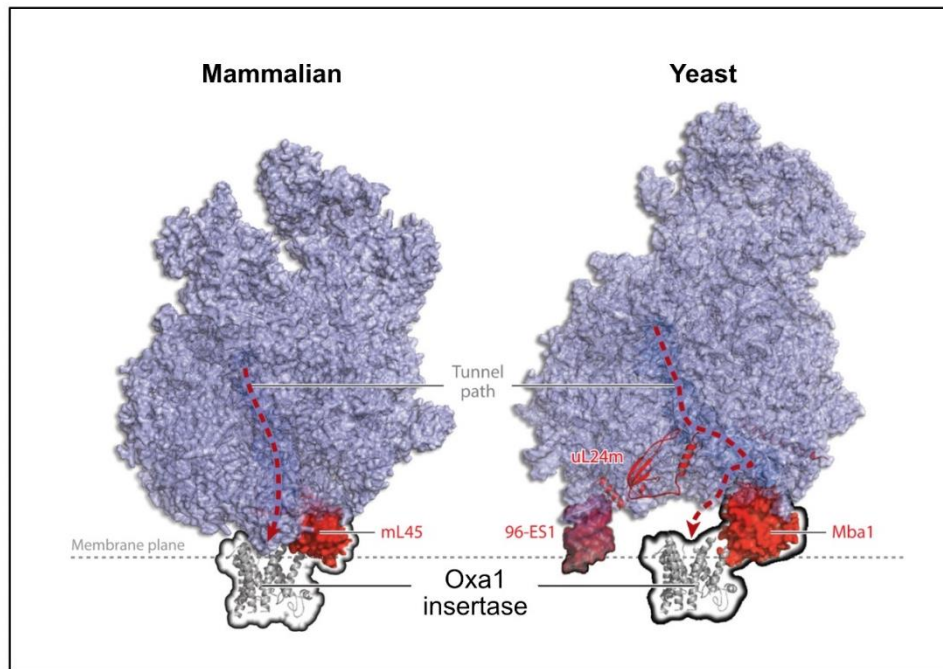
Figure 16: The mitochondrial ribosomes are bound to the inner membrane of mitochondria

Models of the yeast and human mitoribosomes membrane attachment. In both cases the protein mL45/Mba1, shown in red, is involved in the association with Oxa1, the mitochondrial insertase. In yeast the expansion segment 96 of the 21S rRNA seems to contact the membrane hence stabilizing the interaction.

Derived from (Ott et al., 2016)

Adaptation to a membrane-bound system

Only a few number of genes are still retained in mitochondrial genomes. The protein-coding genes are comparatively conserved across eukaryotes, those genes being mostly components of the respiratory chain, at the exception of plants and jakobids which also encode mitochondrial r-proteins along with other proteins (Gray, 2015; Lang et al., 1997; Marienfeld et al., 1999). For example, the two proteins universally present in mt-genomes are *cob* and *cox1,* encoding proteins of the Complex III and IV respectively (Table 1) (Gray, 2015). These respiratory chain components are highly hydrophobic membrane proteins (Sousa et al., 2018). Therefore, mitochondrial ribosomes seem to have evolved to synthesize predominantly, or even exclusively, these highly hydrophobic components of the mitochondrial respiratory chain. Indeed it has been shown, for animals and yeast, that the mitochondria-encoded membrane proteins are co-translationally inserted into the IMM. As a result, mitoribosomes are almost permanently bound to the inner membrane of mitochondria (Bieri et al., 2018; Greber and Ban, 2016; Ott and Herrmann, 2010; Ott et al., 2016; Pfeffer et al., 2015). For the mammalian mitoribosome, this attachment is most likely mediated by the mitochondria-specific protein mL45, located next to the ribosomal tunnel exit, which extends a C-terminal helix into the membrane (Bieri et al., 2018; Liu and Spremulli, 2000). In yeast, Mba1, the homolog protein of mL45, as well as an rRNA fragment, the expansion segment of helix 96, are involved in the mitoribosome attachment to the IMM (Fig 16) (Bieri et al., 2018; Ott et al., 2006; Pfeffer et al., 2015). In both mammals and yeast the nascent protein is thought to be co-translationally inserted into the inner membrane by the insertase Oxa1, thus constituting the central component of IMM protein insertion machinery (Bieri et al., 2018; Haque et al., 2010; Szyrach et al., 2003). This attachment may not be conserved across all eukaryotes. For example in Reclinomomas, the nature of the proteins encoded in mitochondria are much more diverse than in any other eukaryote, therefore mitochondrial ribosomes might not necessary be bound to the IMM (Lang et al., 1997). In plants, the difficulty to purify soluble mitoribosomes also suggested that they are bound to the IMM even if several non-membrane proteins are also encoded by the plant mt-genomes.

Mitoribosomes compositions

Mitoribosomes are evolutionarily derived from ancestral bacterial ribosomes, therefore their composition and structure rely on a prokaryotic scaffold. Indeed, biochemical and initial structural studies show that in spite of their evolutionary origin, mitoribosomes have dramatically diverged from bacterial ribosomes and between different eukaryotic lineages. The common

Table 3: Mitoribosomes composition

A

	Yeast	Animal	Trypa		Bacteria
Ribosome	74S	55S	?		70S
Molecular mass	3.0 MDa	2.7 MDa	4.5 Mda		2.3 MDa
Diameter	330 A	320 A	385 A		260nm
RNA:protein	1:1	1:2	1:6		2:1
Number of rRNAs	2	3	2		3
Number of proteins	80	82	127		54
Large subunit	54S	39S	?		50S
rRNA	21S (3296 nt)	16S (1569nt) / CP tRNA (73–75 nt)	12S (1176nt)		23S (2904nt) / 5S (120nt)
proteins	46	52	70		33
Small subunit	37S	28S	?		30S
rRNA	15S (1649 nt)	12S (962 nt) 30	9S (620nt)		16S (1534nt)
proteins	34	30	57		21

B

Bacterial homologs

Bact.	Mamm.	Yeast	Trypa.
bS1	bS1m	bS1m	–
uS2	uS2m	uS2m	–
uS3	uS3m	uS3m	uS3m
uS4	–	uS4m	–
uS5	uS5m	uS5m	uS5m
bS6	bS6m	bS6m	bS6m
uS7	uS7m	uS7m	–
uS8	–	uS8m	uS8m
uS9	uS9m	uS9m	uS9m
uS10	uS10m	uS10m	uS10m
uS11	uS11m	uS11m	uS11m
uS12	uS12m	uS12m	uS12m
uS13	–	uS13m	–
uS14	uS14m	uS14m	uS14m
uS15	uS15m	uS15m	uS15m
bS16	bS16m	bS16m	bS16m
uS17	uS17m	uS17m	uS17m
bS18	bS18m	bS18m	bS18m
uS19	–	uS19m	uS19m
bS20	–	–	–
bS21	bS21m	bS21	bS21m
bTHX	–	–	–

Bact.	Mamm.	Yeast	Trypa.
uL1	uL1m	uL1m	–
uL2	uL2m	uL2m	–
uL3	uL3m	uL3m	uL3m
uL4	uL4m	uL4m	uL4m
uL5	–	uL5m	uL5m
uL6	–	uL6m	–
bL9	bL9m	bL9m	bL9m
uL10	uL10m	uL10m	uL10m
uL11	uL11m	uL11m	uL11m
bL12	bL12m	bL12m	bL12m
uL13	uL13m	uL13m	uL13m
uL14	uL14m	uL14m	–
uL15	uL15m	uL15m	uL15m
uL16	uL16m	uL16m	uL16m
uL18	uL18m	–	–
bL19	bL19m	bL19m	bL19m
bL20	bL20m	–	bL20m
bL21	bL21m	bL21m	bL21m
uL22	uL22m	uL22m	uL22m
uL23	uL23m	uL23m	uL23m
uL24	uL24m	uL24m	uL24m
bL25	–	–	–
bL27	bL27m	bL27m	bL27m
bL28	bL28m	bL28m	bL28m
uL29	uL29m	uL29m	uL29m
uL30	uL30m	uL30m	uL30m
bL31	bL31m	bL31m	bL31m
bL32	bL32m	bL32m	bL32m
bL33	bL33m	bL33m	bL33m
bL34	bL34m	bL34m	–
bL35	bL35m	bL35m	bL35m
bL36	bL36m	bL36m	bL36m

Mitochondria specific

Bact.	Mamm.	Yeast	Trypa.
	mS22	–	mS22
	mS23	mS23	mS23
	mS25	–	–
	mS26	mS26	mS26
	mS27 *	–	–
	mS29 (DAP3)	mS29	mS29
	mS31	–	–
	mS33	mS33	mS33
	mS34	–	mS34
	mS35	mS35	mS35
	mS37	mS37	mS37
	mS38	mS38	mS38
	mS39 *	–	–
	mS40	–	
–		mS41	mS41
–		mS42	mS42
–		mS43	mS43
–		mS44	–
–		mS45	–
–		mS46	–
–		mS47	mS47
–		–	mS48
–		–	mS49
–		–	mS50
–		–	mS51 *
–		–	mS52
–		–	mS53
–		–	mS54 *
–		–	mS55 *
–		–	mS56
–		–	mS57
–		–	mS58
–		–	mS59
–		–	mS60
–		–	mS61
–		–	mS62 *
–		–	mS63
–		–	mS64
–		–	mS65
–		–	mS66
–		–	mS67
–		–	mS68
–		–	mS69
–		–	mS70
–		–	mS71
–		–	mS72
–		–	mS73
–		–	mS74

Bact.	Mamm.	Yeast	Trypa.
	mL37		
	mL38	mL38	mL38
	mL39	–	–
	mL40	mL40	mL40
	mL41	mL41	mL41
	mL42	–	mL42
	mL43	mL43	mL43
	mL44	mL44	–
	mL45	–	
	mL46	mL46	mL46
	mL48	–	
	mL49	mL49	mL49
	mL50	mL50	–
	mL51	–	
	mL52	–	mL52
	mL53	mL53	mL53
	mL54	mL54	–
		mL57	
		mL58	
		mL59	
		mL60	
		mL61	
	mL62 (ICT1)	–	
	mL63	–	mL63
	mL64 (CRIF1)	–	mL64
	mL65	–	
	mL66	–	
	–	MHR1	–
			mL67
			mL68
			mL69
			mL70
			mL71 *
			mL72 *
			mL73
			mL74
			mL75 *
			mL76
			mL77
			mL78
			mL79
			mL80
			mL81
			mL82
			mL83
			mL84
			mL85
			mL86
			mL87
			mL88
			mL89
			mL90
			mL91
			mL92
			mL93
			mL94
			mL95
			mL96
			mL97
			mL98
			mL99
			mL100

Table 3: Mitoribosomes composition
A Composition et characteristics of the mitochondrial ribosomes compared with the bacterial 70S ribosome.
B Protein composition of the mammalian (**Mamm.**), yeast and trypanosome (**Trypa.**) mitoribosomes compared with the bacterial one (**Bact.**). Proteins highlighted in light blue are universal to prokaryotes and eukaryotes, the ones in dark blue are specific to bacteria. The proteins in yellow are specific to mitochondria and shared by different organisms, the ones in red are specific to the specified organism.
(Amunts et al., 2015; Desai et al., 2017; Greber et al., 2015; Ramrath et al., 2018)

feature of mitochondrial ribosomes is that they contain a considerably increased number of r-proteins compared to the 54 found in *E.coli*. Proteomic analyses have helped to provide detailed catalogues of mitochondrial ribosomal proteins in *Saccharomyces cerevisiae* and mammals (Goldschmidt-Reisin et al., 1998; Kitakawa et al., 1997; Koc et al., 2000, 2001). In plants, such analysis was not performed and candidate bacterial-like ribosomal proteins of Arabidopsis and rice mitochondrial ribosomes were proposed based on sequence homology only (Bonen and Calixte, 2006). Some proteins of the bacterial core were lost, this is the case for bS20, that is absent from all eukaryote investigated, suggesting that it was lost early during eukaryote evolution (Smits et al., 2007). Among the additional mitochondrial r-proteins, some are conserved across eukaryotes (mS29, mS33, mL40, mL41 …) and seem to form a common core of mito-specific r-proteins that were surely acquired during early eukaryote evolution (Table 3). On top of these common proteins, additional r-proteins are found specifically in each eukaryotic lineages (Bieri et al., 2018; Greber and Ban, 2016; Ott et al., 2016). Man, pig, yeast and recently trypanosome are the only organisms for which high-resolution structure confirmed the protein composition of the mitoribosomes. Mammals and yeast respectively possess 36 and 30 ribosomal mito-specific r-proteins, which correspond to 45% and 37% of their total r-proteins. Among these, half are common to both yeast and mammals and the other half are specific to each of them. All these proteins form an extensive interaction network on the surface of the rRNA core and helped toward the specialization of the mitoribosomes (Bieri et al., 2018).

The mitochondrial rRNAs are highly variable in length. Initial appreciations described them as highly reduced, as observed in metazoans, kinetoplastids or apicomplexan (Table 1) (Feagin et al., 2012; Kirby and Koslowsky, 2017; Taanman, 1999). However, in plants and fungi they are significantly extended as compared to bacteria (Marienfeld et al., 1999; Sloan et al., 2018a; Wolters et al., 2015). Their organization is sometime unusual, for example in Chlamydomonas and Plasmodium both LSU and SSU rRNAs are fragmented in the mt-genome (Feagin et al., 2012; Salinas-Giegé et al., 2017). The 5S rRNA is not always conserved. In animals and yeast is it absent from the mt-genome, and was suspected to be imported from the cytosol for long. But it turned out not to be the case. Indeed, in mammals, a tRNA structurally replaces the 5S rRNA. Interestingly in Human, tRNA^Val compensates for the 5S, whereas in *Sus scrofa* mitoribosomes the identity of the mt-tRNA is different as a tRNA^Phe is found (Brown et al., 2014a; Greber et al., 2013; Ott et al., 2016). The incorporation of mt-tRNA into the mitoribosome in both organisms appears to result from its location in the mitochondrial genome, where both mitochondrial tRNAs flank the SSU rRNA with which they are cotranscribed (Chrzanowska-Lightowlers et al., 2017). In yeast, the 5S is

also absent, but this loss is compensated by an extension of the 21S rRNA (Ott et al., 2016). Most likely in plasmodium and trypanosome the 5S is also absent, but in plants and Reclinomonas, the 5S rRNA is still encoded in the mt-genome, suggesting that this components are part of the mitoribosome (Lang et al., 1997; Marienfeld et al., 1999).

As a result, the RNA:proteins ratio is different from the bacterial and cytosolic one. Indeed, for the latter the ratio, is of 2:1, where rRNA contribute much more to the composition compared to proteins. In mammals mitoribosomes this ratio is completely switched to a 1:2, resulting in a less dense ribosome (55S) compared to the bacterial one (70S) (Amunts et al., 2015; Greber et al., 2015). In yeast the ratio is of 1:1 and additional r-proteins associated with the extension of both the large and small rRNA components have been acquired (Bieri et al., 2018; Desai et al., 2017). The most dramatic switch occurred in *Trypanosoma brucei* where the ratio is of a striking 1:6 (Ramrath et al., 2018).

Mitoribosomes structures

Over the last two decades, considerable efforts were deployed to purify and solve the structure of mitoribosomes from different organisms. Several medium resolution structures have become available for diverse organisms: namely the bovine mitoribosome, the mitoribosomes from the fungus *Neurospora crassa* and the trypanosomatid protozoan *Leishmania tarentolae* (Sharma et al., 2003, 2009; van der Sluis et al., 2015). Four high resolution 3D structures of complete mitochondrial ribosomes have only been determined recently by cryo-EM, in yeast and two mammalian species, and this year for trypanosoma (Amunts et al., 2015; Desai et al., 2017; Greber et al., 2015; Ramrath et al., 2018). Obtaining a high resolution structure of mitoribosomes is challenging for multiple reasons. First mitoribosomes are found in low abundance in the cell compared to cytosolic ribosomes (or chlororibosomes in organisms that possess chloroplasts). Therefore, it is necessary to purify mitochondria, which can already be a challenging process. Even from pure mitochondria, several component of similar sedimentation coefficient co-purify with mitoribosomes, in particular, the abundant respiratory chain components, but also cytosolic ribosomes which are found attached to the outer membrane of mitochondria (Gold et al., 2017). Moreover mitoribosomes associate with membranes which can result in low solubility and aggregation when taken out of this hydrophobic environment. For these reasons, classical methods of structure determination, namely X-ray crystallography, which are already challenging for big molecular complexes like ribosomes, are nearly impossible with mitoribosomes. It is mostly
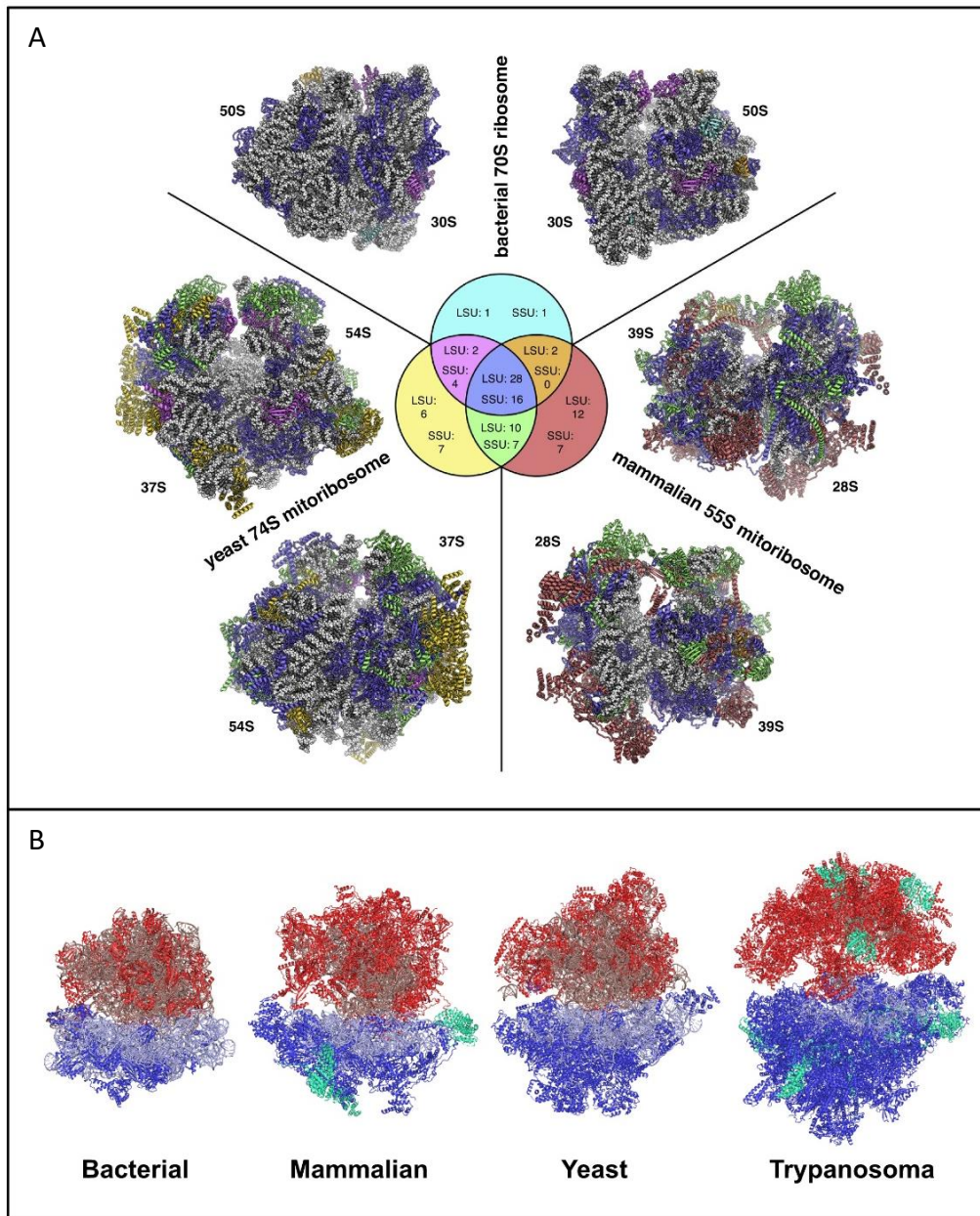
Figure 17: Mitoribosomes structures
A Composition and structural comparison of the 74S yeast and 55S mammalian mitoribosome with the bacterial ribosome. The atomic models of the ribosomes are shown in two views related by a 180° rotation. The rRNAs are represented in gray. The ribosomal proteins are colored according to their species distribution. A diagram in the center indicates the number of proteins that are either shared or specific for each type of ribosomes (color coded as in the structures). From (Bieri et al., 2018)
B The different mitoribosomes structures determined to date, compared with the *E.coli* ribosom. LSU components are showed in red shades, SSU components in blue shades. PPR proteins of the mammalian and trypanosoma mitoribosomes are shown in cyan.
E.coli 4YBB, *S.cerevisiae* 5MRC, *S.scrofa* 5AJ4, *T.brucei* 6HIV

thanks to the technological progress achieved with cryo-electron microscopy that such high resolution structures were obtained (Kühlbrandt, 2014).

In 2015, the team of **Nenad Ban** and the **Ramakrishnan** group both published structures of the 55S mammalian mitoribosomes, respectively a 3.8 Å structure of the porcine mitoribosome and a 3.6 Å structure of the human one (Fig 17) (Amunts et al., 2015; Greber et al., 2015). In 2017, Desai and colleagues published the cryo-EM reconstruction of the complete yeast 74S mitoribosome at 3.3 Å resolution (Desai et al., 2017). These structures revealed the major features of the mitoribosomes and give clues to the functions of their additional components. The 55S mammalian mitoribosome is composed of a total of 82 proteins along with two rRNAs and a tRNA replacing the 5S rRNA, termed CP tRNA. The LSU is termed 39S and is constituted by the 16S rRNA and the CP tRNA (Central protuberance tRNA) along with 52 r-proteins, the SSU is termed 28S and is composed of 30 r-proteins and the 12S rRNA (Amunts et al., 2015; Greber and Ban, 2016; Greber et al., 2015). The 74S yeast mitoribosome is composed of a total of 80 proteins along with two rRNAs, the 5S rRNA is completely absent. The LSU is termed 54S and is constituted by the 21S rRNA, which is larger than its bacterial counterpart, along with 46 r-proteins, the SSU is termed 37S and is composed of 34 r-proteins and the 15S rRNA, which is also larger than the bacterial one (Bieri et al., 2018; Desai et al., 2017). Earlier this year, the teams of Nenad Ban and André Schneider published a 7.8 Å resolution structure of the *Trypanosoma brucei* mitochondrial ribosome, which the is most protein-rich ribosome described to date, with 127 different proteins, 70 in the LSU and 57 in the SSU, along with the highly reduced 9S and 12S rRNAs (Ramrath et al., 2018). The structures revealed that among the additional r-proteins specific to mitochondria, certain fulfill an architectural function by stabilizing and protecting the rRNA core, while others provide additional functionality. More recent work, using proteomic and structural approaches, are now trying to elucidate the mechanisms of mitoribosome assembly (Bogenhagen et al., 2018; Brown et al., 2017; Zeng et al., 2018).

Characteristic structural features of the mitoribosomes

Yeast and mammalian mitoribosomes share common structural features. For example, the central protuberance (CP) of the LSU, which mediates inter-subunit contacts with the SSU head, is heavily remodeled in both organisms due to the absence of the 5S rRNA. Indeed, in yeast the CP lacks homologs of bacterial r-proteins uL18 and bL25 which interact with the 5S in bacteria, but its volume is tripled in comparison to the bacterial CP. This is the results of the acquisition of the 82-ES and 84-ES rRNA expansion segments, which form a scaffold for the mitochondria-specific r-

proteins mL38, mL40, and mL46 (Bieri et al., 2018; Desai et al., 2017). The CP of mammalian mitoribosomes is also different. As mentioned above, it acquired a tRNA as an integral component to the CP, structurally replacing the 5S rRNA. Contrary to yeast, the CP tRNA is anchored by the bacterial homologs uL18m and bL31m and the mito-specific r-proteins mL38, mL40 shared with yeast and mL48, specific to mammals (Amunts et al., 2015; Bieri et al., 2018; Greber et al., 2015). In *T.brucei*, the CP is even further modified due to the total absence of RNA. The bacterial r-proteins uL18 and bL25 are also absent but bL31m was retained. mL38, mL40, mL46 (shared with yeast and mammals) are also present, and two Trypanosoma specific, mL73 and mL96 further structure the CP, hence being almost three times the size of the bacterial one (Ramrath et al., 2018).

Moreover, both yeast and mammals have acquired the r-protein mS29. This protein is embedded in the SSU head where it contributes to the formation of inter-subunit bridges, and function as a guanine nucleotide-binder. It was showed that the mammalian mitoribosomes is able bind the guanine nucleotides GDP and GTP, the SSU exhibiting higher affinity for GTP relative to the complete 55S mitoribosome (Denslow et al., 1991). Therefore, it is suspected that depending on the state of the bound nucleotide, mS29 might regulate subunit association (Bieri et al., 2018). Interestingly in Trypanosoma, mS29 is still present and bind GTP, however it seems to be structurally unable to play a role in subunit association, the protein and GTP being too distant from the LSU (Ramrath et al., 2018).

The functional centers of the ribosome, namely the peptidyl transferase center (PTC) of the LSU as well as the decoding center of the SSU, are still exclusively formed by rRNA. Therefore the basic mechanisms of translation are still conserved and performed by a ribozyme. The A, P and E sites, that accommodates the tRNAs in the ribosomal inter-subunit space are also conserved. However, in human they all co-evolved with the mitochondrial tRNAs to accommodate to their considerable structural variability (Greber et al., 2015; Salinas-Giegé et al., 2015). The yeast sites are comparable to the bacterial ones. In Trypanosoma the tRNA binding sites are surrounded by non-conserved rRNA residues, but occupy equivalent positions in the structure compared to bacteria, thus mediating a conserved structural role. Moreover, even if the several bits of rRNA and proteins were lost in *T.brucei* mitoribosome the PTC and decoding center are still highly conserved (Ramrath et al., 2018).

Both in yeast and *T.brucei*, the mitoribosome has acquired the mS47 protein, which forms one of the two a large body protuberance. The protein shows homology with enzymes involved in
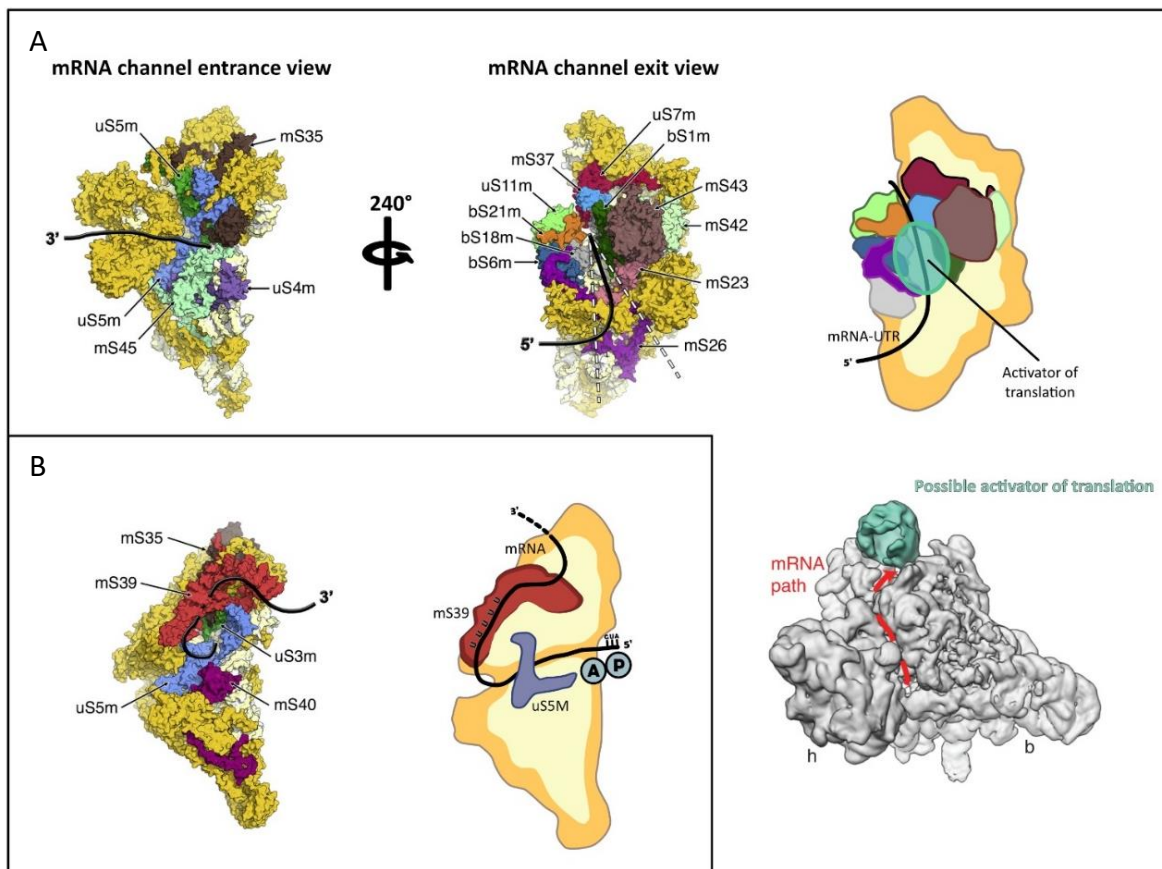
Figure 18: The mRNA recruitment to mitoribosomes

Path of the mRNA on the mitoribosomal small subunit. **A** The yeast 37S SSU is shown in two different orientations exposing the mRNA entry and exit channels. **B** the mammalian 28S SSU is only shown from one side. In each case a schematic representation is provided.

A In the case of yeast, it appears that a number of proteins at the exit channel form a V-shaped canyon. The 5' part of the mRNA pass through this canyon and is stabilized by an activator of translation (shown in cyan in the scheme). This is supported by experimental data from Desai *et al,* were additional densities at the mRNA exit channel was observed.

B In the case of mammalian, the mRNAs are leaderless, hence the stabilization of the transcript is different. It is thought that the PPR protein mS39 is involved in the recruitment and stabilization of the mRNA by recognizing U-rich sequences which are conserved starting from codon 7 in the 11 mRNAs of mammalian mitochondria. In addition, the r-protein uS5m appears to be involved in the stabilization of the mRNA.

Derived from (Bieri et al., 2018; Desai et al., 2017; Kummer et al., 2018)

valine catabolism. The structure showed that the catalytic residues are conserved and the cavity for substrate binding is accessible, therefore mS47 may still perform its enzymatic activity, but its clear function is still not understood (Desai et al., 2017; Ramrath et al., 2018). The fact that it is not present in mammals – *it is also present in plant* – suggests a secondary loss of this protein in animals.

mRNAs recruitment to the mitoribosomes

For a long time, the recruitment of mRNAs to the mitoribosome was enigmatic, but structural insights gave hint to understand it. Indeed, in bacteria the recruitment of mRNA to the SSU and the correct placement of the start codon within the mRNA channel is mediated by the interaction of the Shine-Dalgarno (SD) sequence on the mRNA with the anti-SD sequence of the 16S rRNA (Shine and Dalgarno, 1974). In mitochondria, the mRNAs and the SSU rRNA lost their SD and anti-SD sequences, therefore new structural features needed to be acquired to promote the mechanism of translation initiation (Kuzmenko et al., 2014). In animals, this mechanism was even less understood as the mRNAs do not possess 5' UTR, that usually harbors motifs to recruit the ribosome or factors involved in initiation. The data on mitoribosomes high-resolution structures provided the basis to understand mRNA recruitment in both mammals and yeast.

In mammalian mitoribosome, the r-protein mS39, which is a PPR protein, is located at the SSU head near the mRNA entry channel. It was already speculated that mS39 would be involved in the binding of the leaderless mRNAs to aid their threading through the remodeled mRNA channel entrance toward the decoding center (Amunts et al., 2015; Greber et al., 2015). The recent structure of the translation initiation complex (reconstituted *in vitro*) confirmed these results (Fig 18). Indeed the PPR motifs of mS39 apparently to not bind the mRNAs in a clear sequence-specific manner, as it is usually the case. Rather, it seems that mS39 recognize U-rich sequences which are conserved starting from codon 7 in the 11 mRNAs of mammalian mitochondria. These stretches of Us may be the determinant for the initial binding of the mitochondrial mRNAs mitoribosome (Bieri et al., 2018; Kummer et al., 2018).

In yeast mitoribosome, the situation is different as the mRNAs have long 5' UTRs. It is well established in yeast that these 5' UTRs bind to translation activators. These translation activators are usually specific to a particular transcript and are essential for the translation of the respective transcripts (Derbikova et al., 2018). The mRNA exit channel of the yeast mitoribosome is highly remodeled, composed of mitoribosomal proteins specific to yeast. These proteins form a V-shaped canyon at channel exit hence being larger. The canyon is composed by the protuberance

made of mS42 and mS43 on one side and a series of protein extensions on the other side. In the 2017 3D reconstruction, an additional density was detected on the V-shaped canyon, suggesting that this canyon acts as a binding platform for translation activators (Fig 18 A) (Desai et al., 2017).

Mitochondrial translation:

Initiation

In eubacteria, the process of translation involves three initiation factors, IF1, IF2, and IF3 (Gualerzi and Pon, 1990). IF1 and IF2 are referred to as "universal translation initiation factors" as they have conserved functional and structural homologs in eukaryotes as well as in Archaea (Roll-Mecak et al., 2001). In mitochondria, even if the general process of translation initiation seems to be conserved, where a formylated fMet-tRNA^{Met} participates in the initiation of translation similar to bacteria, the most dramatic difference with bacterial translation lies in the nature of translational factors orchestrating the process. Indeed among the three initiation factors, mitochondrial IF2 (mIF2) is universally present, mIF3 is near-universal (in yeast Aim23p is a mIF3 orthologue), and mIF1 is universally lacking (Atkinson et al., 2012). Therefore, in mitochondria one of the three crucial translation initiation factors was lost. It was shown that, the bovine mIF2 is able complement for the essential biological roles of both IF1 and IF2 in bacteria (Gaur et al., 2008). Hence, the function of IF1 seem to have been transferred to the mIF2 thanks to a conserved 37 amino acid insertion in mIF2 (Gaur et al., 2008; Kuzmenko et al., 2014).

Yeast translational activators

The regulation of mitochondrial translation is best studied in yeast, because yeast is one of the two organisms whose mitochondria can be genetically engineered (Bonnefoy and Fox, 2007). This is crucial for the study of mitochondrial protein biosynthesis, since there is no *in vitro* system for mitochondrial translation. In yeast the extended mRNAs' 5'UTRs are closely involved in mitochondrial translation regulation. Indeed, the regulation of mitochondrial translation is orchestrated by a set of proteins termed as translational activators (Herrmann et al., 2013; Kuzmenko et al., 2014). Each of these proteins specifically regulates translation of just one or two mRNAs, which was demonstrated by genetic approaches (Haffter et al., 1990; Körte et al., 1989). It appears that this unique system has no full analogs in any other domain of life. These translational activators are thought to serve two major functions. First, as some of them are both able to bind to mitochondrial ribosomes and to the inner mitochondrial membrane, they could physically anchor the mitoribosome to the IMM; and second, they could provide a feedback mechanism to sense the availability of interaction partners of the membrane proteins encoded on

the specific mRNAs (Herrmann et al., 2013). These translational activators are generally present in limiting amounts and thus restrain the expression of their target RNA, acting as translational bottlenecks. In yeast, 15 translational activators are described for 7 protein coding genes present in mitochondria. Interestingly, among these translational activators, several are PPR proteins. Indeed Pet111, Atp22 and Aep1 were described to have a clear function in the translation of their targets (Herbert et al., 2013a). Pet111 is absolutely required for the translation of *COX2* (Poutre and Fox, 1987), Aep1 is involved in *ATP9* translation (Ziaja et al., 1993), and Atp22 function in the translation of the bi-cistronic mRNA, *ATP8/6*, by preventing the synthesis of Atp6 it activates the translation of Atp8 (Zeng et al., 2006). Moreover Pet309 and Cbp1 are also two PPR proteins that affect translation, but whether or not the effect is due to their primary role is not clear (Herbert et al., 2013a; Manthey and McEwen, 1995; Staples and Dieckmann, 1994).

Regulators of translation in other eukaryotes

In other eukaryotes, translational regulator are much less characterized, and when one protein affect translation, it is usually an indirect process (e.g mRNA destabilization or transcription). In plant, little is known about mitochondrial translation and the actors of its regulation. Nevertheless it was shown that the PPR protein MTL1 (Mitochondrial Translation Factor 1) promotes the translation of the nad7 mRNA, and is associated with polysomes (Haïli et al., 2016). Two other PPR proteins, PPR336 and PNM1, were found in mitoribosomes containing fractions, but their role in mitochondrial translation and their direct association with mitoribosomes was not established (Hammani et al., 2011; Uyttewaal et al., 2008).

In mammalian, the complex LRPPRC-SLIRP is involved in post-transcriptional regulation of mitochondrial gene expression, and therefore translation. LRPPRC is one of the longest known PPR protein with about 30 PPR motifs and does not contain any other functional domains. Mutation of this protein is responsible for the French-Canadian variant of Leigh syndrome (Lightowlers and Chrzanowska-Lightowlers, 2013; Siira et al., 2017). *In vivo,* it stably interacts with its partner SLIRP which protect LRPPRC from degradation (Lagouge et al., 2015). The LRPPRC-SLIRP complex has been proposed to be a regulator of both mitochondrial protein synthesis and polyadenylation of mitochondrial mRNAs that mediates RNA stability, although the exact target of the complex was not known (Ruzzenente et al., 2012). Recently it was shown through a PAR-CLIP approach that the LRPPRC–SLIRP complex is a global RNA chaperone that decorates mRNAs, resulting in the disruption of local mitochondrial RNA secondary structures. Through this process, the required

sites on the transcripts are exposed for translation, stabilization, and polyadenylation (Siira et al., 2017).

Elongation and termination

Compared to initiation and termination, elongation seem to be the most conserved step of translation in mitochondria. The high resolution structures confirmed that the tRNA binding sites are rather conserved (in structure) even in distant organisms such as Human, yeast and Trypanosoma, even if in Human (and animals in general) the mitoribosomes need to accommodate tRNAs with divergent structures (Bieri et al., 2018; Ott et al., 2016). And as mentioned above, the peptidyl transferase center of the LSU as well as the decoding center of the SSU, are still exclusively formed by rRNA, and structurally conserved.

Termination on the other hand, has diverged significantly, one of the reason being the genetic code drift, which lead to the apparition of new stop codons. Moreover additional mitochondria specific termination factors are recruited compared to bacteria. Indeed, in bacteria, two separate release factors (RF1 and RF2) recognize stop codons (RF1 for UAA and UAG, and RF2 for UAA and UGA) in the ribosomal A site to promote release of the nascent polypeptide and recycling of the ribosome (Scolnick et al., 1968). Homologs of both these release factors are found in eukaryotes, and are termed as "canonical release factors". mtRF1a is the most widespread of all organellar release factors. Every eukaryotic organism with a mitochondrial genome, harbors a mitochondrial type RF1 encoded in the nucleus (Chrzanowska-Lightowlers et al., 2011). For the mitochondrial mtRF2a, it is different as it has a relatively narrow phylogenetic distribution, when compared to its mtRF1a counterpart. It has been lost at least five times during the eukaryotic evolution, coevolving together with the mitochondrial genetic code. It is only consistently found in streptophytes (land plants), red algae, dictyosteliida, and some stramenopiles (Duarte et al., 2012). Aside from canonical release factors, non-canonical release factors are also found. mtRF1 is probably the most studied non-canonical release factor, and yet its molecular function fails to be determined. It is the longest protein of the RF family, and is a vertebrate-specific mitochondrial protein (Duarte et al., 2012; Ott et al., 2016). It has been suggested that mtRF1 could be responsible for decoding the nonstandard mitochondrial stop codons, AGG and AGA, predicted to terminate numerous vertebrate mitochondrial ORFs. This is mainly supported by the fact that this protein's origin and the AGG/ AGA stop codons' origin both root in the vertebrate lineage. Nevertheless, this hypothesis has never been experimentally confirmed. Moreover, it was shown that at least in Human, on entry of the single AGA or AGG triplet to the mitoribosomal A-site, a -1

frame shifting event occurs, placing a standard UAG into the A-site, hence bypassing the need for an extra RF protein (such as mtRF1) (Temperley et al., 2010).

An interesting case is the one of ICT1, which is also classed in the non-canonical release factors. It was shown that this factor has, *in vitro*, a ribosome-dependent peptidyl-tRNA hydrolase activity. Moreover it was shown to be an integral subunit of the mitoribosome, sitting on the LSU (Richter et al., 2010). It was therefore speculated that ICT1 could act when abortive elongation occurs, to permit the release of the truncated peptide from the mitochondrial monosome (Richter et al., 2010). But the high-resolution structure of the human mitochondrial ribosome showed that ICT1 is indeed stably incorporated into the mitoribosome but in a position incompatible with an access to the ribosomal A site (Amunts et al., 2015; Greber et al., 2015). Thus, for a putative function as a peptidyl-tRNA hydrolase, ICT1 would either need to be released from the mitoribosome or to be present in an additional free pool (Greber and Ban, 2016; Ott et al., 2016). If it is the case, experiments suggested that ICT1 may act in mitochondrial translation termination when the nonstandard AGA and AGG termination codons would enter the A-site, which would allow an alternative to the −1 ribosomal frameshifting (Akabane et al., 2014).

Aims of this study

As described during this introduction, mitochondria is a crucial component of eukaryotes, where mitochondrial translation plays a central role. In plants, while numerous pentatricopeptide repeat (PPR) proteins are involved in all steps of gene expression, their function in mitochondrial translation remained unclear. Recent works has revealed that mitochondrial protein synthesis structurally and functionally diverged between eukaryotes, with many aspects carrying organelle-specific features. In plants, the translation mechanism and the composition of the mitoribosome remained particularly elusive. Hence, to fully understand mitochondria biology in plants it was necessary to determine the composition of their specialized ribosomes. Therefore my PhD project focused on the characterization of the Arabidopsis mitochondrial translation apparatus. My work was divided in two parts:

The first part consisted in the biochemical and biophysical characterization of the mitochondrial ribosome of *A.thaliana.* To do so, I developed an adequate protocol for purification of mitochondrial ribosomes by complementary approaches, either through classical biochemical purification or by co-immunoprecipitation using a plant specific mitoribosomal r-protein, rPPR1-HA/336-HA/mS49-HA. The aim was to obtain sufficiently pure samples to analyze them by mass spectrometry to accurately determine the composition of Arabidopsis mitoribosome, but also by cryo-electron microscopy in order to determine the structure of a plant mitoribosome.

The second part focused on the functional characterization of the specific protein factors associated with the mitochondrial ribosome of *A.thaliana*. In this case I studied the role of a specific set of PPR proteins, termed here rPPR. Even if previous studies indicated that some of these proteins were involved in mitochondrial translation, these proteins were never directly identified as true ribosomal proteins. Along with these PPR proteins, other plant specific mitochondrial r-proteins were discovered. To characterize these factors, I have analyzed several mutant lines to identify their functions. Finally, the function of one of these novel proteins, rPPR1, was analyzed at the molecular level by ribosome profiling.

Results

Figure 19: Mitochondrial ribosomes purification protocols used during this study
This scheme represent the overall workflows of the two main protocols used during this study.

Figure 20: Occurrence of mitoribosomes and cytoribosomes in crude ribosome fractions
A SDS-PAGE of a purification of crude ribosomes with **1.** Arabidopsis total cells **2.** Crude mitochondria **3.** Pure mitochondria **4.** Crude ribosomes.
B Western blot using antibodies directed against the cytosolic rps6 and mitochondrial rpl16 r-proteins.

Results

Optimization of Arabidopsis mitoribosomes purification

At the beginning of this study, no defined protocol was established for the preparation of highly pure plant mitoribosomes. Hence, my first goal was to develop an optimal procedure for the purification of plant mitoribosomes. The work of the Ban team on mammalian mitoribosomes was the starting point to establish an adequate purification procedure (Greber et al., 2013, 2015). Moreover, previous work in the laboratory showed that by using the Greber et al protocol up to the crude mitoribosome step (Fig 19) (before separation on sucrose gradients), the final product was sufficient to obtain mitoribosomes in reasonable quantities for preliminary cryo-EM analysis (even though not pure enough to obtain a high-resolution reconstruction).

Mitoribosomes are found in very low quantities in the cell as compared to cytosolic ribosomes and chloroplast ribosomes. Therefore, the first step of mitoribosome purification is to purify mitochondria. For mitochondria purification, it quickly appeared that a way of extracting quickly large amounts of mitochondria was mandatory. As a result, the classical purification of mitochondria from Arabidopsis flowers was only used for the analysis of mutants (see below). For the optimization of mitoribosome purification and for structural analysis, which requires large quantities of starting material, we chose to work with Arabidopsis dark grown cell cultures. In terms of yields, when starting from flower material, it first takes 2 months to grow plants to the flowering stage, then 30 g of flowers can be harvested from about 100 plants to obtain 2-5 mg of pure mitochondria. With the cell suspension, a 1 L culture grown for 1 week, results in the equivalent of 100 g of dry cells, and yields around 10-15 mg of pure mitochondria. Furthermore, chloroplastic contaminations are almost inexistent with dark grown cells.

Using the Greber et al protocol, I first tested by western blot, using an antibody for the cytosolic rps6 and one for the mitochondrial rpl16 r-proteins, that my crude ribosome pellet was indeed enriched in ribosomes (Fig 20). This confirmed that both mitochondrial and cytosolic ribosomes were enriched in the final crude ribosomes pellet. Even though the samples did contain mitoribosomes, mass spectrometry analyses revealed that the samples were highly contaminated by 1) cytosolic ribosomes, most probably attached at the surface of mitochondria and 2) other mitochondrial components (mainly respiration chain proteins). The cytosolic ribosomes were not problematic for proteomic analyses as they could easily be sorted out, but in terms of structural analyses it appeared that the data sets contained a majority of individualized cytosolic ribosomes
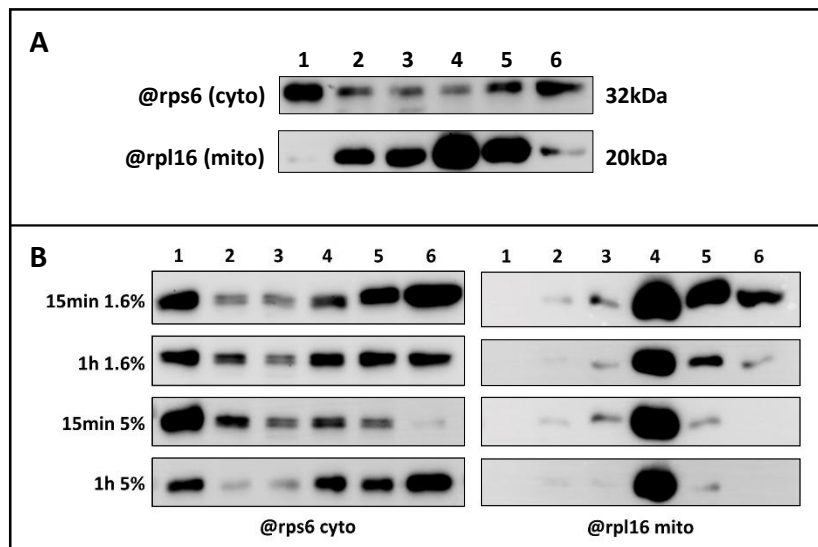
Figure 21: Mitochondria lysis optimization
A Ribosome purification monitored by western blot analysis using antibodies against the cytosolic rps6 and mitochondrial rpl16 r-proteins. **1.** Arabidopsis total cells **2.** Crude mitochondria **3.** Pure mitochondria **4.** First clarification pellet after lysis **5.** Second clarification pellet **6.** Crude ribosomes. As observed here, the majority of mitochondrial ribosomes are lost during the first clarification step (4)
B Mitochondria lysis optimization. Similar to **A**, the same purifications were performed and monitored, but different mitochondrial lysis conditions were used.

(more than 70% of particles representing individualized ribosomes (i.e. monosomes) corresponded to cytosolic ribosomes) hence hampering the 3D reconstruction and only yielding low resolution maps of mitoribosomes. In the case of mitochondrial components such as respiratory complexes, they were easily sorted out during EM analysis, but as the composition of the plant mitoribosome was completely unknown, any of those proteins could have been in theory a novel uncharacterized plant-specific mitochondrial r-protein. Further purification steps were thus required to determine *bona fide* core mitoribosome proteins.

Moreover the whole ribosome purification process was monitored. During the mitoribosome purification, the final yield of mitoribosomes was not satisfying as seen by western analyses (Fig 21 A). Indeed, most of the mitoribosomes were lost in the first clarification pellet, which follows mitochondria lysis. This is most likely due to the fact that, similar to yeast and mammalian ones, plant mitoribosomes are bound to the inner membrane of mitochondria. Beside immuno-detections, qPCR on rRNA mt / rRNA chloro / rRNA cyto was also used to try to assess the purity of the samples. However, results were not informative enough and this method was therefore abandoned. It also appeared clearly that immuno-detections on a limited number of proteins was not sufficient to assess the purity of the final samples. Therefore, the equivalent of 20 µg of proteins were almost systematically analyzed by LC-MS/MS to determine the extensive protein content of the respective purification fractions.

As the initial mitochondria solubilization step seemed to be crucial for mitoribosome recovery, different solubilization conditions were tested. I first tested different times of solubilization (15 min or 1h) and different concentrations of detergent (1.6% or 5% Triton X-100) (Fig 21 B). It appeared that 5% of Triton was clearly not efficient. The difference between 15 min and 1 h of solubilization using 1.6% of detergent was not significative. Therefore to avoid degradation and to keep the protocol short, the "1.6%, 15 min" condition was kept. **Triton X-100** was the detergent used in the standard condition, but I also tested two other detergents, **nDM** and **digitonin**. In the end, both mass spectrometry and electron microscopy showed that digitonin was not efficient at all to solubilize mitochondrial ribosomes. Indeed, by MS, 2803 spectra were obtained for 270 individual proteins with **digitonin**, whereas 7080 spectra for 347 different proteins were obtained with **Triton X-100** and 8714 spectra for 390 proteins with **nDM**. Given the MS results, nDM appeared to be similar or better than Triton X-100, but EM screening showed that Triton resulted in better quality particles than nDM. Triton was thus retained as a standard detergent for all subsequent analyses.

A

Accession	Description	100mM	200mM	400mM
AT2G37230	Tetratricopeptid	54	93	67
AT1G60770	Tetratricopeptid	41	81	53
AT5G60960	PNM1 Pentatric	34	50	39
AT4G36680	Tetratricopeptid	32	53	32
AT1G19520	NFD5 pentatrico	34	70	27
AT5G15980	Pentatricopeptic	34	47	
AT3G02650	Tetratricopeptid	23	47	21
AT1G61870	336 pentatricop	18	34	5
AT1G55890	Tetratricopeptid	22	27	
AT3G13160	Tetratricopeptid	22	13	
AT3G49240	Pentatricopeptic	11	5	
AT1G80270	Pentatricopeptic	6	10	
AT1G26460	Tetratricopeptid	7	6	
AT4G35850	Pentatricopeptic	8	7	
AT2G15690	Pentatricopeptic	6	4	
AT1G15480	Tetratricopeptid	4	3	
AT1G11630	336L Tetratricop	3	6	2
AT3G61520	Pentatricopeptic	3	2	
AT3G54980	Pentatricopeptic	2	2	
AT5G28460	Pentatricopeptic	2	2	
AT3G15590	Tetratricopeptid	3	4	2
AT3G02490	Pentatricopeptic	4	2	
AT5G18950	Tetratricopeptid	2	3	
AT4G28080	Tetratricopeptid	2	3	
AT1G01320	Tetratricopeptid	2	2	
AT3G14110	Tetratricopeptid	2	1	

B

	Accession	Description	100mM	200mM	400mM
uL1m	AT2G42710	Ribosomal prote	34	69	42
bL25m	AT5G66860	Ribosomal prote	22	34	42
uL3m	AT3G17465	ribosomal protei	23	38	39
uL4m	AT2G20060	Ribosomal prote	24	38	30
bL9m	AT5G53070	Ribosomal prote	15	29	27
mS29	AT1G16870	mitochondrial 28	19	48	26
uS5m	AT1G64880	Ribosomal prote	31	58	22
uL29m	AT1G07830	ribosomal protei	21	28	22
bL25m	AT4G23620	Ribosomal prote	17	25	19
uL13m	AT3G01790	Ribosomal prote	11	25	19
uL2m	AT2G44065	Ribosomal prote	14	27	18
uL30m	AT5G55140	ribosomal protei	13	20	18
mL41	AT5G39800	Mitochondrial rib	13	19	18
mL43	AT3G59650	mitochondrial rib	7	10	14
uS11m	AT1G31817	Ribosomal L18p	12	34	13
uL22m	AT4G28360	Ribosomal prote	12	19	13
bL21m	AT4G30930	Ribosomal prote	18	25	12
bL17m	AT5G09770	Ribosomal prote	2	11	12
uL22m	AT1G52370	Ribosomal prote	13	21	11
mS35	AT3G18240	Ribosomal prote	10	20	11
uS2m	AT3G03600	ribosomal protei	11	23	9
uS10m	AT3G22300	ribosomal protei	8	18	8
uL14m	AT5G46160	Ribosomal prote	8	16	8
uS19m	AT5G47320	ribosomal protei	6	14	7
bS18m	AT1G07210	Ribosomal prote	4	12	7
uS3m	ATMG00090	structural consti	10	34	6
bS16m	AT5G56940	Ribosomal prote	18	29	6
mS35	AT4G21460	Ribosomal prote	6	12	6
uS13m	AT1G77750	Ribosomal prote	5	13	2
bS21m	AT3G26360	Ribosomal prote	5	11	2

Table 4: Purification optimization using different salt concentrations

To optimize mitochondria lysis and mitoribosome purification, different salt concentrations were tested. The final crude ribosome fractions were analyzed by mass spectrometry and the results are presented in **A** and **B**.

In **A** the list of PPR proteins identified in the three different conditions is presented. In **B** the most abundant mitoribosomal proteins are listed. Some PPR proteins, highlighted in red, were resistant to the highest salt treatment and found in similar abundance compared to mitochondrial r-proteins.

For the purification process I also tested different buffer conditions, notably by testing different salt concentrations. Three different conditions were tested using 100 mM of KCl (the standard one), 200 mM or 400 mM which were used all along the purification process. The final samples were analyzed by mass spectrometry and showed that even though the 100 mM and 200 mM gave comparable results, 200 mM seemed nonetheless to be the best of the three conditions. With the 400 mM salt condition very few proteins were retrieved but the remaining ones were in majority ribosomal proteins (Table 4). Interestingly by performing this and comparing the proteins in the three different samples, I was able to see that a certain group of PPR proteins were retained all along the purification even in the 400mM KCl condition. These PPR proteins that were resisting the salt treatment were also found in similar amount compared to the "true" mitochondrial r-proteins. This was the first hint that PPR proteins were in fact part of the plant core mitoribosomes, and the proteins listed in red in Table 4, constituted my first list of candidate plant-specific r-proteins.

To further purify my samples, separation by size exclusion chromatography columns was also tested. However, the final samples were too diluted to be analyzed rapidly and conveniently as required for further structural analyses. This purification step was thus abandoned.

Alternative protocol to purify plant mitoribosomes

After trying all the different conditions of purification described above, it appeared clearly that highly pure mitoribosomes could not be obtained with this strategy and that an alternative protocol had to be used in order to obtain much purer samples. Hence, I decided to change the protocol and use the procedure used in the Ramakrishnan lab for the purification of the human and yeast mitoribosomes (Amunts et al., 2014, 2015; Desai et al., 2017). This protocol did not involve PEG precipitation which I found particularly tricky and non-consistent (Fig 19). Moreover during the purification optimization, I only started from 1-3 mg of mitochondria to purify mitoribosomes, but it quickly appeared that it was not sufficient for cryo-EM and proteomic studies. Hence, to obtain much purer samples, I chose to start from a lot more material, about 20-25 mg of mitochondria, and to have an additional sucrose gradient step of purification. Using this method I was able to separate cytoribosomes from mitoribosomes with reduced contamination by respiratory complexes proteins. These results are presented in the manuscript "Small is big in Arabidopsis mitochondrial ribosome" by Waltz et al.
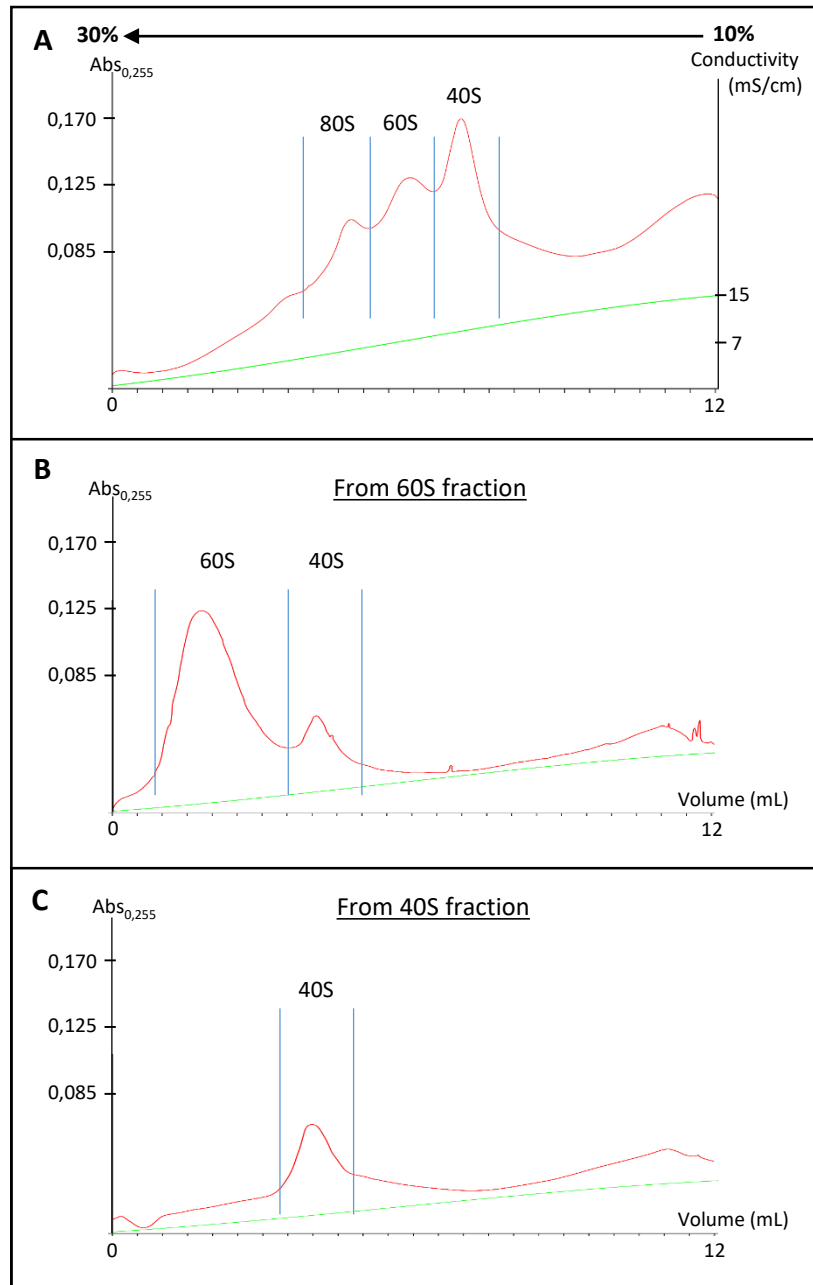
Figure 22: Attempts to resolve mitoribosome subunits using cauliflower mitochondria
A Crude ribosomes purified from cauliflower were separated on a continuous 10-30% sucrose gradient (16 h, 20 krpm). Fractions corresponding to the 60S peak, composed of the full mitoribosome, and the 40S peak, composed of the dissociated mitoribosome, were then further separated under high salt conditions (500 mM KCl) and in more resolutive conditions (20 h, 25 krpm), to promote subunit separation. The resulting separation are presented in **B** and **C**.
B The separation resulting from the 60S fraction. Two peaks are observed at 40S and 60S respectively
C The separation resulting from the 40S fraction. Only one peak is observed. No further separation of the two mitoribosome subunits could be achieved.
Altogether, results suggest that the separated plant mitoribosome SSU and LSU both sediment at 40S.

Separation of the small and large mitoribosome subunits

One of the main problem that we had with the Arabidopsis mitoribosome was that we could not clearly separate the large and the small mitochondrial subunits on sucrose gradients, both migrating in a 40S peak. As a result, it was particularly difficult to assign the newly identified proteins to each subunit. Therefore, I wanted to try the separation of the two subunits on more resolutive gradients. To do so, after retrieving the peak corresponding to the "full mitoribososomes" and the one corresponding to the "dissociated mitoribosomes" I added another step of purification. Each fractions were separated on more resolutive gradient under dissociative conditions (higher salt and lower magnesium). As this procedure required an additional step of purification, I chose to optimize this procedure using cauliflower as a starting material which yield a lot more mitochondria than Arabidopsis cells (60 mg/kg of cauliflower).

The analysis of sucrose gradients showed that besides the "full mitoribosomes" fraction (60S), only one additional peak (40S) was detected. This peak corresponds to the dissociated form of the mitoribosomes, containing both the large and small subunit (Fig 22). Then, when attempting to separate the fraction containing the already-dissociated mitoribosomes, only a single peak could be observed as well. Therefore I was not able to further separate the two subunits. However, this confirmed that both mitoribosome subunits are of similar size and molecular weight, corresponding to about 40S, similar to the small subunit of the cytoribosome (40S) that co-migrates with the two mitochondrial subunits. With the purification samples obtained with cauliflower mitochondria, I was not able to produce good quality proteomic results, mainly because of the poor annotation of the cauliflower genome and proteomic databases. However, when Arabidopsis was used, as described in the Waltz et al manuscript, the fine analysis of the peak containing the mitoribosome dissociated subunits, allowed to show that even though both subunits co-migrate in the same peak, the small subunit remains slightly lighter than the large subunit and a clear gradient of SSU and LSU proteins can be observed from the start to the end of the 40S peak. These results also allowed me to assign each r-proteins to their respective subunit.

Figure 23: The *336/336L* double mutant and construction of the 336-HA (rPPR1-HA) line
A Macroscopic phenotype of the *336/336L* mutant and the partial phenotype complementation observed in the *336/336L* – 336-HA line.
B T-DNA insertion verified by PCR. **WT PCR** was performed with primers to detect the WT 336 locus, **mt PCR** was performed with primers to detect the mutant 336 locus and **336HA PCR** was performed using primers specific to the HA sequence. **1** and **2** correspond to the 336-HA lines 1 and 2. **A** correspond to WT gDNA, **B** to *336* gDNA and **C** to *336L* gDNA.
C 336-HA expression verified on total plant extract by western blot analysis using HA specific antibodies. A total plant extract expressing YacP-HA, a chloroplastic protein tagged with an HA epitope, was used as western positive control.

Figure 24: BN Page analyses
Mitochondrial complexes of the *336/336L* mutant, as well as the ones of the *336/336L* – 336-HA line were analyzed by BN-PAGE and compared to WT.
The *336/336L* mutant has reduced accumulation and activity of Complex I. The *336/336L* – 336-HA line also has reduced accumulation and activity of Complex I but not to the level of *336/336L.* Altogether, this suggests that both PPR336 (rPPR1) and PPR336L are required for adequate complex I accumulation but that their functions are not redundant.

Immuno-purification of mitoribosomes using rPPR1 (PPR336) as a bait

At the beginning of this project, two PPR proteins had already been proposed to be non-canonical r-proteins, associated to the plant mitoribosome, PPR336 and PNM1, also called rPPR1 and rPPR9 hereafter (Hammani et al., 2011; Uyttewaal et al., 2008). PNM1 was already well studied and the fact that its knock-out mutation was lethal (which made its function more difficult to investigate at the molecular level) resulted in the choice of PPR336 as a tool to try to purify mitoribosomes.

Characterization of the *336/336L* line

Initial analysis of the *ppr336* knock out mutant did not allow to recognize any striking macroscopic phenotype (as described by Uyttewaal et al., 2008). However, Uyttewaal et al., phylogenic analyses revealed that PPR336 defines a small subfamily of P-class PPR proteins. This subfamily is composed of eight PPR proteins resembling PPR336. They are all short (423 amino acids on average) and share an unusually high percentage of sequence identity with PPR336 (average of 35%). Among them, the one that was the most similar to PPR336 with 70% identity in amino acids, was called "PPR336-Like" or PPR336L. Similar to *ppr336,* the *ppr336L* mutant had no clear macroscopic phenotype. Given the similarity of the two proteins, it was suspected that PPR336 and PPR336L might have redundant functions. Therefore I studied the phenotype of the double mutant line that was already available in the lab. Contrary to single mutants, the double mutant plants did have a clear growth delay phenotype, thus suggesting that the two proteins might indeed have redundant function or might be involved in a same pathway (Fig 23 A).

Creation of 336HA plants

Since it was expected that the two proteins might be associated with the mitoribosome, the double mutant *336/336L* was transformed by floral-dip using a construct allowing the expression of PPR336 under the control of its own promoter (1kb upstream the AUG) and fused with one HA-tag in C-terminus of the protein. Two plants were selected in the progeny for hygromycin resistance. The T-DNA insertion was confirmed by PCR and the expression of the protein confirmed by anti-HA immuno-detection (Fig 23 B - C). Additionally, both lines showed a restored phenotype, from the *336/336L* phenotype to almost wild-type, indicating that the protein was correctly expressed and that it was fulfilling its original function (Fig 23 A).

A

	WT		M	336-HA L1		336-HA L2			WT		M	336-HA L1		336-HA L2	

WT — Inp El M 336-HA L1 — Inp El 336-HA L2 — Inp El

70 kDa
55 kDa
42 kDa
35 kDa
25 kDa

← 45 kDa

Coomassie staining **Western @HA**

B

	Accession	Description	Col0	336-HA	
				L1	L2
BAIT 336HA	AT1G61870	pentatricopeptide	0	233	341
mS77 (rPPR2)	AT1G19520	pentatricopeptide	0	410	413
mS81 (rPPR8)	AT5G15980	Pentatricopeptide	0	404	420
mS29	AT1G16870	mitochondrial 28S	0	277	349
mS80 (rPPR6)	AT3G02650	Tetratricopeptide	0	254	353
uS15m	AT1G15810	S15/NS1, RNA-b	0	222	220
mS85	AT1G53645	hydroxyproline-ri	0	222	236
mS78 (rPPR3a)	AT1G55890	Tetratricopeptide	0	211	242
uS9m	AT3G49080	Ribosomal protei	0	201	210
mS75	AT5G62270	unknown protein	0	186	215
uL1m	AT2G42710	Ribosomal protei	0	165	196
uS11m	AT1G31817	Ribosomal L18p/	0	159	178
mS23	AT1G26750	unknown protein	0	159	176
uS3m	ATMG00090	structural constit	0	153	159
uS15m	AT1G80620	S15/NS1, RNA-b	0	144	150
uS2m	AT3G03600	ribosomal protein	0	122	165
bS16m	AT5G56940	Ribosomal protei	0	121	173
uS4m	ATMG00290	Alpha-L RNA-bin	0	119	113
mS35	AT3G18240	Ribosomal protei	0	118	152
NA	AT4G31210	DNA topoisomera	0	107	105
NA	ATMG00560	Nucleic acid-bind	0	103	151
uS10m	AT3G22300	ribosomal protein	0	102	111
NA	AT5G08610	P-loop containing	0	94	100
bL25m	AT4G23620	Ribosomal protei	0	91	79
NA	AT3G03060	P-loop containing	0	90	93
bS18m	AT1G07210	Ribosomal protei	0	90	90
mS79 (rPPR3b)	AT3G13160	Tetratricopeptide	0	89	124
uL3m	AT3G17465	ribosomal protein	0	89	90

Figure 25: Mitoribosome co-immunoprecipitation using 336HA

A SDS-PAGE and western blot analysis of the initial co-IP performed to assess the 336-HA lines. Compared to WT, proteins are already observed by SDS-PAGE in the elution fractions (**El**) of the 336-HA lines. Western analyses confirmed the presence of 336-HA in the elution fractions for both lines. Line 2 consistently yielded higher levels of 336-HA protein. **M** represent molecular weight markers, **Inp** the inputs and **El** the elution fractions. The red arrow indicates the signal corresponding to PPR336-HA.

B The elutions presented in **A** were analyzed by mass spectrometry. The table present a list of the 20 proteins detected at the highest levels by LC-MS/MS in the two different lines and compared with wild type Col0 plants. In both lines, the highest ranking proteins identified are in majority mitochondrial r-proteins.

BN-PAGE analyses

To investigate the cause of the macroscopic phenotype of the *336/336L* mutant, a BN-PAGE analysis was performed in order to study the state of mitochondrial respiratory complexes in this mutant. Interestingly, the double mutant is highly impaired for Complex I abundance which results in a lower activity. The *336/336L* – 336HA still has a low Complex I abundance and activity, but it is higher than that of the *336/336L* mutant. Hence, the majority of this Complex I deficiency seems to be caused by the PPR336L mutation. (Fig 24)

Mitoribosome co-Immunoprecipitation

The purification of mitoribosomes through sucrose gradients was efficient, but it was heavily contaminated i) by cytoribosomes and to a lesser extent ii) by components of the respiratory chain. Therefore, in order to have a more precise idea of the composition of the Arabidopsis mitoribosome, I used the 336HA plant line as a tool to co-immunoprecipitate the mitoribosome. It was a way to confirm that PPR336 was indeed associated with the mitoribosome, as shown by the preliminary salt treatment and sucrose gradients results, but also a way to specifically purify the mitoribosome devoid of cytoribosome contamination.

First, I purified mitochondria from the two independent 336HA lines that I obtained and performed co-IP with them. The immunoprecipitation of 336HA was confirmed by western blot analyses and the final elutions were analysed by MS (Fig 25). From the data it was clear that both lines were able to co-IP the mitoribosome, with line 2 being more efficient than line 1 (with higher numbers of spectra consistently retrieved for line 2 than from line 1). This correlated with the protein expression level observed by western. Line 2 was thus selected for further analyses. To confirm that 336HA was indeed able to co-IP the entire mitoribosome, the co-IP were then performed in triplicates and with different conditions. Control IPs were performed in parallel of each experiments to allow statistical analyses. All the results are presented in the Waltz et al manuscript.

Using this method I was able to i) confirm that PPR336 is indeed associated with the mitoribosome and ii) to IP the full mitoribosome and identify its different core components. By performing the co-IP with increased salt concentrations, I was also able to show that PPR336 is associated to the small subunit. Compared to the sucrose gradient purification I had no contamination from cytoribosomes or from components of the respiratory chain. However, as I was purifying active translating mitoribosomes, I co-purified translation and RNA maturation
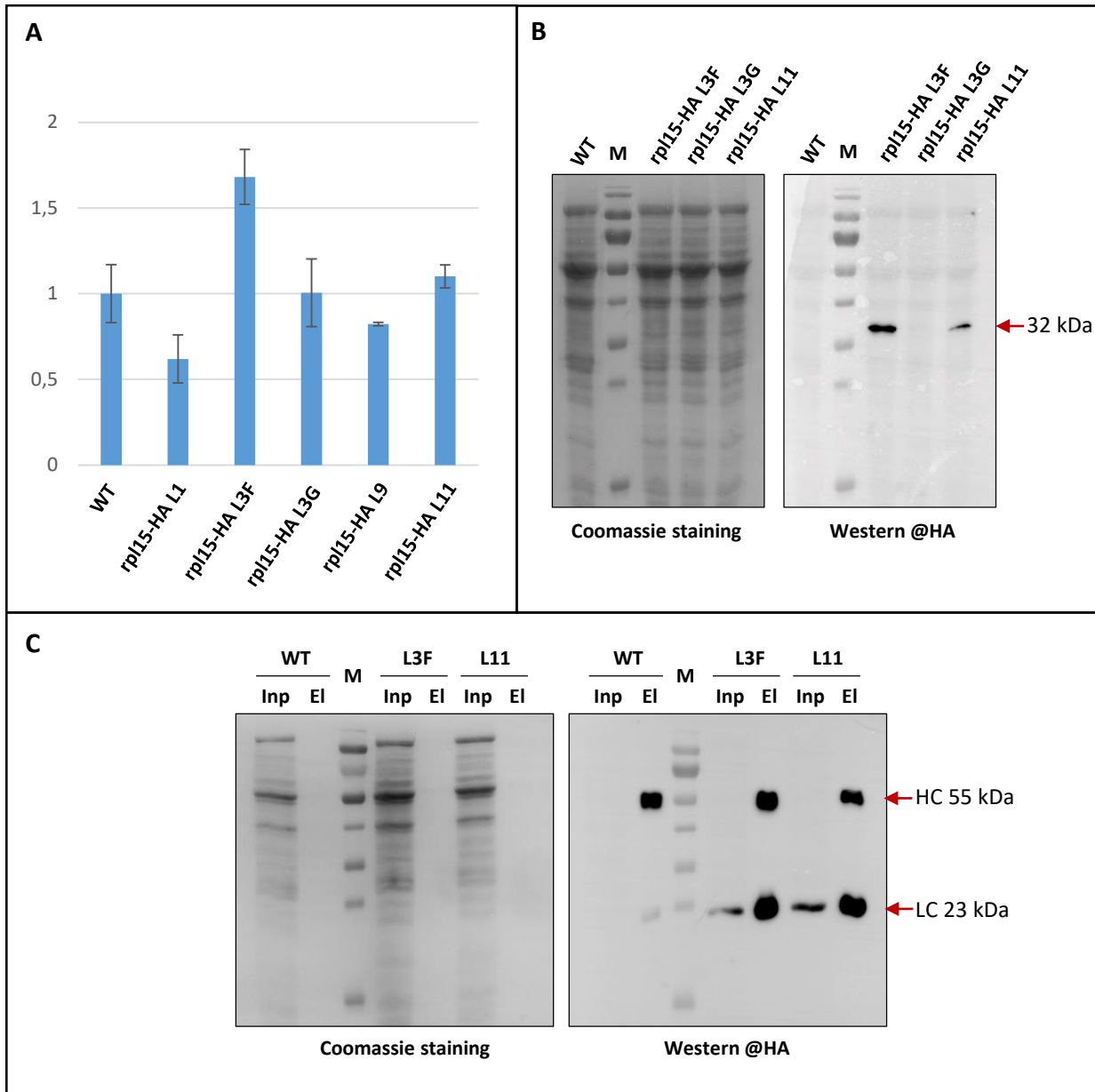
Figure 26: Mitochondrial rpl15-HA IP
A rpl15-HA transcript levels analyzed by quantitative RT PCR compared to wild-type. Five independent plant lines were tested. L3F, L3G and L11 were retained for further analysis as their transcript levels were similar or higher to that of the WT.

B rpl15-HA protein levels were investigated by western blot analyses in the three lines selected. Protein extracts from purified mitochondria were used because the rpl15-HA was undetectable in total plant extracts. rpl15-HA was clearly detected in L3F and L11 lines, therefore these lines were selected for co-IP. The red arrow indicates the signal corresponding to rpl15-HA.

C Co-IP were performed on the rpl15-HA lines selected, starting from purified mitochondria. The same experiment was performed using WT mitochondria. Inputs (**Inp**) and elutions (**El**) were analyzed by SDS-PAGE and immuno-detections. rpl15-HA was undetectable in the elution fractions. These results were also confirmed by mass spectrometry analyses.

Red arrows indicate the signals corresponding to the Light and Heavy chain of the @HA antibodies. **M** represent molecular weight markers.

related proteins, which were not present in sucrose gradient purifications, where free monosomes were separated.

As a control and to further confirm the protein composition of mitoribosomes, I tried to perform co-IPs using mitochondrial rpl16 antibody, but the result were negative, most probably because the antibody epitopes were masked in the ribosome structure. I also tried to perform co-IPs using complemented mitochondrial rpl15HA lines obtained from Hakim Mireau. From five independent plant lines, two were selected for their rpl15HA protein expression. These two lines were tested for co-IP, with the same protocol used successfully with 336HA plants. However, here as well, IPs attempts were unsuccessful, i.e. even the tagged bait protein was not found (Fig 26), most probably because the tag was not accessible and masked in the ribosome structure.

Small is big in Arabidopsis mitochondrial ribosome

Florent Waltz[1], Tan-Trung Nguyen[2], Mathilde Arrivé[1$], Anthony Bochler[4$], Johana Chicher[3$], Philippe Hammann[3$], Lauriane Kuhn[3$], Martine Quadrado[2$], Hakim Mireau[2*], Yaser Hashem[4*] & Philippe Giegé[1*]

[1]Institut de biologie de moléculaire des plantes, UPR2357 du CNRS, Université de Strasbourg, 12 rue du général Zimmer, F-67084 Strasbourg, France

[2]Institut Jean-Pierre Bourgin, INRA, AgroParisTech, CNRS, Université Paris-Saclay, RD10, 78026 Versailles Cedex, France

[3]Plateforme protéomique Strasbourg Esplanade, FRC1589 du CNRS, Université de Strasbourg, 15 rue René Descartes, Strasbourg, F-67084, France

[4]Institut Européen de Chimie et Biologie, U1212 Inserm, Université de Bordeaux, 2 rue R. Escarpit, F-33600 Pessac, France

[$]by alphabetical order

[*]authors for correspondence

Small is big in *Arabidopsis* mitochondrial ribosome

Florent Waltz [1], Tan-Trung Nguyen[2], Mathilde Arrivé[1], Anthony Bochler[3], Johana Chicher[4],
Philippe Hammann[4], Lauriane Kuhn[4], Martine Quadrado[2], Hakim Mireau [2]*, Yaser Hashem [3]* and
Philippe Giegé [1]*

Mitochondria are responsible for energy production through aerobic respiration, and represent the powerhouse of eukaryotic cells. Their metabolism and gene expression processes combine bacterial-like features and traits that evolved in eukaryotes. Among mitochondrial gene expression processes, translation remains the most elusive. In plants, while numerous pentatrico-peptide repeat (PPR) proteins are involved in all steps of gene expression, their function in mitochondrial translation remains unclear. Here we present the biochemical characterization of *Arabidopsis* mitochondrial ribosomes and identify their protein subunit composition. Complementary biochemical approaches identified 19 plant-specific mitoribosome proteins, of which ten are PPR proteins. The knockout mutations of ribosomal PPR (rPPR) genes result in distinct macroscopic phenotypes, including lethality and severe growth delay. The molecular analysis of *rppr1* mutants using ribosome profiling, as well as the analysis of mitochondrial protein levels, demonstrate rPPR1 to be a generic translation factor that is a novel function for PPR proteins. Finally, single-particle cryo-electron microscopy (cryo-EM) reveals the unique structural architecture of *Arabidopsis* mitoribosomes, characterized by a very large small ribosomal subunit, larger than the large subunit, bearing an additional RNA domain grafted onto the head. Overall, our results show that *Arabidopsis* mitoribosomes are substantially divergent from bacterial and other eukaryote mitoribosomes, in terms of both structure and protein content.

Mitochondria are essential components of the eukaryotic cell. They were acquired through the endosymbiosis of an α-proteobacterial ancestor that evolved into compartmentalized energy conversion powerhouses, and they produce most of the energy of eukaryotic cells through oxidative phosphorylation involving an electron transport chain in their inner membrane[1]. Their dysfunction has been associated with an increasingly large number of inherited disorders and is implicated in common diseases including neurodegenerative disorders, cardiomyopathies, metabolic syndrome, cancer and obesity[2,3]. In plants, these organelles have also attracted considerable attention since they specify a widely expanded trait leading to an inability of plants to produce functional pollen, called 'cytoplasmic male sterility'. Plant mitochondria are thus of huge medical and agronomical interest[4,5].

Even though mitochondria have retained a genome and possess fully functional gene expression machinery, because most of the proteins acting in mitochondria (>95%) are nuclear encoded these organelles are only semi-autonomous[6]. While most genes retained in mitochondrial genomes are comparatively conserved across eukaryotes, the structure of genomes and gene expression machinery are highly divergent. These complex processes, involving many post-transcriptional steps, are poorly understood at the molecular level[7,8]. Over recent years, substantial progress has been made and key mitochondrial factors have been identified in humans, yeasts and plants[9–11]. In particular, in plants, a rapidly increasing number of reports identify PPR proteins as the major player in mitochondrial gene expression[10,12]. These proteins comprise one of the largest gene families in plants, with over 450 members in *Arabidopsis* alone[10]. They are involved in RNA editing, splicing, messenger RNA (mRNA) stabilization, maturation of transcript ends and translation. In plant mitochondria, PPR proteins were

also proposed to be involved in translation and two proteins were identified in high-molecular weight mitochondrial fractions containing ribosomes[13–15].

Nevertheless, mitochondrial translation, the last level of mRNA expression that is also the least amenable to simple molecular analyses, remains largely unexplored. Translation is the fundamental process of decoding the genetic message present on mRNAs into proteins. Despite its prokaryotic origin, mitochondrial translation machinery differs profoundly from its bacterial counterpart in many essential aspects, and major translation-associated components (mRNA, transfer RNA (tRNA) and ribosomes) are substantially different from their counterparts found in bacteria[16–18]. Indeed, mitochondria can use a genetic code distinct from that universally conserved. Moreover in animals, mitochondrial tRNAs do not always adopt the conventional clover leaf-shaped secondary structures[19,20]. Additionally, the constraints of translating the few mRNAs essentially encoding hydrophobic membrane subunits have also strongly influenced the evolution of mitochondrial gene expression machinery. Such differences have yielded highly specialized translation mechanisms guided by membrane-associated mitochondrial ribosomes (mitoribosomes), thus facilitating co-translational insertion of mitochondria-encoded proteins into the inner mitochondrial membrane[21]. This divergence from bacteria is also particularly obvious for translation initiation, since plant mitochondrial mRNAs lack the typical ribosome-binding sites in their 5′ leaders, also called the Shine–Dalgarno sequence, used in prokaryotes to aid in the correct positioning of the start codon at the P-site of the small ribosomal subunit (SSU). In addition, mammalian and *Chlamydomonas* mitochondrial mRNAs are devoid of 5′-untranslated regions (5′ UTRs)[7,8,22,23]. Consequently, the mechanisms by which ribosomes are recruited onto mitochondrial mRNA

5′ UTRs and the correct translation initiation codon recognized by the SSU remain elusive in most eukaryotes. Overall, the organization and regulation of protein synthesis in mitochondria is dependent on a highly degenerate prokaryotic scaffold animated by a large number of host-derived, co-evolved factors of which little is known.

Proteomic analyses have provided detailed catalogues of mitochondrial ribosomal proteins in *Saccharomyces cerevisiae* and mammals[24–27]. In plants, such analysis has not been carried out and the characterization of mitochondrial ribosomal proteins in *Arabidopsis* and rice is based solely on homology to bacterial ribosomal proteins[28]. High-resolution, three-dimensional (3D) structures of complete mitochondrial ribosomes have only recently been determined by cryo-EM, in yeast and two mammalian species[29–31]. These structures revealed that mitoribosomes are comparatively protein rich and that ribosomal RNAs (rRNAs) have been reduced compared to prokaryotic ribosomes, especially in animal mitoribosomes. Moreover, a tRNA molecule has structurally replaced 5S rRNA in the latter[32,33].

Here we determined experimentally the exhaustive protein content of *Arabidopsis* mitoribosomes and identified its global architecture, unravelling the unique structural characteristics of plant mitoribosomes. Our analysis reveals that plant mitoribosomes are larger than bacterial ribosomes and both animal and yeast mitoribosomes. Such greater molecular weight is in part due to a substantially large SSU following the inclusion of an additional rRNA domain and the integration of numerous PPR proteins, in agreement with the prevalence of this family of proteins in all plant organelle gene expression machinery.

Results

***Arabidopsis* mitochondrial ribosomal RNAs are significantly extended compared to those of bacteria.** *Arabidopsis* mitochondrial genome encodes 26S, 18S and 5S rRNAs (3,169, 1,935 and 118 nucleotides (nt) in length, respectively). Because of the prokaryote origin of mitochondrial rRNA genes, these ribosomal RNAs were compared to those of *Escherichia coli* to identify *Arabidopsis* rRNA-specific features. *Arabidopsis* mitoribosome SSU 18S rRNA is characterized by the presence of an additional rRNA domain issued from a large, 370 nt insertion located in helix 39 (h39) of the 3′ major domain (Supplementary Fig. 1). Phylogenetic analysis shows that this large h39 expansion is conserved in angiosperms, and expansions of shorter size are sometimes found in more basal plants (Supplementary Fig. 2a). In addition, other smaller insertions can be identified at various regions on the SSU. For instance, h6 of the *Arabidopsis* 18S rRNA 5′ domain is 56 nt larger compared to that of *E. coli*. This expansion is conserved in angiosperms, where it ranges from 56 nt in dicotyledons to 102 nt in monocotyledons (Supplementary Fig. 2a). Another short insertion of 47 nt is also found in h44 of the 3′ minor domain. In contrast, short rRNA segments present in *E. coli* SSU are missing from *Arabidopsis* 18S rRNA, that is in h8, h9, h10 and h17 of the 5′ domain, along with the missing anti-Shine–Dalgarno sequence at the 3′ end of this rRNA. *Arabidopsis* 26S large subunit (LSU) rRNA also diverges significantly from its prokaryotic counterpart, with domain III containing several short insertions in H52–H55 (Supplementary Fig. 1a). Altogether, as compared to prokaryotes, *Arabidopsis* mitochondria have larger SSU and LSU rRNAs, with the former being 20% larger and the latter 9% larger than those of *E. coli*.
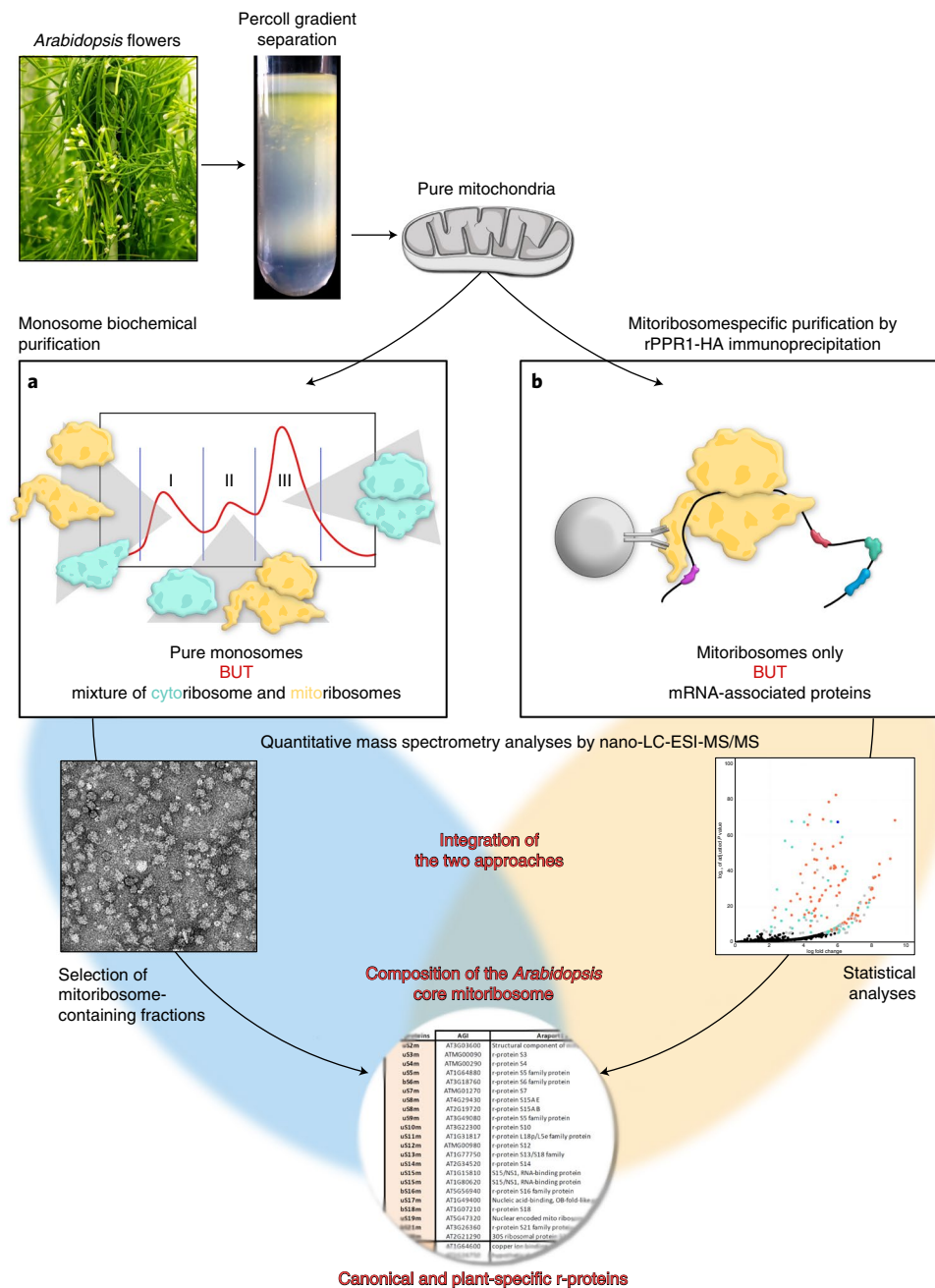
Strategy used to determine *Arabidopsis* mitochondrial ribosome protein content. Early biochemical studies described plant mitoribosomes as protein rich and identified approximate sedimentation coefficients of 77–78S, thus very close to 80S cytosolic ribosomes[34]. The precise characterization of plant mitoribosome protein content has been hampered by (1) the low solubility of mitoribosomes attached to the inner mitochondrial membrane[21] and (2) the

occurrence of cytosolic ribosomes on the outer surface of plant mitochondria[35], resulting in the co-purification of both mitoribosomes and cytoribosomes of almost equivalent sedimentation coefficients from purified plant mitochondria. In order to identify the comprehensive composition of *Arabidopsis* mitoribosomes, a strategy based on complementary approaches was used. Classical biochemical purification of ribosomes was combined with immuno-precipitation of mitoribosomes, using a specific *Arabidopsis* mitoribosome protein as bait and quantitative proteomics (Fig. 1).

In a previous study, the plant-specific PPR protein PPR336 was proposed to be associated with *Arabidopsis* mitoribosomes[13]. PPR336, renamed rPPR1 hereafter (for ribosomal PPR protein 1), has a close paralogue, At1g11630 (PPR336L), with 70% sequence identity between the two proteins. Because it was expected that these two proteins would have redundant functions, a *rppr1/At1g11630* double-knockout mutant line was transformed with a HA-HA-tagged rPPR1 construct, resulting in the rPPR1-HA plant line. Mitochondria were extracted from this line and used to assess the ability to immuno-purify rPPR1 with HA-specific antibodies. Immuno-purification was were then performed with increasing stringency (that is, with 100, 400 or 600 mM KCl for the wash step). As controls, immuno-purification was performed with HA antibodies and mitochondria extracted from WT Col-0 plants, and with Myc-specific antibodies in mitochondria extracted from rPPR1-HA plants. For each condition, experiments were performed in triplicate and immuno-purification proteins were identified by quantitative nano-LC-ESI-MS/MS. Proteins significantly enriched in rPPR1-HA in 12 immuno-precipitation experiments were identified by statistical analysis and are visualized as a volcano plot (Fig. 2 and Supplementary Table 1). Among a total of 1,625 proteins identified, 178 were significantly enriched (adjusted $P < 1 \times 10^{-5}$) in rPPR1-HA samples. Among these proteins, 81 were annotated as putative canonical mitoribosome proteins in databases (for example, Araport11, Pfam, UniProt), confirming that rPPR1 can indeed immuno-precipitate *Arabidopsis* mitoribosomes. Moreover, 44 PPR proteins were found among the enriched proteins. Among these, a specific group of nine PPR proteins was most significantly enriched (adjusted $P < 1 \times 10^{-50}$) (Fig. 2). Interestingly, when immuno-precipitation was performed under dissociative conditions (with 800 mM KCl), canonical SSU proteins were significantly enriched (Supplementary Fig. 3 and Supplementary Table 2), suggesting that rPPR1 may be associated with mitoribosome S

To distinguish between core mitoribosome proteins and other proteins co-purified with the translating mRNA, classic biochemical purification of mitochondrial monosomes (that is, free ribosomes) was performed. The ribosome purification procedure was derived from previous work[33], starting from purified WT *Arabidopsis* mitochondria. Ribosome separation on high-resolution 10–30% continuous sucrose gradients allowed the separation of 80S from 77–78S mitoribosomes, as well as from dissociated mitoribosome subunits that co-sedimented with the 40S cytoribosome SSU (Supplementary Fig. 4). All purification fractions were analysed by quantitative nano-LC-ESI-MS/MS. Analysis of ten samples from two independent monosome purifications identified 160 cytoribosome proteins and 75 proteins annotated as putative mitochondrial ribosomal proteins and additional proteins, including ten PPR proteins (Supplementary Table 3). Interestingly, these PPR proteins were always found in all the fractions, that is peaks I and II, containing the canonical mitoribosome proteins. They include the nine most highly enriched PPR proteins found by immuno-purification analysis.

Composition of the *Arabidopsis* mitoribosome proteome. Proteins were considered bona fide core mitoribosome proteins only if they were (1) significantly enriched in rPPR1 immuno-purification (adjusted $P < 1 \times 10^{-5}$) and (2) found in ribosome separation

Fig. 1 | Strategy for the identification of *Arabidopsis* mitochondrial ribosome protein composition. The mitoribosome proteome was determined by a combination of approaches. **a**, The classical purification of monosomes from WT plants; **b**, the immuno-purification of mitoribosomes from plants expressing a human influenza haemagglutinin (HA)-tagged rPPR1. Supplementary Table 4 explains the rationale used to integrate the two approaches and to define bona fide *Arabidopsis* core mitochondrial ribosome proteins.

peaks I and/or II (Supplementary Fig. 4 and Supplementary Table 4) in all biological replicates. This allowed the compilation of a list of 92 proteins (Table 1). Among these, some belong to families of identical function—for example, three examples of L12 were found. Thus, if a single protein is considered for each function, the *Arabidopsis* mitochondrial ribosome contains 81 proteins. Among these, 13 were previously found specifically in yeast and/or animal mitoribosomes[29–31]. Furthermore, 19 are entirely specific to plant mitoribosome, of which ten are PPR proteins (Table 1). Surprisingly, At1g11630 (PPR336L), a protein showing very high similarity to rPPR1, is not found in the *Arabidopsis* mitochondrial ribosome (that is, it was never identified in sucrose gradient fractions containing

mitoribosome proteins). The relative abundances of plant-specific proteins were then compared to those of canonical r-proteins. This revealed that both categories of protein are found at similar levels, that is 6.26 peptides detected on average for plant-specific proteins (numbers corrected for protein length) and 6.72 peptides for other proteins (Supplementary Fig. 5, Supplementary Table 6), thus suggesting that plant-specific and canonical r-proteins are present at similar stoichiometries in plant mitoribosomes. All novel rPPR proteins belong to the P-type subgroup of PPR proteins[36]. Structure predictions suggest that these PPR proteins fold into elongated super-helices (Supplementary Fig. 6). With the exception of five proteins, all mitoribosome proteins are predicted
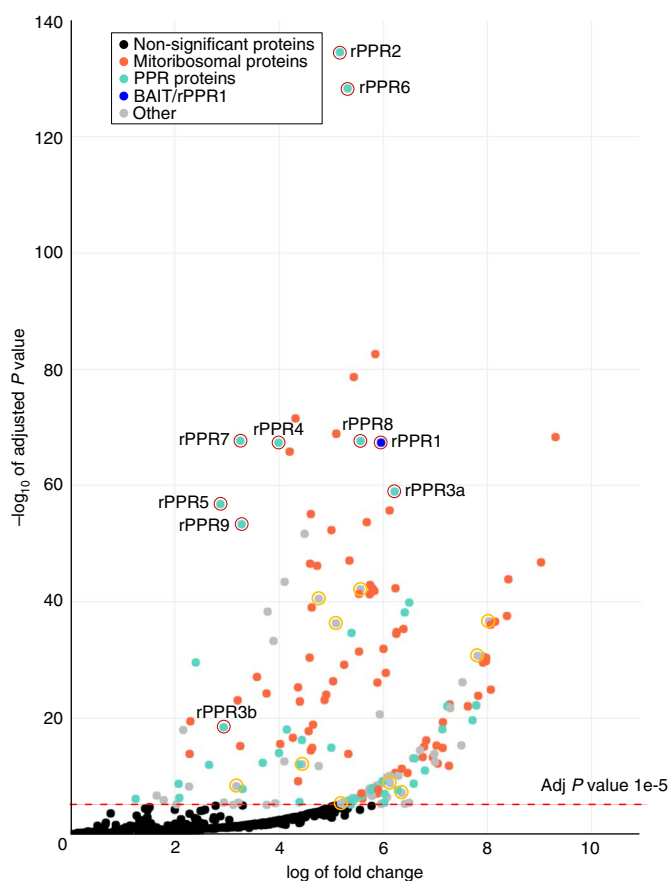
Fig. 2 | rPPR1 immuno-precipitates the *Arabidopsis* mitochondrial ribosome. Proteins statistically over-represented in specific rPPR1 immuno-purification as compared to control experiments are visualized as a volcano plot. Over-representation of proteins in rPPR1 IPs is represented on the *x* axis as log fold change, and statistical confidence in their enrichment is shown by decreasing *P* values on the *y* axis ($-\log_{10}$ of adjusted *P* values). The red dotted line indicates the significance threshold of 5 ($-\log_{10}$ (1×10^{-5})). A total of 178 proteins have an adjusted $P<1\times10^{-5}$ and are thus considered as bona fide immuno-purification partners of rPPR1. The bait rPPR1 is represented in blue, canonical mitoribosome proteins in orange, PPR proteins in turquoise, other proteins in grey and non-significant proteins in black. Nine rPPR proteins belong to the most significantly enriched proteins, with adjusted $P<1\times10^{-50}$. Additional plant-specific mitoribosomal proteins are highlighted in yellow, and rPPR proteins in red.

to localize to mitochondria (Table 1). Analysis of transcriptome databases revealed that transcripts encoding the ten ribosomal PPR proteins, hereafter referred to as rPPR1–9, as well as the additional plant-specific mitoribosome proteins, have expression levels similar to those of canonical mitochondrial r-proteins and follow similar expression patterns during plant development (Supplementary Fig. 7). Moreover, co-expression patterns of r-proteins were investigated with ATTED-II. The numbers of mitochondrial canonical r-protein genes co-expressed with each plant-specific r-protein were similar to those of mitochondrial canonical r-protein genes co-expressed with representative canonical r-proteins. This suggested that novel plant-specific and canonical r-proteins share similar co-expression patterns (Supplementary Fig. 8).

As a next step, the respective compositions of *Arabidopsis* mitoribosome individual subunits were investigated. For this, the 10–30% sucrose gradient peak I was further separated into four fractions as analysed by quantitative nano-LC-ESI-MS/MS (Supplementary Fig. 4).

Fractions 1–4 clearly revealed a decrease in SSU r-protein abundance and an increase in LSU r-protein abundance. For instance, fraction 1 contained mostly SSU and fraction 4 mainly LSU (Supplementary Fig. 4b). The distribution of novel rPPR proteins and additional plant-specific mitoribosome proteins among fractions 1–4 was investigated. Proteins following the distribution of SSU r-proteins were considered to be SSU, while those following the distribution of LSU r-proteins were considered LSU. This allowed us to conclude, for instance, that rPPR1 (PPR336[12]) is SSU while rPPR9 (PNM1[14]) is LSU (Table 1 and Supplementary Fig. 4c). Identified rPPR proteins distributed fairly equally between SSU and LSU.

Knockout of *rPPR* genes can result in lethality or impaired growth. To gain insights into the functions of novel rPPR proteins as part of the *Arabidopsis* mitoribosome, knockout mutants of the respective *rPPR* genes were analysed. Mutants were obtained for nine *rPPR* genes, displaying a wide array of distinct macroscopic phenotypes (Fig. 3). While *rppr3a* and *rppr8* homozygous mutants did not show any recognizable macroscopic phenotype, *rppr2* and *rppr9* resulted in embryo lethality[14,37]. In contrast, *rppr4*, *rppr7* and *rppr5* homozygous mutations, although viable, resulted in increasingly severe growth delay. In particular, rPPR7 mutation led to a severe phenotype characterized by dwarf plants, distorted leaves and yellowish cotyledons during the very first stage of development. The most severe growth delay was observed for *rppr5*, associated with lower seed production. *rppr1* mutation also resulted in growth delay. Although minor, this growth retardation was closely monitored at all rosette stages, revealing shorter roots as well as a reduction of 18% in projected leaf area ($P=0.0066$) 29 days after sowing of plants (Supplementary Fig. 9). Finally, homozygous mutation in *rppr3b*, a gene closely related to *rppr3a* (67% amino acid identity between rPPR3a and rPPR3b proteins), did not show any macroscopic phenotype but resulted in impaired seed production (Fig. 3). In summary, the variety of macroscopic phenotypes observed for *rPPR* mutants suggests that ribosomal PPR protein performs an array of distinct and non-redundant functions as part of the *Arabidopsis* mitoribosome.

Mitochondrial mRNAs are less efficiently translated in *rPPR1* mutants. In the next step, the molecular function of rPPR1 was investigated. While rPPR1 identification as a core mitoribosome protein confirmed previous assumptions proposing that it is associated with plant mitoribosomes[10,13], its function as part of the translation apparatus was not established. The minor, yet global, growth delay observed in *rppr1* mutants as compared to WT plants (Supplementary Fig. 9) could be explained by reduced translation efficiency in *Arabidopsis* mitochondria. In order to assess this hypothesis, ribosome-profiling experiments were performed. Here, RNA footprints from stalled mitoribosomes were prepared from both *rppr1* mutants and Col-0 WT plants and analysed by next-generation sequencing (Ribo-Seq analysis). Translation efficiencies were evaluated after normalization by mitochondrial mRNA abundances (Supplementary Fig. 10) as described previously[18]. This revealed relative mitoribosome densities along all mitochondria-encoded mRNAs in both mutant and WT plants. Interestingly, a two- to fourfold reduction in mitoribosome occupancy was observed for most mitochondrial mRNAs in *rppr1* mutant plants, strongly suggesting a general decrease of mitochondrial translation in the *rppr1* mutant (Fig. 4a). This global translational decrease was confirmed by analysis of the steady-state levels of mitochondria-encoded proteins. For instance, mitochondria-encoded Nad9, Cox1 and Cox2 showed reduced levels in *rppr1* while levels of mitochondrial, nuclear-encoded RISP, CYT*c*, AOX and PORIN were not affected (Fig. 4b). Together, our results support a function of rPPR1 in *Arabidopsis* mitochondrial translation. Growth reduction in the *rppr1* mutant

did not seem to be related to a reduction in the translation of a specific mitochondrial mRNA, although decrease in translation efficiency was variable among mitochondria-encoded mRNAs. In all cases, rPPR1 appears to be a generic translation factor required to obtain optimal translation levels of most, or all, mRNAs encoded in *Arabidopsis* mitochondria. This represents a novel function for PPR proteins. The absence of rPPR1 appears to limit the production of mitochondria-encoded proteins and, in turn, negatively impacts mitochondrial function and the global growth of *Arabidopsis* plants.

Arabidopsis **mitoribosome—unique structural features.** Finally, the global architecture of the *Arabidopsis* mitoribosome was determined by single-particle cryo-EM using an *Arabidopsis* monosome fraction prepared from purified mitochondria. While the majority of the sample could not be visualized as individualized particles, probably because of the low solubility of mitoribosomes and their tendency to aggregate, some individualized particles were identified. Image analysis and two-dimensional classification identified classes of particle with structural features clearly attributable to SSU- or LSU-like ribosome subunits. The 3D reconstructions

**Table 1 | *Arabidopsis* mitochondrial ribosome core proteins**

Proteins	AGI	Araport11	Protein amino acids	SUBA4	Note(s)
uS2m	At3g03600	Structural component of mito SSU	219	Mito	
uS3m	AtMg00090	r-protein S3	556	Mito	
uS4m	AtMt00290	r-protein S4	362	Mito	a
uS5m	At1g64880	r-protein S5 family protein	515	Mito	
bS6m	At3g18760	r-protein S6 family protein	139	Cyto	l
uS7m	AtMg01270	r-protein S7	148	Mito	a
uS8m	At4g29430	r-protein S15A E	129	Cyto	i, l
uS8m	At2g19720	r-protein S15A B	129	Cyto	i, l
uS9m	At3g49080	r-protein S5 family protein	430	Mito	
uS10m	At3g22300	r-protein S10	241	Mito	
uS11m	At1g31817	r-protein L18p/L5e family protein	314	Mito	j
uS12m	AtMg00980	r-protein S12	125	Mito	a
uS13m	At1g77750	r-protein S13/S18 family	154	Mito	
uS14m	At2g34520	r-protein S14	164	Mito	
uS15m	At1g15810	S15/NS1, RNA-binding protein	419	Mito	
uS15m	At1g80620	S15/NS1, RNA-binding protein	414	Mito	
bS16m	At5g56940	r-protein S16 family protein	135	Mito	
uS17m	At1g49400	Nucleic acid-binding, OB-fold-like protein	116	Cyto	l
bS18m	At1g07210	r-protein S18	261	Mito	
uS19m	At5g47320	Nuclear encoded mito ribosome subunit	212	Mito	
bS21m	At3g26360	r-protein S21 family protein	101	Mito	
bTHXm	At2g21290	30S ribosomal protein S31	98	Mito	
mS22	At1g64600	copper ion binding / methyltransferase	537	Mito	b,c
mS23	At1g26750	hypothetical protein	195	Mito	c
mS29	At1g16870	mito 28S ribosomal prot, S29-like protein	480	Mito	c
mS33	At5g44710	37S ribosomal protein S27	102	Mito	c
mS34	At5g52370	28S ribosomal S34 protein	142	Mito	c
mS35	At3g18240	Ribosomal protein S24/S35	419	Mito	c
mS35	At4g21460	Ribosomal protein S24/S35	415	Mito	c
mS47	At4g31810	Clp protease/crotonase family protein	409	Mito	b,c
mS75	At5g62270	ribosomal protein L20	420	Mito	d, g
mS76 (rPPR1)	At1g61870	PPR protein (PPR336)	405 (10)	Mito	d
mS77 (rPPR2)	At1g19520	PPR protein (NFD5)	725 (19)	Mito	d
mS78 (rPPR3a)	At1g55890	PPR protein	398 (10)	Mito	d
mS79 (rPPR3b)	At3g13160	PPR protein	394 (10)	Mito	d, f
mS80 (rPPR6)	At3g02650	PPR protein	576(13)	Mito	d
mS81 (rPPR8)	At5g15980	PPR protein	668 (16)	Mito	d
mS82	At4g22000	tyrosine sulfotransferase-like protein	130	Nucleus	d, l
mS83	At4g15640	adenylyl cyclase	390	Mito	d
mS83	At3g21465	adenylyl cyclase	388	Mito	d
mS84	At1g53645	hydroxyproline-rich glycoprotein family	523	Mito	d
mS85	At1g18630	glycine-rich RNA binding protein	155	Mito	d
mS86	At1g47278	hypothetical protein	91	Mito	d
mS87	At5g26800	xaa-pro aminopeptidase P	112	Mito	d
mrpX	Ag5g49210	stress response NST1-like protein	195	Mito	d,e

Continued

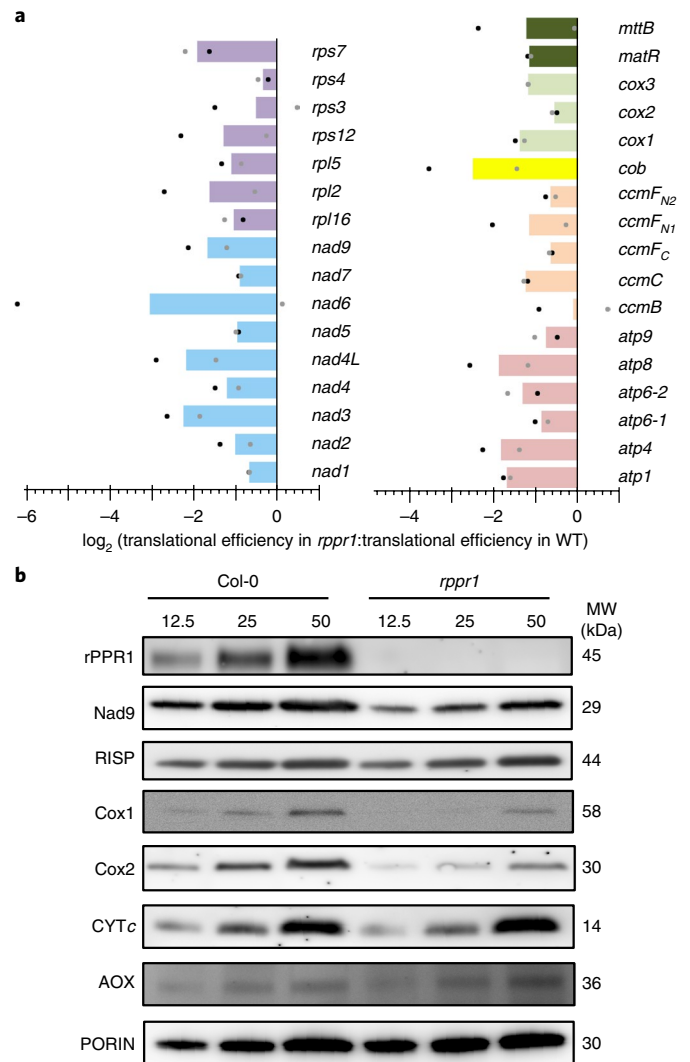## Table 1 | *Arabidopsis* mitochondrial ribosome core proteins (contined)

LSU					
uL1m	At2g42710	r-protein L1p/L10e family	415	Mito	
uL2m	At2g44065	r-protein L2 family	214	Mito	h
uL3m	At3g17465	Putative r-protein L3	324	Mito	
uL4m	At2g20060	r-protein L4/L1 family	300	Mito	
uL5m	AtMg00210	r-protein L5	185	Mito	a
uL6m	At2g18400	r-protein L6 family protein	102	Mito	
bL9m	At5g53070	r-protein L9/RNase H1	221	Mito	
uL10m	At3g12370	r-protein L10 family protein	171	Mito	
uL11m	At4g35490	r-protein L11	155	Mito	
bL12m	At3g06040	r-protein L12/ Clp protease family	186	Mito	
bL12m	At1g70190	r-protein L7/L12 domain- protein	208	Mito	
bL12m	At4g37660	r-protein L12/ Clp protease family	167	Mito	
uL13m	At3g01790	r-protein L13 family protein	205	Mito	
uL14m	At5g46160	r-protein L14p/L23e family protein	173	Mito	
uL15m	At5g64670	r-protein L18e/L15 superfamily protein	281	Mito	
uL16m	AtMg00080	r-protein L16	179	Mito	
bL17m	At5g09770	r-protein L17 family protein	160	Mito	
bL17m	At5g64650	r-protein L17 family protein	160	Mito	
uL18m	At5g27820	r-protein L18p/L5e family protein	114	Mito	
bL19m	At1g24240	r-protein L19 family protein	222	Mito	
bL20m	At1g16740	r-protein L20	126	Mito	
bL21m	At4g30930	r-protein L21	270	Mito	
uL22m	At1g52370	r-protein L22p/L17e family protein	269	Mito	
uL22m	At4g28360	r-protein L22p/L17e family protein	271	Mito	
uL23m	At4g39880	r-protein L23/L15e family protein	178	Mito	
uL24m	At5g23535	KOW domain-containing protein	159	Mito	
bL25m	At4g23620	r-protein L25/Gln-tRNA synthetase-like	277	Mito	
bL25m	At5g66860	r-protein L25/Gln-tRNA synthetase-like	249	Mito	
bL27m	At2g16930	r-protein L27 family protein	154	Mito	k
bL28m	At4g31460	r-protein L28 family	212	Mito	
uL29m	At1g07830	r-protein L29 family protein	144	Mito	
uL30m	At5g55140	r-protein L30 family protein	109	Mito	
bL31m	At5g55125	r-protein L31	76	Mito	
bL31m	At1g27435	hypothetical protein	81	Mito	
bL33m	At5g18790	r-protein L33 family protein	58	Mito	
bL36m	At5g20180	r-protein L36	103	Mito	k
mL40	At4g05400	copper ion binding protein	250	Mito	b,c
mL41	At5g40080	r-protein L27	94	Mito	c
mL41	At5g39800	r-protein L27	94	Mito	c
mL43	At3g59650	r-protein L51/S25/CI-B8 family protein	146	Mito	c
mL46	At1g14620	decoy	233	Mito	c
mL53	At5g39600	39S ribosomal protein	127	Mito	c
mL54	At3g01740	r-protein L37	126	Mito	c
mL101 (rPPR4)	At1g60770	PPR protein	491 (12)	Mito	d
mL102 (rPPR5)	At2g37230	PPR protein	757 (18)	Mito/Chloro	d
mL103 (rPPR7)	At4g36680	PPR protein	412 (10)	Mito	d
mL104 (rPPR9)	At5g60960	PPR protein (PNM1)	521 (12)	Mito	d
mL105	At3g51010	protein translocase subunit	188	Mito	d
mL106	At1g73940	tumor necrosis factor receptor protein	151	Mito	d

Light colours indicate canonical r-proteins, mid-colours (mSx or mLx) show mitoribosome-specific proteins previously found in animal and/or yeast mitoribosomes and dark colours show the novel plant-specific mitoribosome proteins. Proteins were allocated to either SSU or LSU according to the analysis presented in Supplementary Fig. 4. According to this distribution, proteins are annotated with the nomenclature recently established for r-proteins[72]. For each protein, the *Arabidopsis* gene identifier (AGI) is indicated as well as its Araport11 database annotation, the protein length in amino acids and the predicted subcellular localization as determined by SUBA4. For rPPR proteins, the numbers of PPR motifs of the respective proteins are indicated in parentheses. Notes: a, mitochondrial-encoded proteins with a nuclear duplicate gene on chromosome 2[73]. b, At1G64600, At4G31810 and At4G05400. These proteins were used to validate our criteria to identify novel core plant mitoribosome proteins. They were not initially annotated as *Arabidopsis* orthologues of mS22, mS47 and mL40, and were recognized as such only after in-depth sequence comparisons. c, mitochondria-specific r-proteins with orthologues in either yeasts, animals or both. d, novel plant-specific ribosomal proteins whose assignation to either SSU or LSU was determined as shown in Supplementary Fig. 4. e, the assignation of At5g49210 to LSU or SSU is unclear. f, the asignation of rPPR3b is also unclear, but it was assigned to SSU because of its high similarity to rPPR3a. g, At5g62270 is annotated as an L20 r-protein in databases, but its biochemical distribution is similar to that of SSU r-proteins. h, *Arabidopsis* has two uL2 genes, the mitochondrial AtMg00560 and the nuclear At2g44065; interestingly, only the nuclear-encoded uL2 protein is found in the *Arabidopsis* mitoribosome. i, Ancestral mitochondrial uS8 was lost in angiosperms and replaced by a cytosolic S15A protein[74]. j, At1g31817 is annotated as an L18p but has a clear S11 protein domain (IPR001971). k, On account of their low abundance, bL27 and bL36 did not match the defined criteria but were retained as core mitoribosome proteins because of their conservation in bacterial ribosomes and other eukaryote mitoribosomes, and also because they were always found in monosome peak II fractions (Supplementary Fig. 4). l, proteins not predicted to localize to mitochondria.
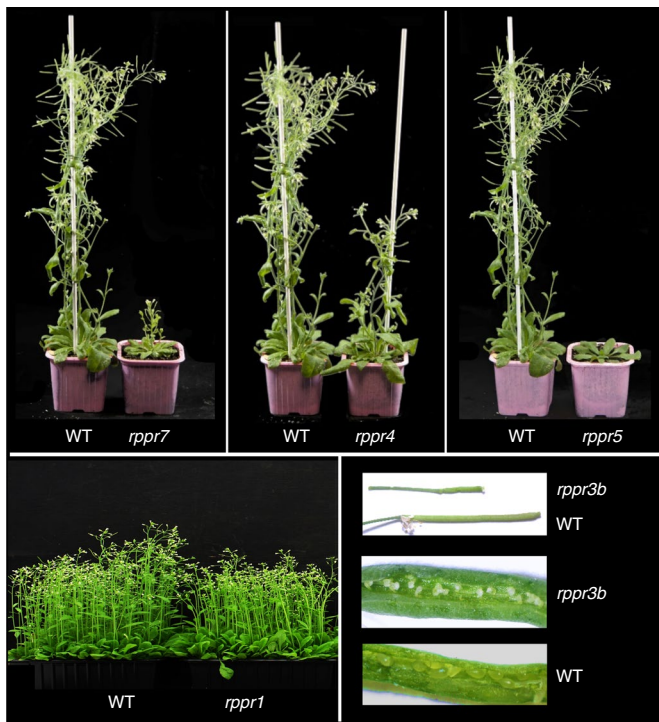
Fig. 3 | Macroscopic phenotyping of rPPR mutants. Seven-week-old *rppr7*, *rppr4*, *rppr5* and *rppr1* mutants and WT plants were compared to reveal macroscopic phenotypes. *rppr4* and *rppr5* both display a delayed growth phenotype, more marked in the case of the latter. Alongside delayed growth, *rppr7* mutants show a more severe phenotype with plants remaining stunted and pale leaves at the cotyledon stage. The *rppr1* mutant has a mildly delayed growth phenotype, quantified in Supplementary Fig. 9. The *rppr3b* mutant does not display retarded growth, but analysis of its siliques revealed that it is unable to produce seeds. Mutants *rppr2*, *rppr3a*, *rppr8* and *rppr9* are not shown here because *rppr2* and *rppr9* are lethal mutants while *rppr3a* and *rppr8* did not show any visible phenotype. The same phenotypes were observed for all plants (20 plants per genotype) over three independently repeated experiments (five independent experiments for *rppr1*).

Fig. 4 | *Arabidopsis* rPPR1-deficient plants have lower ribosome density along mitochondrial mRNAs. a, Ribo-Seq analysis comparing average mitoribosome density along mitochondria-encoded transcripts in WT (Col-0) and *rppr1* mutant plants. The histograms show $\log_2$ ratios according to Ribo-Seq RPKM, normalized by the abundance of each mitochondrial mRNA, for the *rppr1* mutant to the WT. Values are the means derived from two biological replicates, with dots indicating values for individual experiments. Purple, blue, dark green, light green, yellow, beige and pink bars represent mRNAs encoding ribosomal, complex I, maturase and protein transport, complex IV, complex III, *c*-type cytochrome and complex V proteins, respectively. **b**, Steady-state level analysis of mitochondrial proteins in WT and *rppr1* plants. Indicated protein amounts (µg) from crude mitochondrial preparations were loaded in each lane and probed with antibodies specific to indvidual proteins. Porin was used as control to check equal loading among samples. Two independent immmuno-detection experiments were performed. Molecular weight (MW) markers are indicated in kDa.

of these complexes at ~21 and 16 Å, respectively, identified them as *Arabidopsis* mitochondrial ribosomal SSUs and LSUs (Fig. 5). Despite numerous attempts to improve the resolution of both subunits, the very high structural heterogeneity present hindered any significant improvement. Such structural heterogeneity can be conformational and compositional. In fact we believe that the mitoribosome of *Arabidopsis thaliana* is substantially fragilized once extracted from its mitochondrial membrane, which explains the impossibility of resolving the structure of the full mitoribosome with both subunits present. Nevertheless, 3D reconstructions revealed the unique features of the *Arabidopsis* mitoribosome. In particular, it is characterized by a very large SSU, larger than the LSU. This SSU has a distinctive foot extension (expansion segment at h44) and a distinctive body protuberance. Moreover, its most remarkable feature is the occurrence of a very large (>200 Å) head extension (Fig. 5). *Arabidopsis* mitoribosome architecture was compared to that of an animal (*Sus scrofa*) mitoribosome[30]. While the LSUs of both animal and plant mitoribosomes seem to share a similar overall size and shape, plant SSUs are significantly larger—for instance, D_{max} is 1.5-fold higher (Fig. 5b). These distinctive head and foot extensions, as well as the body protuberance, are not present in the animal mitoribosome. Similarly, the *Arabidopsis* mitoribosome was compared to its cytoribosome. The mitoribosome SSU

is considerably larger, with a D_{max} 1.4-fold higher than that of the cytoribosome, and neither extensions (rRNA head domain and h44 expansion segment) nor protuberance are present in the cytoribosome (Fig. 5c). The *Arabidopsis* mitoribosome was also compared to that of yeast[31]. Both yeast proteins and rRNAs fitted within the *Arabidopsis* density (Fig. 6a). While foot and head extensions are not present in the yeast mitoribosome, the *Arabidopsis* body protuberance fits with yeast protein mS47, a protein of unknown
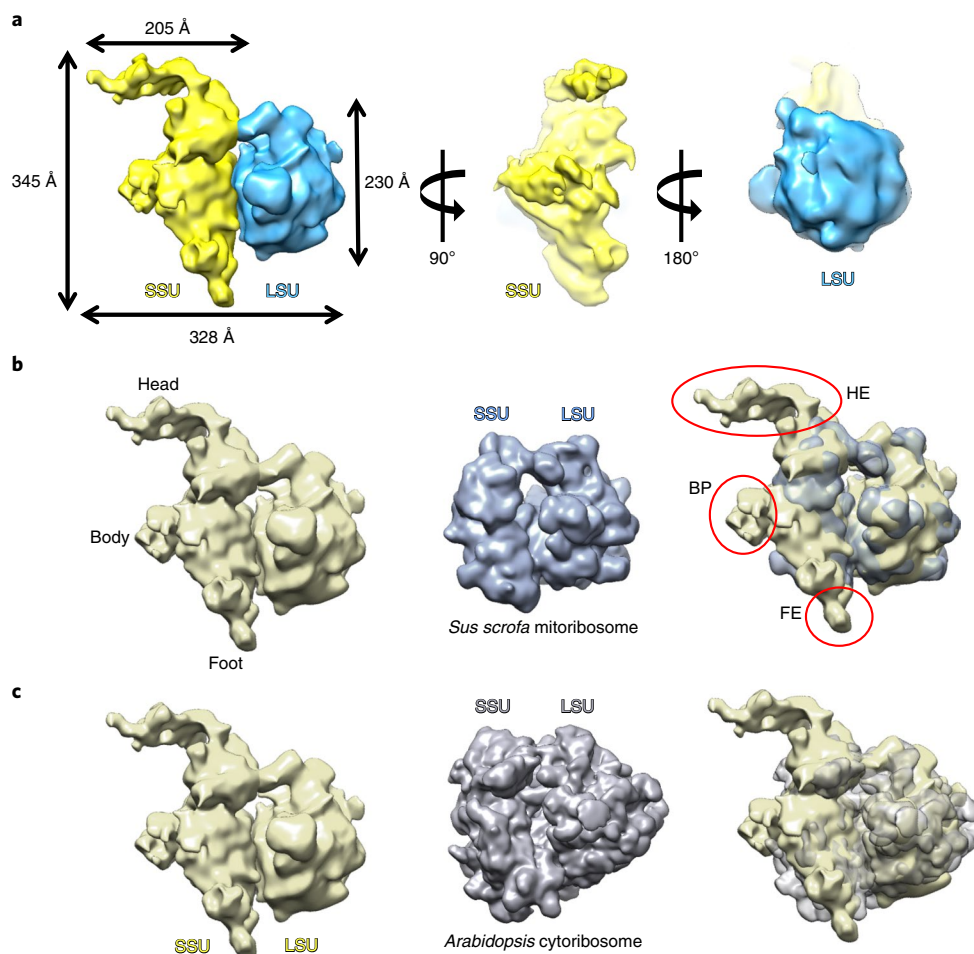
Fig. 5 | *Arabidopsis* mitoribosome cryo-EM map compared to animal mitoribosome and *Arabidopsis* cytoribosome. a, Segmented cryo-EM architecture of *A. thaliana* mitoribosome, seen from the SSU (yellow) and LSU (blue) aspects. Ribosome dimensions are indicated in angstroms. **b,** Structural comparison of *Arabidopsis* (ivory) and *Sus scrofa* mitoribosomes (grey/blue) (PDB:5AJ4[30]). The *Arabidopsis* LSU is comparable to animal LSU, but the *Arabidopsis* SSU shows major differences—a large head extension (HE) and a small foot extension (FE)—as well a clear body protuberance (BP). **c,** Structural comparison between *A. thaliana* mitoribosome and cytoribosome shows that plant mitoribosome-specific HE, FE and BP are also absent from the plant cytoribosome. The *A. thaliana* 80S cytoribosome architecture presented here was also determined during this analysis. It illustrates all the structural characteristics typical of eukaryotic ribosomes. This reconstruction is not discussed in the manuscript.

function previously found specifically in the yeast mitoribosome. Because an *Arabidopsis* mS47 protein is also associated with the *Arabidopsis* mitoribosome (Table 1), the body protuberance most probably corresponds to *Arabidopsis* mS47. Finally, the *Arabidopsis* mitoribosome was fitted with the *E. coli* rRNA structure (Fig. 6b). Interestingly, *Arabidopsis* SSU foot and head extensions are found at sites where specific *Arabidopsis* 18S rRNA insertions are localized. In particular, the site in h39 where the Arabidopsis-specific 370 nt domain is inserted (Supplementary Fig. 1) corresponds exactly to the root of the head expansion in the *Arabidopsis* mitoribosome SSU. This expansion is predicted to form an elongated RNA structure (Supplementary Fig. 1c). It is thus very probable that the large head extension corresponds to the 370 nt *Arabidopsis*-specific 18S rRNA additional specific domain, possibly coated by plant-specific mitoribosome proteins.

Discussion

This analysis shows that the *Arabidopsis* mitochondrial ribosome differs substantially from both prokaryotic ribosomes and yeast and animal mitochondrial ribosomes, both structurally and in its composition. Although of bacterial origin, these ribosomes have significantly and specifically diverged from their prokaryotic ancestor.

For instance in plants, mitoribosomes have a massive SSU characterized by a large head extension, show major changes in rRNAs and have recruited many eukaryotic, plant-specific, additional proteins of which ten are PPR proteins.

Diversity of mitoribosome rRNAs in evolution. Major rearrangements of rRNAs occurred during eukaryote evolution. In particular, 5S rRNA has been lost from yeast and animal mitochondrial genomes, is absent in the respective mitoribosomes and is functionally replaced by a tRNA in animals[32,33,38]. In contrast, 5S rRNA has been maintained in plant mitochondrial genomes and probably occurs in its ribosome, as suggested by its detection in sucrose gradient fractions containing plant mitoribosomes[13]. Interestingly rearrangements of rRNAs correlate with mitoribosome protein contents. For instance, as compared to *E. coli*, the loss of h8, h9 and h44 from *Arabidopsis* 18S rRNA correlates with the loss of r-protein S20 from the plant mitoribosome SSU (in *E. coli* ribosome, S20 interacts tightly with h8, h9 and h44). The plant mitoribosome 18S rRNA head expansion segment observed here is the largest described to date in mitoribosomes. We qualify this head extension as an additional rRNA domain because it appears to fold into a domain, as seen in our low-resolution structure, unlike the
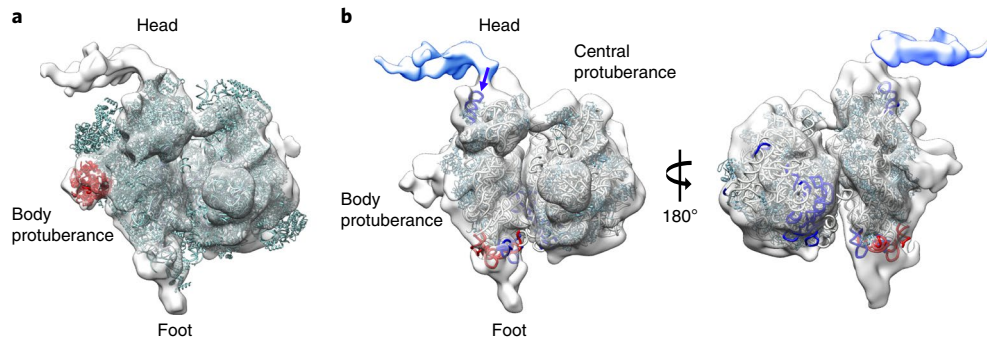
Fig. 6 | Architecture of *Arabidopsis* mitoribosome fitted with *E. coli* and yeast mitochondria ribosomes. a, Architecture of *Arabidopsis* mitoribosome fitted with yeast mitoribosome high-resolution structure (PDB:5MRC[31]). Proteins and RNA chains of the yeast mitoribosome are shown in pale green. This strongly suggests that the yeast-specific mS47 protein, shown in red, is also present at a density corresponding to the *Arabidopsis* mitoribosome body protuberance. **b**, *Arabidopsis* mitoribosome fitted with *E. coli* ribosome high-resolution structure (PDB:Y4BB[75]). *E. coli* r-proteins are shown in light blue and rRNAs in white. Major differences between *Arabidopsis* mitochondria and *E. coli* ribosomal rRNAs are indicated as follows: blue ribbons indicate sites where *Arabidopsis* mitochondria-specific rRNA segments are inserted, and red ribbons show *E. coli* rRNA segments not present in *Arabidopsis*. Major differences can be observed for the SSU. In particular, *Arabidopsis* 18S rRNA has a major expansion segment of 370 nt inserted at h39 (blue arrow), predicted to be folded into an elongated RNA structure (Supplementary Fig. 1c). This is located near the area of additional density on the map, at the head of the *Arabidopsis* mitoribosome. The density corresponding to the head extension is shown in light blue. Other *Arabidopsis* expansions on the map can also be seen at the foot, close to the site where extensions h6 and h44 are located.

majority of expansion segments that are smaller and often incomparably more flexible. In eukaryotic ribosomes, expansion segments form intricate networks with eukaryote-specific proteins or protein extensions[39]. Therefore this plant-specific rRNA domain is likely to bind mitochondrial plant-specific r-proteins, such as elongated rPPR proteins (Supplementary Fig. 6) and/or other translation factors related, for example, to translation initiation, which remains elusive in plant mitochondria.

Occurrence of PPR protein as part of mitochondrial translation apparatus. Given the huge number of PPR proteins (>450) found in plant genomes and their wide implication in all organellar gene expression processes, it may not be surprising to find PPR proteins associated with plant mitoribosomes. However, in chloroplasts where PPR proteins are as prominent in regard to gene expression as in plant mitochondria, the ribosome strongly resembles its bacterial counterpart and does not involve any PPR protein[40,41]. Likewise, the chlororibosome does not contain the SSU additional domains observed here in the plant mitoribosome. In other eukaryote groups PPR proteins are also found, although in lower numbers, that is ~30 in kinetoplastids and ~10 in animals. Among these, *Trypanosoma brucei* PPR proteins were also proposed to be associated with mitoribosomes, because one PPR protein co-purifies with the LSU rRNA[42] and several PPR proteins were found during affinity purification of ribosome particles[43]. It is unclear whether these proteins are directly associated with mitoribosomes or whether they were co-purified with the translating mRNA, because they were also found in the polyadenylation complex[44]. Functional analysis of two of these PPR proteins suggests that they are at the interface between polyadenylation and translation, and that they are essential in activating mRNA for translation[44]. A recent study on the structure of the trypanosome mitoribosome confirmed that seven PPR proteins are indeed core mitoribosome proteins[45]. Likewise, structural analysis of the animal mitoribosome revealed the occurrence of two PPR proteins (mS39 and mS27) in the small mitoribosomal subunit. For instance, mS39 is localized close to the mRNA entry channel; its molecular function is unknown but is suspected to be involved in mRNA recruitment[30,46]. In contrast, investigation of yeast mitoribosome structure did not identify any PPR protein[31]. Nevertheless, other yeast, animal and plant PPR proteins were described as regulators of translation

but are apparently not part of mitoribosomes[15,47–51]. It is noteworthy that plant rPPR proteins are neither related to other trypanosome, yeast and animal PPR proteins found in mitoribosomes, nor related to translation. However, all those plant and eukaryote rPPR proteins belong to the P-class of PPR proteins, a sub-class of PPR proteins believed to represent their ancestral form[10].

Functions of novel rPPR proteins in *Arabidopsis* mitoribosome. Among the ten rPPR proteins identified in this analysis, rPPR1, rPPR3a, rPPR3b and rPPR7 are evolutionarily related as they belong to the plant-specific subgroup of rPPR1-like proteins (also named the PPR336 family[13]). This group of proteins, which has seemingly arisen in gymnosperms (Supplementary Fig. 11), is characterized by their short length (with ten PPR motifs) in comparison to most other PPR proteins. All other rPPRs are larger, rPPR2 and rPPR5 being the largest with 18–19 PPR motifs. Besides rPPR1-like proteins, other rPPRs are more ancient because probable orthologues can be found in more basal plant groups such as mosses and ferns (Supplementary Fig. 11). Given the PPR protein mode of RNA recognition, with individual PPR motifs recognizing individual ribonucleotides[10,12], shorter rPPR proteins may be less specific or less tightly bound to RNA. In all cases, the diversity of *rppr* knockout mutant phenotypes showed that individual rPPRs do not share redundant functions but rather seem to maintain an array of independent and essential functions as part of the mitoribosome. Ribosome-profiling analysis of the *rppr1* mutant showed that rPPR1 is a generic mitochondrial translation factor, because its depletion resulted in impaired translation for all *Arabidopsis* mitochondria-encoded mRNAs. This is in contrast to most characterized PPR proteins that specifically recognize a single or small group of RNA targets[10]. As part of the plant mitoribosome, rPPR protein functions may help in the recruitment and docking of mRNAs to the SSU. In addition, they may be involved directly in the translation reaction, for example through the stabilization of rRNAs. Some rPPRs may be generic translation factors similar to rPPR1, while others may have functions related to the specific recruitment or translation of individual mitochondrial mRNAs.

Diversity of mitochondrial translation machinery in eukaryotes. Contrary to cytosolic ribosomes that seem to share similar and

structure across eukaryotes[52,53], mitochondrial ribosomes appear to have evolved a remarkable array of varying architecture. Despite a common bacterial origin, mitoribosomes from plants, animals and fungi have surprisingly different RNA and protein compositions. This evolution may have been driven by the specific divergence of mitochondrial genomes and gene contents, along with the need to translate mRNAs encoding different subsets of mitochondrial proteins in the respective eukaryote groups. Alternatively, it may have resulted from an adaptation to specific environmental constraints encountered by different eukaryotes. Nonetheless, despite this diversity, some proteins (for instance, mS23–mS33 and mL40–mL54) are found in all animal, yeast and plant mitoribosomes. These proteins may represent ancestral eukaryote-specific factors that were acquired early in the evolution of the mitochondrial translation apparatus. In plants, the larger size and specific features of the mitoribosome SSU probably reflect specific processes for the initiation and/or regulation of translation not involving the Shine–Dalgarno bacterial-like translation initiation signal or other recognizable conserved sequence or structure elements[54]. The conservation of rRNA structural features—that of the large head insertion as well as the conservation of specific rPPR proteins—suggests that the plant mitoribosome composition and organization described here are conserved, at least for all flowering plants.

In conclusion, animal mitoribosomes, as compared to prokaryotes, have reduced rRNAs compensated by higher numbers of proteins[17] whereas plant mitoribosomes have the largest rRNAs and equivalent or higher numbers of protein subunits as compared to other mitoribosomes. The plant mitoribosome is thus significantly larger than the bacterial and animal or yeast mitoribosome, with a size comparable to that of the typical eukaryote cytosolic ribosome. The *Arabidopsis* mitoribosome thus represents one of the largest and most complex ribosomes described to date. The in-depth structural and functional characterization of its components will reveal specificities of the plant mitochondria translation process, and will help in understanding the diversity and evolution of translation machinery from prokaryotes to eukaryotes.

Methods

Plant materials. *Arabidopsis thaliana* Col-0 plants were obtained from the Institut National de la Recherche Agronomique Stock Centre, Versailles. Plants were grown in soil under long-day conditions (16 h light:8 h dark). Mutant plants were obtained from SALK T-DNA, SAIL T-DNA and GABI-Kat insertion line collections. Transfer DNA insertion lines were selected and retained if the insertions were located in an exon in gene-coding sequences. The insertions were confirmed by PCR amplification and sequencing. The *rppr1* T-DNA insertion mutants are Salk_037390 and Salk_139562[13]. One mutant for At1g19520 (*rppr2*) was previously characterized as *nfd5*[37]. Another line where the insertion is located in exon2 of the gene, Salk_120951, confirmed the lethal phenotype. Mutant At1g55890 (*rppr3a*) is Salk_113426; At3g13160 (*rppr3b*) is Salk_009440; At1g60770 (*rppr4*) is GK_645B03; At2g37230 (*rppr5*) is Sail_1146_C06; At4g36680 (*rppr7*) is Sail_358_D12; and At5g15980 (*rppr8*) mutants are Salk_122059 and GK_803H10. Mutant At5g60960 (*rppr9/pnm1*) was previously characterized[14]. The primers used for genotyping are presented in Supplementary Table 7. For the rPPR1-HA plant line, At1g61870 was cloned with its endogenous promoter (1 kb upstream of its initiation codon) into the pGWB1 binary vector, in fusion with a single C-terminal HA tag. The *rppr1/At1g11630* knockout line was transformed by floral dip in *Agrobacterium tumefaciens* GV3101[55]. Plants were selected for hygromycin resistance and rPPR1-HA expression.

Purification of *A. thaliana* mitochondria. Mitochondria were purified starting either from *Arabidopsis* flowers or *Arabidopsis* cell culture grown under dark conditions in cell medium (4.41 g l⁻¹ Murashige and Skoog basal medium (Duchefa M0256) supplemented with 30 g l⁻¹ sucrose, 500 μl l⁻¹ NAA (2 mg ml⁻¹) and 25 μl l⁻¹ Kinetin (2 mg ml⁻¹), pH 5.6). Starting material was blended in extraction buffer containing 0.45 M mannitol, 50 mM sodium pyrophosphate (10.H$_2$O), 0.5% BSA, 0.5% (w/v) polyvinylpyrrolidone-40, 20 mM MgCl$_2$ and 20 mM cysteine, pH 8.0. Lysate was filtered and clarified by centrifugation at 1,500 g for 10 min at 4 °C. The supernatant was kept and centrifuged at 18,000 g for 15 min at 4 °C. The organelle pellet was re-suspended in wash buffer (0.3 M mannitol and 10 mM TES-KOH, pH 7.5) and the above centrifugation processes were repeated once. The resulting organelle pellet was re-suspended in wash buffer, loaded on a 10–23–40% Percoll

gradient (in wash buffer) and run for 45 min at 40,000 g. Mitochondria were collected at the 23–40% inter-phase.

Purification of mitochondrial ribosomes. Mitochondria were re-suspended in lysis buffer (20 mM HEPES-KOH, pH 7.6, 100 mM KCl, 30 mM MgCl$_2$, 1 mM DTT, 0.5 mg ml⁻¹ heparin, 1.6% Triton X-100, 100 μg ml⁻¹ chloramphenicol, supplemented with protease inhibitors (Complete EDTA-free)) to a concentration of 1 mg ml⁻¹ and incubated for 15 min at 4 °C. Lysate was clarified by centrifugation at 30,000 g, for 20 min at 4 °C. The supernatant was loaded on a 50% sucrose cushion in Monosome buffer (as lysis buffer but with no Triton X-100, and 50 μl ml⁻¹ chloramphenicol) and centrifuged at 235,000 g for 3 h at 4 °C. The crude ribosome pellet was re-suspended in Monosome buffer, loaded onto a 10–30% sucrose gradient in the same buffer and run for 16 h at 65,000 g. Fractions corresponding to mitoribosomes were collected, pelleted and re-suspended in Monosome buffer.

Immuno-precipitation of mitochondrial ribosomes. Immuno-precipitation were performed with protein extracts using the μMACS HA-Tagged Protein Isolation Kit (Miltenyi Biotec). Mitochondria corresponding to 1 mg protein were lysed in lysis buffer (20 mM HEPES-KOH, pH 7.6, 100–800 mM KCl, 30 mM MgCl$_2$, 1 mM DTT, 0.5 mg ml⁻¹ heparin, 1% Triton X-100 supplemented with protease inhibitors (Complete EDTA-free)) for 30 min at 4 °C on a rotating wheel. The lysate was clarified at 10,000 g, 10 min at 4 °C and the supernatant was retained and supplemented with 50 μl of anti-HA magnetic beads. The mix was incubated for 30 min on a rotating wheel and loaded onto the column. After loading, the column was washed four times with 200 μl Wash buffer (as lysis buffer but with 0.1% Triton X-100), and elution was performed with 120 μl elution buffer (Miltenyi Biotec) and with magnetic beads. The resulting co-immuno-precipitation proteins were analysed by liquid chromatography–mass spectrometry.

Proteomic and statistical analyses of mitochondrial ribosome composition. Mass spectrometry analyses of ribosome fractions were performed on the Strasbourg–Esplanade proteomic platform. In brief, protein extracts were precipitated (cold 0.1 M ammonium acetate in 100% methanol) and proteins were further digested with sequencing-grade trypsin. Each sample was analysed by nano-LC-ESI-MS/MS on a Q Exactive Plus mass spectrometer coupled to an EASY-nanoLC-1000 (Thermo-Fisher Scientific) with a 160 min gradient. Peptides were re-suspended in 30 μl of water containing 0.1% (v/v) formic acid (solvent A). Five microlitres of each sample were loaded onto a C-18 pre-column (75 μm ID × 20 mm nanoViper, 3 μm Acclaim PepMap; Thermo-Fisher Scientific) at 800 bars in solvent A. After de-salting and concentration, the pre-column was switched online with the C18 analytical column (75 μm ID × 25 cm nanoViper, 3 μm Acclaim PepMap; Thermo-Fisher Scientific) equilibrated in solvent A:solvent B (95:5; v/v). Peptides were eluted at a flow rate of 300 nl min⁻¹ using a gradient of 5–20% B for 120 min, 20–32% B for 15 min, 32–95% B for 1 min and 95–95% B for 24 min. Q Exactive Plus was operated in data-dependent acquisition mode with Xcalibur software. Survey MS scans were acquired at a resolution of 70 K at 200 m z⁻¹ (mass range 350–1,250), with maximum injection time of 100 ms and an automatic gain control set at 3e6. At least ten of the most intense multiply charged ions (≥2) were selected for HCD (higher energy collisional dissociation) fragmentation, with normalized collision energy set at 27, at 17.5 K resolution, maximum injection time 100 ms and automatic gain control 1 × 10⁻³. A dynamic exclusion time of 20 s was applied during the peak selection process. Raw files were finally transformed into mgf files using Proteome Discoverer software (v2.0). Data were searched against the TAIR *A. thaliana* database with a decoy strategy (release TAIRv10, 27281 forward protein sequences). Peptides and proteins were identified with the Mascot algorithm (version 2.5.1, Matrix Science) and data were further imported into Proline v1.4 software (http://proline.profiproteomics.fr/). Proteins were validated on Mascot pretty rank equal to 1, and 1% FDR on both peptide spectrum matches (PSM score) and protein sets (protein set score). The total number of MS/MS fragmentation spectra was used to quantify each protein from three independent biological replicates (spectral count label-free relative quantification). Proline was further used to align the spectral count values across all samples. To identify significantly enriched proteins, a statistical analysis by the msmsTests R package using spectral counts was performed[56]. The entire MS dataset was first normalized by the total number of MS/MS spectra (column-wise normalization). The negative binominal model, which is based on the solution provided by the edgeR package, was used[57]. *P* values were then adjusted using the Benjamini–Hochberg method. Proteins that were over-represented in rPPR1-HA immuno-precipitation were visualized as a volcano plot showing log$_2$-fold change and –log$_{10}$ *P* value on the x and y axes, respectively. The graphic was plotted using the Plotly R graphing library.

Protein immuno-detection. Mitochondrial proteins were separated using SDS–PAGE and their apparent molecular mass estimated with a Pageruler pre-stained protein ladder (Fermentas). Gels were transferred to polyvinylidene difluoride membrane and incubated with antibodies directed against PORIN (1:300 dilution; a gift of D. Day, University of Western Australia), Nad9 (1:2,000 dilution[58]), AOX (1:100 dilution[59]), Nad7 (1:2,000 dilution[60]), Cox1 (1:500 dilution; Abcam), Cox2 (1:1,000 dilution; Agrisera, AS04-053A), RISP (1:5,000 dilution[61]), CYT*c* (1:5,000 dilution; Agrisera) and rPPR1(PPR336) (1:2,000 dilution[13]).

Ribo-Seq analysis. Mitoribosome footprints were prepared as previously described[18]. Next-generation sequencing was performed by the I2BC sequencing facility (Illumina NextSeq technology, single end, 75 nt). reads per kilobase per million mapped reads (RPKM) values were calculated based on the number of reads mapping to mitochondrial coding sequences after normalization to the number of reads mapping to nuclear coding sequences. Mitochondrial mRNA translation efficiency was evaluated by dividing RPKM values by mitochondrial mRNA abundance, which was determined by RT-quantitative PCR as described by Planchard et al.[18].

Quantitative RT-PCR. The abundance of mitochondrial mRNAs (mtRNAs) was determined by RNA extracted from aliquots of Ribo-Seq initial lysates before the addition of RNase I, and by quantitative RT-PCR as described in ref. [18].

Bioinformatic analyses. Subcellular localization predictions were determined with SUBA4[62]. Protein similarities were analysed by MUSCLE alignment[63]. Predictions of PPR domain were performed with TPRpred[64]. Tri-dimensional predictions were built with Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2)[65] and molecular representations were prepared with PyMol. Gene co-expression data were acquired with ATTED-II[66].

Single-particle cryo-EM data collection. An *Arabidopsis* ribosome fraction (4 μl of 70 nM) prepared from purified mitochondria was applied to 400-mesh holey carbon Quantifoil 2/2 grids (Quantifoil Micro Tools), blotted with filter paper on both sides for 1.5 s in a temperature- and humidity-controlled Vitrobot Mark IV (FEI) (temperature 4 °C, humidity 100%, blot force 5, blot waiting time 30 s) and vitrified in liquid ethane pre-cooled by liquid nitrogen. Data were collected with a Titan Krios S-FEG instrument (FEI) operating at 300 kV acceleration voltage and a nominal under-focus of Δz = −0.6 to −4.5 μm, using the second-generation, back-thinned direct electron detector CMOS (Falcon II) 4096 × 4096 camera and automated data collection with EPU software (FEI). The Falcon II camera was calibrated at a nominal magnification of ×59,000. The calibrated magnification on the 14 μm pixel camera was ×127,272, resulting in pixel size of 1.128 Å at the specimen level. The camera was programmed to collect seven frames (starting from the second) out of the 17 possible. Total exposure was 1 s, with a dose rate of 60 ē per Å² (that is, 3.5 ē per Å² per frame).

EM image processing. A framework for image processing with an integrated software package (SCIPION)[67] was used to obtain 3D reconstructions of the large and small ribosomal subunits of the *A. thaliana* mitochondrial ribosome. Before particle picking, seven frames in the stack were aligned using the optical flow algorithm integrated in Xmipp3 and Scipion[67,68]. An average image of the whole stack was then used to determine the contrast transfer function by CTFFIND4[69] and to select, semi-automatically, ~318,000 particles in SCIPION[70]. Particle sorting was done by 3D classification using RELION[71], resulting in three key 3D classes comprising ~19,000 (~6%), ~24,000 (~7.5%) and ~107,000 (~34.5%) of particles showing structural features attributable to SSU-like, LSU-like and 80S particles, respectively. The remainder of the particles selected (~52%) was of a very heterogeneous composition, with high noise, and particle aggregation could not be utlized. The SSU- and LSU-like particles were interpreted as being the *A. thaliana* mitochondrial ribosomal subunits that appeared to represent a high degree of structural heterogeneity, which explains their apparent low resolution. Both classes were refined using RELION's 3D auto-refine, and the final refined classes (~16 Å for mitoribosome LSU and ~21 Å for SSU; particles were binned twice to yield a pixel size of 2.256 Å) were then post-processed using the procedure implemented in RELION and were applied to the final maps for appropriate masking, B-factor sharpening and resolution validation to avoid over-fitting[71]. Homogenous, fully assembled mitochondrial ribosome particles were retrieved in insufficient amounts for practical reconstruction. Cryo-EM data collection, refinement and validation statistics are given in Supplementary Table 8.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Mass spectrometric data were deposited with the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD010324. Ribo-Seq sequencing data were deposited in the NCBI Gene Expression Omnibus under accession number GSE119655. Cryo-EM data were deposited in EMDataBank under accession number EMDB-4408 for *Arabidopsis* mitoribosome SSU, and EMDB-4409 for *Arabidopsis* mitoribosome LSU.

References

1. Roger, A. J., Muñoz-Gómez, S. A. & Kamikawa, R. The origin and diversification of mitochondria. *Curr. Biol.* **27**, R1177–R1192 (2017).
2. Nunnari, J. & Suomalainen, A. Mitochondria: in sickness and in health. *Cell* **148**, 1145–1159 (2012).
3. De Silva, D., Tu, Y. T., Amunts, A., Fontanesi, F. & Barrientos, A. Mitochondrial ribosome assembly in health and disease. *Cell Cycle* **14**, 226–2250 (2015).
4. Chen, L. & Liu, Y.-G. Male sterility and fertility restoration in crops. *Annu. Rev. Plant. Biol.* **65**, 579–606 (2014).
5. Horn, R., Gupta, K. J. & Colombo, N. Mitochondrion role in molecular basis of cytoplasmic male sterility. *Mitochondrion* **19**, 198–205 (2014).
6. Giegé, P., Sweetlove, L. J., Cognat, V. & Leaver, C. J. Coordination of nuclear and mitochondrial genome expression during mitochondrial biogenesis in *Arabidopsis*. *Plant Cell* **17**, 1497–1512 (2005).
7. Lightowlers, R. N., Rozanska, A. & Chrzanowska-Lightowlers, Z. M. Mitochondrial protein synthesis: figuring the fundamentals, complexities and complications of mammalian mitochondrial translation. *FEBS Lett.* **588**, 2496–2503 (2014).
8. Hammani, K. & Giegé, P. RNA metabolism in plant mitochondria. *Trends Plant Sci.* **19**, 380–389 (2014).
9. Lopez Sanchez, M. I. G. et al. RNA processing in human mitochondria. *Cell Cycle* **10**, 2904–2916 (2011).
10. Barkan, A. & Small, I. Pentatricopeptide repeat proteins in plants. *Annu. Rev. Plant Biol.* **65**, 415–442 (2014).
11. Chinnery, P. F. et al. The challenges of mitochondrial replacement. *PLoS Genet.* **10**, e1004315 (2014).
12. Giegé, P. Pentatricopeptide repeat proteins. *RNA Biol.* **10**, 1417–1418 (2013).
13. Uyttewaal, M. et al. PPR336 is associated with polysomes in plant mitochondria. *J. Mol. Biol.* **375**, 626–636 (2008).
14. Hammani, K. et al. An *Arabidopsis* dual-localized pentatricopeptide repeat protein interacts with nuclear proteins involved in gene expression regulation. *Plant Cell* **23**, 730–740 (2011).
15. Haïli, N. et al. The MTL1 pentatricopeptide repeat protein is required for both translation and splicing of the mitochondrial NADH dehydrogenase subunit 7 mRNA in *Arabidopsis*. *Plant Physiol.* **170**, 354–366 (2016).
16. Mai, N., Chrzanowska-Lightowlers, Z. M. A. & Lightowlers, R. N. The process of mammalian mitochondrial protein synthesis. *Cell Tissue Res.* **367**, 5–20 (2017).
17. Greber, B. J. & Ban, N. Structure and function of the mitochondrial ribosome. *Annu. Rev. Biochem.* **85**, 103–132 . (2016).
18. Planchard, N. et al. The translational landscape of *Arabidopsis* mitochondria. *Nucleic Acids Res.* **46**, 6218–6228 (2018).
19. Barrell, B. G., Bankier, A. T. & Drouin, J. A different genetic code in human mitochondria. *Nature* **282**, 189–194 (1979).
20. Helm, M. et al. Search for characteristic structural features of mammalian mitochondrial tRNAs. *RNA* **6**, 1356–1379 (2000).
21. Pfeffer, S., Woellhaf, M. W., Herrmann, J. M. & Förster, F. Organization of the mitochondrial translation machinery studied in situ by cryoelectron tomography. *Nat. Commun.* **6**, 6019 (2015).
22. Ott, M., Amunts, A. & Brown, A. Organization and regulation of mitochondrial protein synthesis. *Annu. Rev. Biochem.* **85**, 77–101 (2016).
23. Salinas-Giegé, T. et al. Polycytidylation of mitochondrial mRNAs in *Chlamydomonas reinhardtii* Thalia Salinas-Gieg e. *Nucleic Acids Res.* **45**, 12963–12973 (2017).
24. Kitakawa, M. et al. Identification and characterization of the genes for mitochondrial ribosomal proteins of *Saccharomyces cerevisiae*. *Eur. J. Biochem.* **245**, 449–456 (1997).
25. Goldschmidt-Reisin, S. et al. Mammalian mitochondrial ribosomal proteins. N-terminal amino acid sequencing, characterization, and identification of corresponding gene sequences. *J. Biol. Chem.* **273**, 34828–34836 (1998).
26. Koc, E. C. et al. A proteomics approach to the identification of mammalian mitochondrial small subunit ribosomal proteins. *J. Biol. Chem.* **275**, 32585–32591 (2000).
27. Koc, E. C. et al. The large subunit of the mammalian mitochondrial ribosome: analysis of the complement of ribosomal proteins present. *J. Biol. Chem.* **276**, 43958–43969 (2001).
28. Bonen, L. & Calixte, S. Comparative analysis of bacterial-origin genes for plant mitochondrial ribosomal proteins. *Mol. Biol. Evol.* **23**, 701–712 (2006).
29. Amunts, A., Brown, A., Toots, J., Scheres, S. H. W. & Ramakrishnan, V. The structure of the human mitochondrial ribosome. *Science* **348**, 95–98 (2015).
30. Greber, B. J. et al. The complete structure of the 55S mammalian mitochondrial ribosome. *Science* **348**, 303–308 (2015).
31. Desai, N., Brown, A., Amunts, A. & Ramakrishnan, V. The structure of the yeast mitochondrial ribosome. *Science* **355**, 528–531 (2017).
32. Chrzanowska-Lightowlers, Z., Rorbach, J. & Minczuk, M. Human mitochondrial ribosomes can switch structural tRNAs – but when and why? *RNA Biol.* **14**, 1668–1671 (2017).
33. Brown, A. et al. Structure of the large ribosomal subunit from human mitochondria. *Science* **346**, 718–722 (2014).
34. Leaver, C. J. & Harmey, M. A. Higher-plant mitochondrial ribosomes contain a 5S mitochondrial ribonucleic acid component. *Biochem. J.* **157**, 275–277 (1976).
35. Gold, V. A., Chroscicki, P., Bragoszewski, P. & Chacinska, A. Visualization of cytosolic ribosomes on the surface of mitochondria by electron cryo-tomography. *EMBO Rep.* **18**, 1786–1800 (2017).

36. Lurin, C. et al. Genome-wide analysis of *Arabidopsis* pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell* **16**, 2089–2103 (2004).

37. Portereiko, M. F. Nuclear fusion defective 1 encodes the *Arabidopsis* RPL21M protein and is required for karyogamy during female gametophyte development and fertilization. *Plant Physiol.* **141**, 957–965 (2006).

38. Amunts, A. et al. Structure of the yeast mitochondrial large ribosomal subunit. *Science* **343**, 1485–1489 (2014).

39. Ramesh, M. & Woolford, J. L. Eukaryote-specific rRNA expansion segments function in ribosome biogenesis. *RNA* **22**, 1153–1162 (2016).

40. Bieri, P., Leibundgut, M., Saurer, M., Boehringer, D. & Ban, N. The complete structure of the chloroplast 70S ribosome in complex with translation factor pY. *EMBO J.* **36**, 475–486 (2017).

41. Boerema, A. P. et al. Structure of the chloroplast ribosome with chl-RRF and hibernation-promoting factor. *Nat. Plants* **4**, 212–217 (2018).

42. Pusnik, M., Small, I., Read, L. K., Fabbro, T. & Schneider, A. Pentatricopeptide repeat proteins in *Trypanosoma brucei* function in mitochondrial ribosomes. *Mol. Cell. Biol.* **27**, 6876–6888 (2007).

43. Aphasizheva, I., Maslov, D., Wang, X., Huang, L. & Aphasizhev, R. Pentatricopeptide repeat proteins stimulate mRNA adenylation/uridylation to activate mitochondrial translation in trypanosomes. *Mol. Cell* **42**, 106–117 (2011).

44. Aphasizheva, I. et al. Ribosome-associated pentatricopeptide repeat proteins function as translational activators in mitochondria of trypanosomes. *Mol. Microbiol.* **99**, 1043–1058 (2016).

45. Ramrath, D. et al. Evolutionary shift toward protein-based architecture in trypanosomal mitochondrial ribosomes. *Science* **7735**, 422 (2018).

46. Bieri, P., Greber, B. J. & Ban, N. High-resolution structures of mitochondrial ribosomes and their functional implications. *Curr. Opin. Struct. Biol.* **49**, 44–53 (2018).

47. Kühl, I., Dujeancourt, L., Gaisne, M., Herbert, C. J. & Bonnefoy, N. A genome wide study in fission yeast reveals nine PPR proteins that regulate mitochondrial gene expression. *Nucleic Acids Res.* **39**, 8029–8041 (2011).

48. Tavares-Carreón, F. et al. The pentatricopeptide repeats present in Pet309 are necessary for translation but not for stability of the mitochondrial Cox1 mRNA in yeast. *J. Biol. Chem.* **283**, 1472–1479 (2008).

49. Ruzzenente, B. et al. LRPPRC is necessary for polyadenylation and coordination of translation of mitochondrial mRNAs. *EMBO J.* **31**, 443–456 (2012).

50. Uyttewaal, M. et al. Characterization of *Raphanus sativus* pentatricopeptide repeat proteins encoded by the fertility restorer locus for Ogura cytoplasmic male sterility. *Plant Cell* **20**, 3331–3345 (2008).

51. Herbert, C. J. et al. Yeast PPR proteins, watchdogs of mitochondrial gene expression. *RNA Biol.* **10**, 1477–1494 (2013).

52. Wilson, D. N. & Doudna Cate, J. H. The structure and function of the eukaryotic ribosome. *Cold Spring Harb. Perspect. Biol.* **4**, a011536–a011536 (2012).

53. Jobe, A., Liu, Z., Gutierrez-Vargas, C. & Frank, J. New insights into ribosome structure andfunction. *Cold Spring Harb. Perspect. Biol.* **7**, a032615 (2018).

54. Hazle, T. & Bonen, L. Comparative analysis of sequences preceding protein-coding mitochondrial genes in flowering plants. *Mol. Biol. Evol.* **24**, 1101–1112 (2007).

55. Clough, S. J. & Bent, A. F. Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735–743 (1998).

56. Gregori, J., Villareal, L., Sanchez, A., Baselga, J. & Villanueva, J. An effect size filter improves the reproducibility in spectral counting-based comparative proteomics. *J. Proteom.* **16**, 55–65 (2013).

57. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

58. Lamattina, L., Gonzalez, D., Gualberto, J. & Grienenberger, J. M. Higher plant mitochondria encode an homologue of the nuclear-encoded 30-kDa subunit of bovine mitochondrial complex I. *Eur. J. Biochem.* **217**, 831–838 (1993).

59. Elthon, T. E., Nickels, R. L. & McIntosh, L. Monoclonal antibodies to the alternative oxidase of higher plant mitochondria. *Plant Physiol.* **89**, 1311–1317 (1989).

60. Pineau, B., Layoune, O., Danon, A. & De Paepe, R. l-Galactono-1,4-lactone dehydrogenase isrequired for the accumulation of plant respiratory complex I. *J. Biol. Chem.* **283**, 32500–32505 (2008).

61. Carrie, C. et al. Conserved and novel functions for *Arabidopsis thaliana* MIA40 in assembly of proteins in mitochondria and peroxisomes. *J. Biol. Chem.* **285**, 36138–36148 (2010).

62. Hooper, C. M., Castleden, I. R., Tanz, S. K., Aryamanesh, N. & Millar, A. H. SUBA4: the interactive data analysis centre for *Arabidopsis* subcellular protein locations. *Nucleic Acids Res.* **45**, D1064–D1074 (2017).

63. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

64. Karpenahalli, M. R., Lupas, A. N. & Söding, J. TPRpred: a tool for prediction of TPR-, PPR- and SEL1-like repeats from protein sequences. *BMC Bioinformatics* **8**, 2 (2007).

65. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845 (2015).

66. Obayashi, T., Aoki, Y., Tadaka, S., Kagaya, Y. & Kinoshita, K. ATTED-II in 2018: a plant coexpression database based on investigation of the statistical property of the mutual rank index. *Special Issue – Databases* **59**, 1–7 (2018).

67. de la Rosa-Trevín, J. M. et al. Scipion: a software framework toward integration, reproducibility and validation in 3D electron microscopy. *J. Struct. Biol.* **195**, 93–99 (2016).

68. de la Rosa-Trevín, J. M. et al. Xmipp 3.0: an improved software suite for image processing in electron microscopy. *J. Struct. Biol.* **184**, 321–328 (2013).

69. Rohou, A. & Grigorieff, N. CTFFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).

70. Abrishami, V. et al. A pattern matching approach to the automatic selection of particles from low-contrast electron micrographs. *Bioinformatics* **29**, 2460–2468 (2013).

71. Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).

72. Ban, N. et al. A new system for naming ribosomal proteins. *Curr. Opin. Struct. Biol.* **24**, 165–169 (2014).

73. Stupar, R. M. et al. Complex mtDNA constitutes an approximate 620-kb insertion on *Arabidopsis thaliana* chromosome 2: Implication of potential sequencing errors caused by large-unit repeats. *Proc. Natl Acad. Sci. USA* **98**, 5099–5103 (2001).

74. Adams, K. L., Daley, D. O., Whelan, J. & Palmer, J. D. Genes for two mitochondrial ribosomal proteins in flowering plants are derived from their chloroplast or cytosolic counterparts. *Plant Cell* **14**, 931–943 (2002).

75. Noeske, J. et al. High-resolution structure of the *Escherichia coli* ribosome. *Nat. Struct. Mol. Biol.* **22**, 336–341 (2015).

Author contributions
P.G., F.W., H.M. and Y.H. designed and coordinated the experiments. F.W., T.T.N., M.A., A.B., J.C., M.Q., P.H. and L.K. performed the experiments and analysed the results. P.G., F.W., H.M., T.T.N. and Y.H. wrote the manuscript.

Competing interests
The authors declare no competing interest.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41477-018-0339-y.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to H.M. or Y.H. or P.G.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Small is big in *Arabidopsis* mitochondrial ribosome

Florent Waltz [1], Tan-Trung Nguyen[2], Mathilde Arrivé[1], Anthony Bochler[3], Johana Chicher[4], Philippe Hammann[4], Lauriane Kuhn[4], Martine Quadrado[2], Hakim Mireau [2]*, Yaser Hashem [3]* and Philippe Giegé [1]*

[1]Institut de biologie de moléculaire des plantes UPR2357 du CNRS, Université de Strasbourg, Strasbourg, France. [2]Institut Jean-Pierre Bourgin INRA, AgroParisTech, CNRS, Université Paris-Saclay, Versailles, France. [3]Institut Européen de Chimie et Biologie U1212 Inserm, Université de Bordeaux, Pessac, France. [4]Plateforme protéomique Strasbourg Esplanade FRC1589 du CNRS, Université de Strasbourg, Strasbourg, France. *e-mail: hakim.mireau@inra.fr; yaser.hashem@ubordeaux.fr; giege@unistra.fr
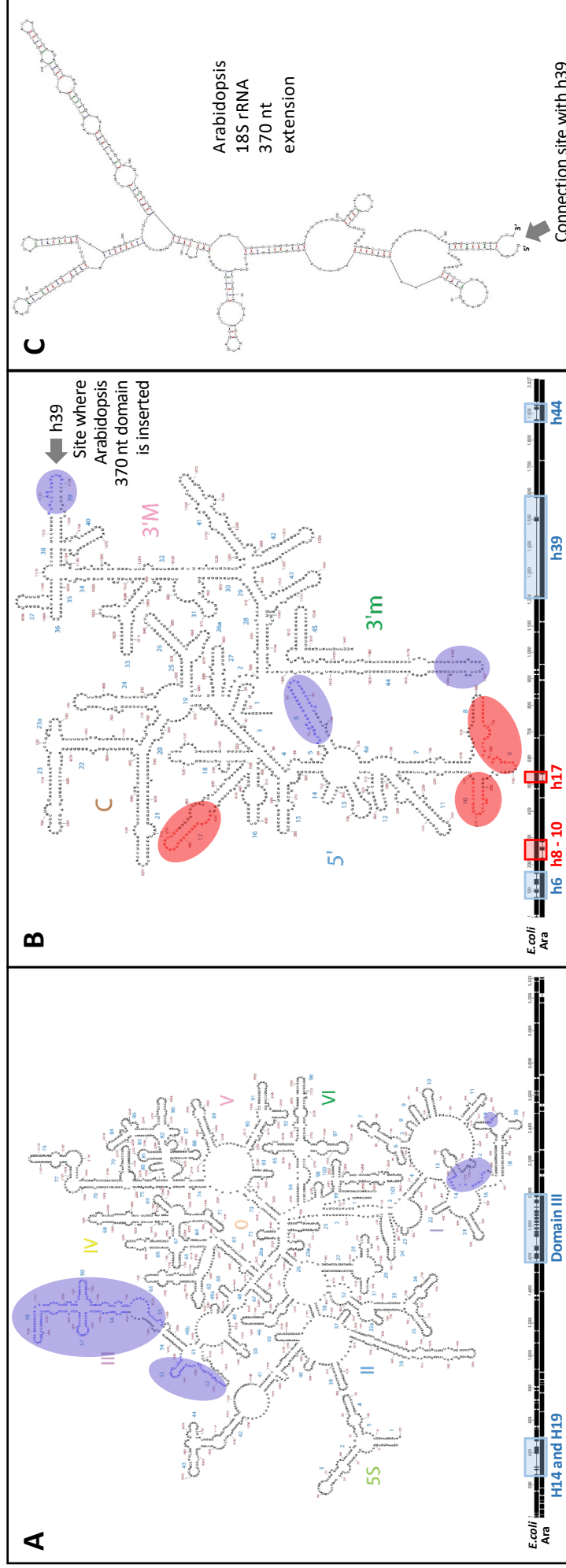
Supplementary Figure 1 Arabidopsis mitochondrial rRNAs specific features. Specific features of Arabidopsis mitochondrial rRNAs based on the secondary structure diagram of the *Escherichia coli* rRNA and sequence alignment of *E.coli* and Arabidopsis rRNAs. Expansion segment positions of Arabidopsis rRNAs are indicated in blue and deletions are indicated in red on the secondary structure of the 23S **(A)** and 16S **(B)** *E.coli* rRNAs. **(A)** In the large subunit rRNA no deletion is observed in Arabidopsis compared to *E.coli* but 127 nt and 35 nt extension are respectively observed in H14 and H19 of domain I. H52 and H53 are two times larger in Arabidopsis and modifications in H54 to 59 makes domain III significantly larger than *E.coli* rRNA. **(B)** Arabidopsis SSU rRNA is significantly larger than the *E.coli* one mainly because of a 370 nt expansion inserted in h39 of 3' major domain (grey arrow). Smaller extensions are also present in h6 and h44. Some part of the rRNA have also been reduced compared to *E.coli* in h8 to 10 and h17. **(C)** Secondary structure prediction of the Arabidopsis 18S rRNA h39 expansion domain determined with mFOLD. dG = -116.88 [Initially -133.60]. Secondary structure templates were obtained from the RiboVision suite (http://apollo.chemistry.gatech.edu/RiboVision).
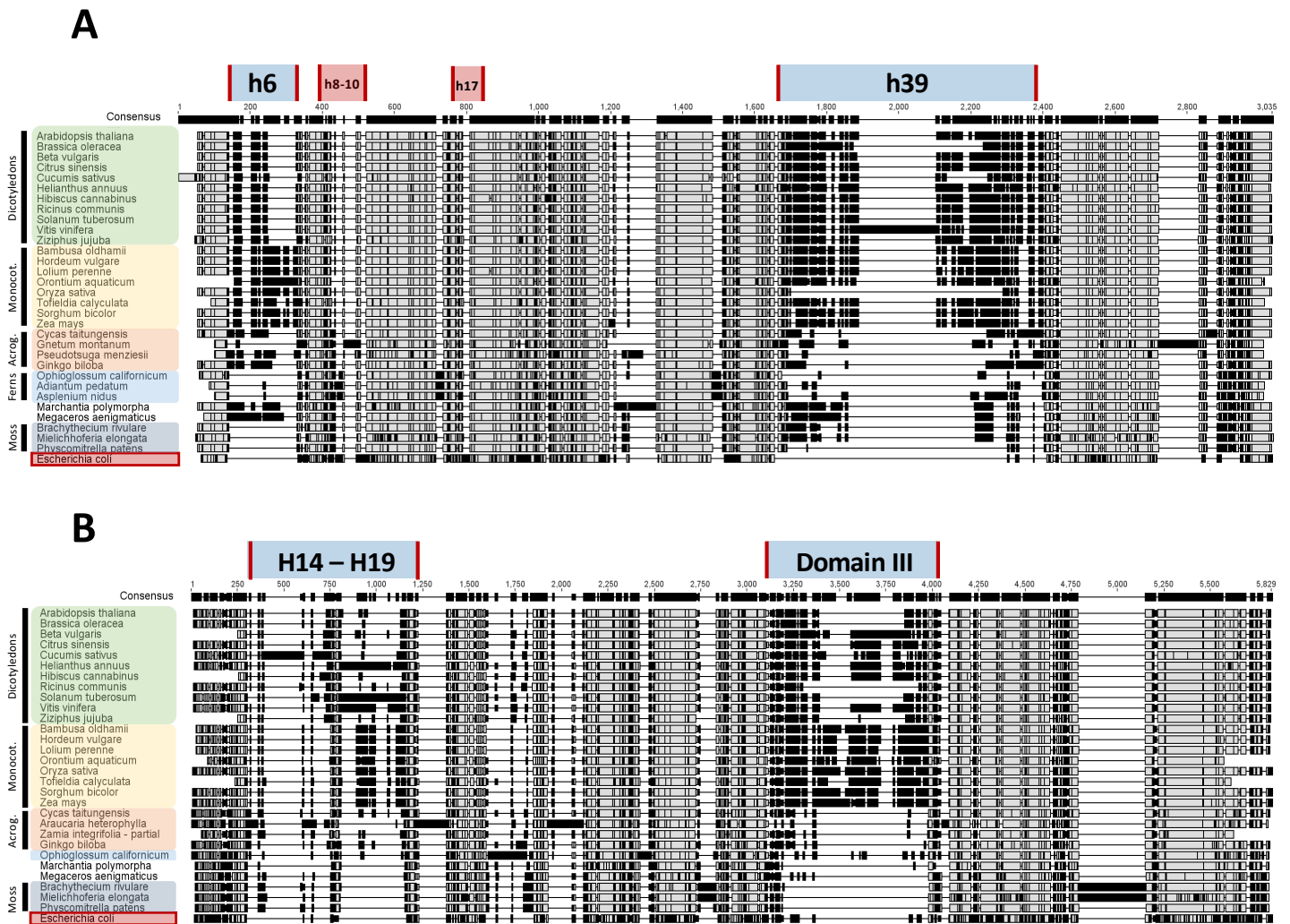
Supplementary Figure 2 Conservation of mitochondrial large and small subunit rRNAs features in representative Viridiplantae species. **(A)** Sequence comparison of Arabidopsis mitochondrial 18S rRNA and *E.coli* 16S rRNA revealed the presence of expansion segments in h6 and h39. A phylogenic analysis revealed that both expansions are conserved in angiosperms, h6 being up to two times larger in monocotyledons than in dicotyledons. The conservation of these expansions is almost lost in more basal plants and variation is observed in angiosperms as well. For instance, the h39 expansion is 200 nt larger in *Vitis vinifera* and much shorter in *Oryza sativa*. Deletions in h8 to 10 and h17 seems to be conserved all across Viridiplantae. **(B)** Sequence comparison of Arabidopsis mitochondrial 26S rRNA and *E.coli* 23S rRNA revealed the presence of expansion segments in domain I and domain III. Again these modifications are mostly conserved in angiosperms, the extensions being significantly larger in monocotyledons than in dicotyledons. Sequences in agreement with the consensus (>85% threshold) are indicated in grey and the ones not in agreement in black.
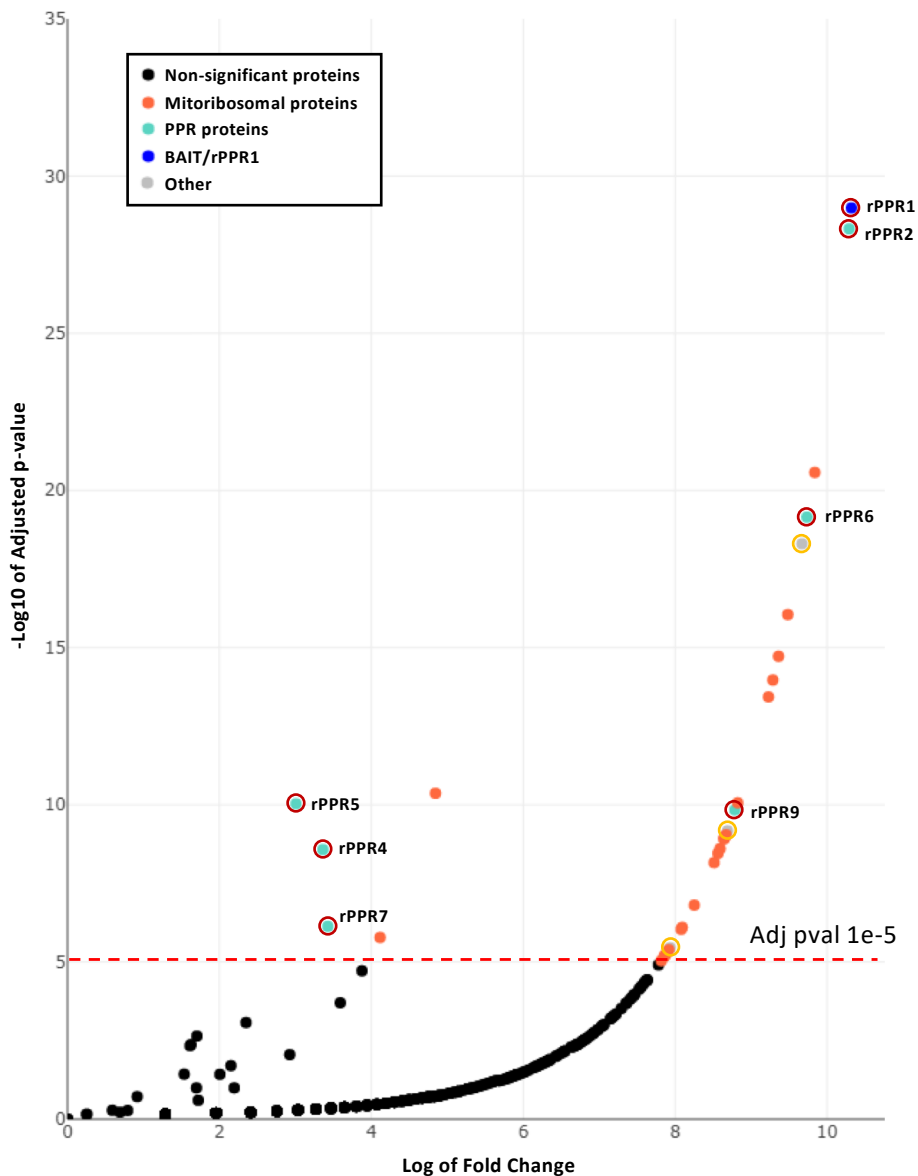
Supplementary Figure 3 rPPR1 preferentially co-immunoprecipitates the small subunit of the mitoribosome. The proteins statistically over-represented in rPPR1 immuno-purifications (IP) performed in dissociative conditions (800mM KCl and 8mM Mg^{2+}) as compared to control experiments are visualized as a volcano plot. The over representation of proteins in rPPR1 IPs is represented on the x-axis as the Log Fold Change and the statistical confidence in their enrichment is shown by decreasing p-values on the y-axis, (-Log10 of Adjusted p values). The red dotted line indicates the significance threshold of 5 ($-log10(1e^{-5})$). 33 proteins have an adjusted p-values "adjp" below $1e^{-5}$ and thus are considered as *bona fide* IP partners of rPPR1 under dissociative conditions. The bait rPPR1 is shown in blue, non-significant proteins in black, mitoribosomal proteins in orange, PPRs in turquoise and the rest in grey. Among the 33 proteins, 23 are mitoribosomal proteins, 17 being SSU proteins, 7 are PPR proteins, all being rPPRs, and 3 are additional ribosomal proteins. Plant specific additional mitoribosomal proteins (3) are circled in yellow and rPPR proteins (7) are circled in red.

A

I   II   III

254nm OD

1 2 3 4

SSU
LSU
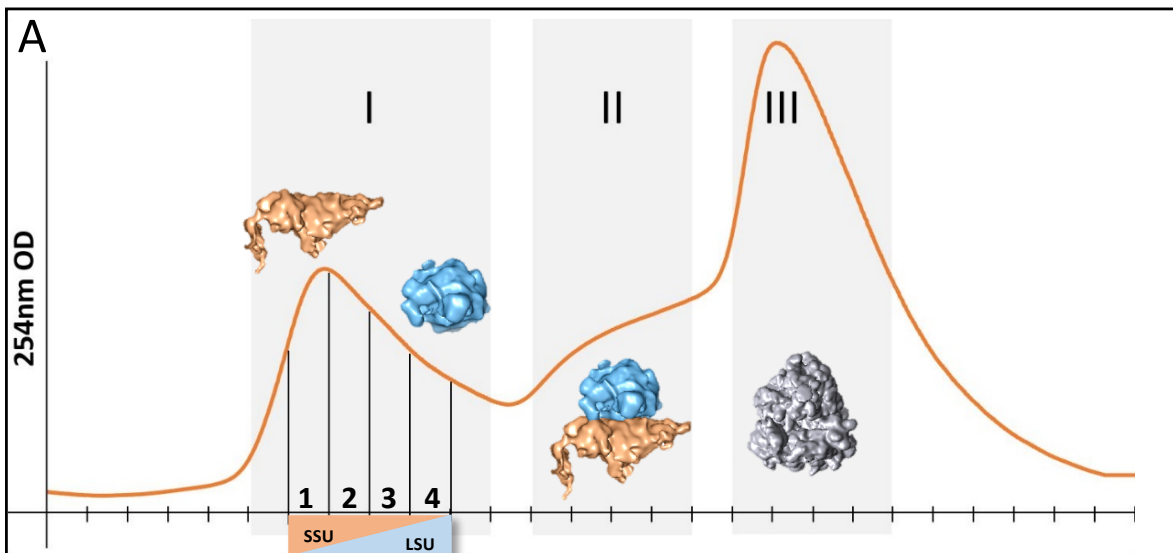
B — Canonical mitochondrial r-proteins

|	AGI	1	2	3	4	
uS2m	AT3G03600	45	40	39	34	SSU
uS3m	ATMG00090	47	39	41	21	
uS4m	ATMG00290	24	14	17	12	
uS5m	AT1G64880	88	79	79	54	
bS6m	AT3G18760	18	19	18	12	
uS7m	ATMG01270	22	14	14	9	
uS8m	AT2G19720	6	6	5	3	
uS8m	AT4G29430	11	11	11	10	
uS9m	AT3G49080	65	49	49	32	
uS10m	AT3G22300	26	26	16	22	
uS11m	AT1G31817	62	53	55	39	
uS12m	ATMG00980	9	4	5	6	
uS13m	AT1G77750	24	21	17	15	
uS14m	AT2G34520	15	15	12	12	
uS15m	AT1G15810	52	39	36	29	
uS15m	AT1G80620	47	34	28	21	
uS17m	AT1G49400	25	22	17	19	
uS19m	AT5G47320	27	16	15	11	
bS16m	AT5G56940	24	23	15	21	
bS18m	AT1G07210	41	34	25	23	
bS21m	AT3G26360	12	9	6	5	
bTHXm	AT2G21290	5	2	0	0	
mS23	AT1G26750	59	46	39	32	
mS29	AT1G16870	62	50	56	30	
mS33	AT5G44710	17	15	16	9	
mS34	AT5G52370	15	9	7	7	
mS35	AT3G18240	73	63	59	38	
mS35	AT4G21460	39	33	33	19	
uL1m	AT2G42710	37	53	63	57	LSU
uL2m	AT2G44065	0	0	0	8	
uL3m	AT3G17465	8	14	20	18	
uL4m	AT2G20060	14	20	23	28	
uL5m	ATMG00210	0	0	3	5	
bL9m	AT5G53070	11	19	23	22	
uL11m	AT4G35490	3	5	9	12	
bL12m	AT3G06040	0	0	0	1	
uL13m	AT3G01790	10	18	29	37	
uL14m	AT5G46160	4	14	25	15	
uL15m	AT5G64670	8	14	19	25	
bL17m	AT5G09770	6	16	20	22	
bL17m	AT5G64650	0	9	10	13	
uL18m	AT5G27820	0	0	0	3	
bL19m	AT1G24240	5	8	14	11	
bL20m	AT1G16740	4	6	9	10	
bL21m	AT4G30930	10	20	30	23	
uL22m	AT1G52370	0	9	16	16	
uL22m	AT4G28360	4	8	15	13	
uL23m	AT4G39880	8	13	20	25	
uL24m	AT5G23535	13	20	22	24	
bL25m	AT4G23620	0	0	6	7	
bL25m	AT5G66860	23	23	43	32	
bL28m	AT4G31460	0	0	7	9	
uL29m	AT1G07830	11	15	13	26	
uL30m	AT5G55140	13	25	20	29	
bL31m	AT1G27435	8	12	14	14	
bL31m	AT5G55125	0	0	2	7	
bL33m	AT5G18790	0	2	3	3	
mL41	AT5G39800	7	11	12	14	
mL41	AT5G40080	6	11	11	13	
mL43	AT3G59650	9	13	16	22	
mL46	AT1G14620	0	0	6	15	
mL54	AT3G01740	0	3	6	7	

C — Arabidopsis specific mitochondrial r-proteins

	AGI	1	2	3	4	
rPPR2	AT1G19520	122	101	88	76	SSU
rPPR3a	AT1G55890	22	15	16	9	
rPPR8	AT5G15980	79	56	61	42	
rPPR1	AT1G61870	57	37	34	21	
rPPR6	AT3G02650	84	68	64	38	
rPPR9	AT5G60960	29	51	60	68	LSU
rPPR7	AT4G36680	33	49	60	59	
rPPR5	AT2G37230	49	64	86	72	
rPPR4	AT1G60770	32	47	57	51	
rPPR3b	AT3G13160	17	23	28	17	Unclear

	AGI	1	2	3	4	
	AT1G18630	25	22	19	14	SSU
	AT1G47278	10	6	3	3	
	AT1G53645	116	110	99	83	
mS22	AT1G64600	23	18	11	9	
	AT3G21465	10	8	8	0	
	AT5G26800	15	11	12	8	
	AT4G15640	50	34	44	26	
	AT4G22000	49	1	2	13	SSU/unclear
	AT1G73940	0	3	4	5	LSU
	AT3G51010	0	0	1	3	
mL40	AT4G05400	0	8	18	29	
mS47	AT4G31810	14	31	32	19	Unclear
	AT5G49210	0	10	15	4	Unclear

Supplementary Figure 4 Biochemical purification of Arabidopsis mitochondrial ribosomes on high resolution 10-30% linear sucrose gradients. **(A)** Schematic representation of mitochondrial ribosome purification chromatogram. The x-axis represents individual fractions and the y-axis the absorbance at 254 nm. Three main zones, indicated as peak I to III were analyzed, I being the lightest and III the heaviest. **(B)** Peak I was further separated in 4 fractions (1 to 4). The abundance of the respective proteins is represented by the number of spectra of all r-proteins of the small and the large subunit of the mitochondrial ribosome. Data are reported as absolute spectra values colored such that higher spectra values are red and lower are green. R-proteins of the LSU are more abundant in fraction 4 than in fraction 1 and r-proteins of the SSU are more abundant in fraction 1 than in fraction 4. This general tendency is respected for all r-proteins. **(C)** Additional mitoribosomal proteins such as rPPRs were attributed to the SSU or LSU according to distributions as described in (B). For three proteins, rPPR3b, At4g31810 and At5g49210 assignations are unclear.
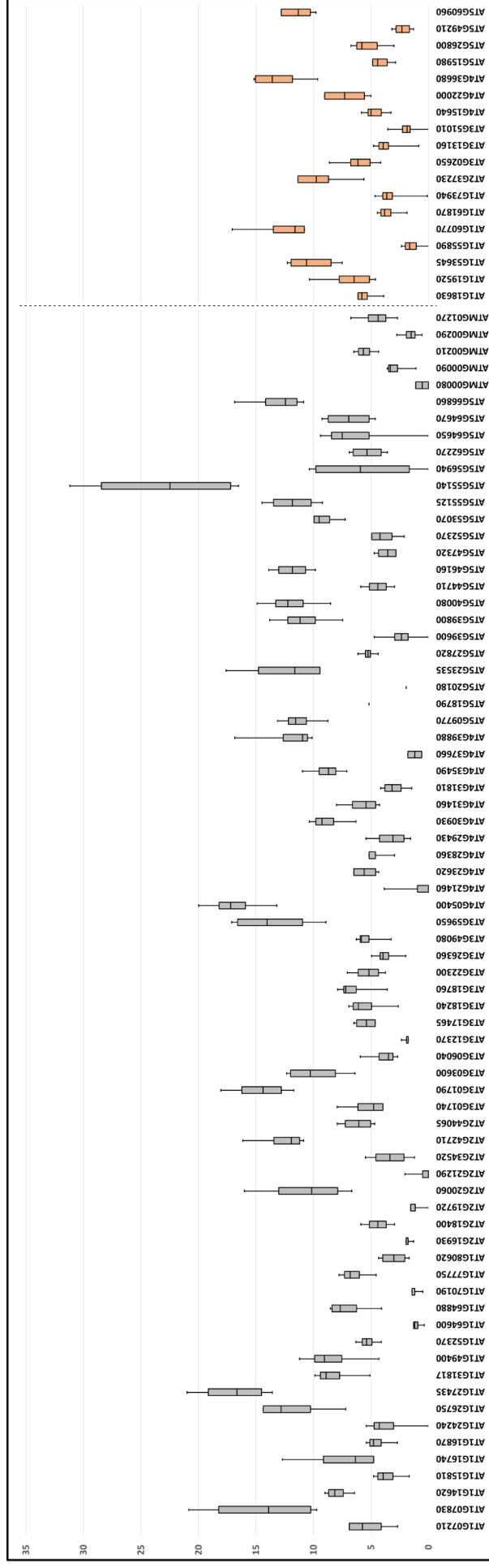
Supplementary Figure 5 Comparison of mitochondrial r-proteins abundance with plant specific mitochondrial r-proteins. To compare the relative abundance of the newly identified plant specific r-proteins with the canonical r-proteins, the "corrected abundance" of each proteins was determined for the four fractions (hence n=4) of peak II containing full mitoribosomes, and are represented as boxplots (with the band inside the box representing the median and the error bars representing the highest and lowest values). The corrected abundance, on the y-axis, indicates the average number of spectra obtained for each protein divided by the length of the protein in amino acids multiplied by 100. This normalizes spectra numbers with protein lengths, long proteins generating more peptides than short ones (average normalized numbers are indicated in Supplementary Table 6). Proteins represented in grey are canonical r-proteins as well as mitochondria specific r-proteins, while proteins represented in pink are the novel plant-specific mitochondrial r-proteins. The average value for the corrected abundance of canonical and mitochondria specific r-proteins is 6.26, while the average value for plant-specific proteins is 6.72. Interestingly, proteins such as L12 (At3g06040, At1g70190 and At4g37660) present in multiple copies are found at lower levels (with values of 1.2, 3.8 and 1.1 on average).
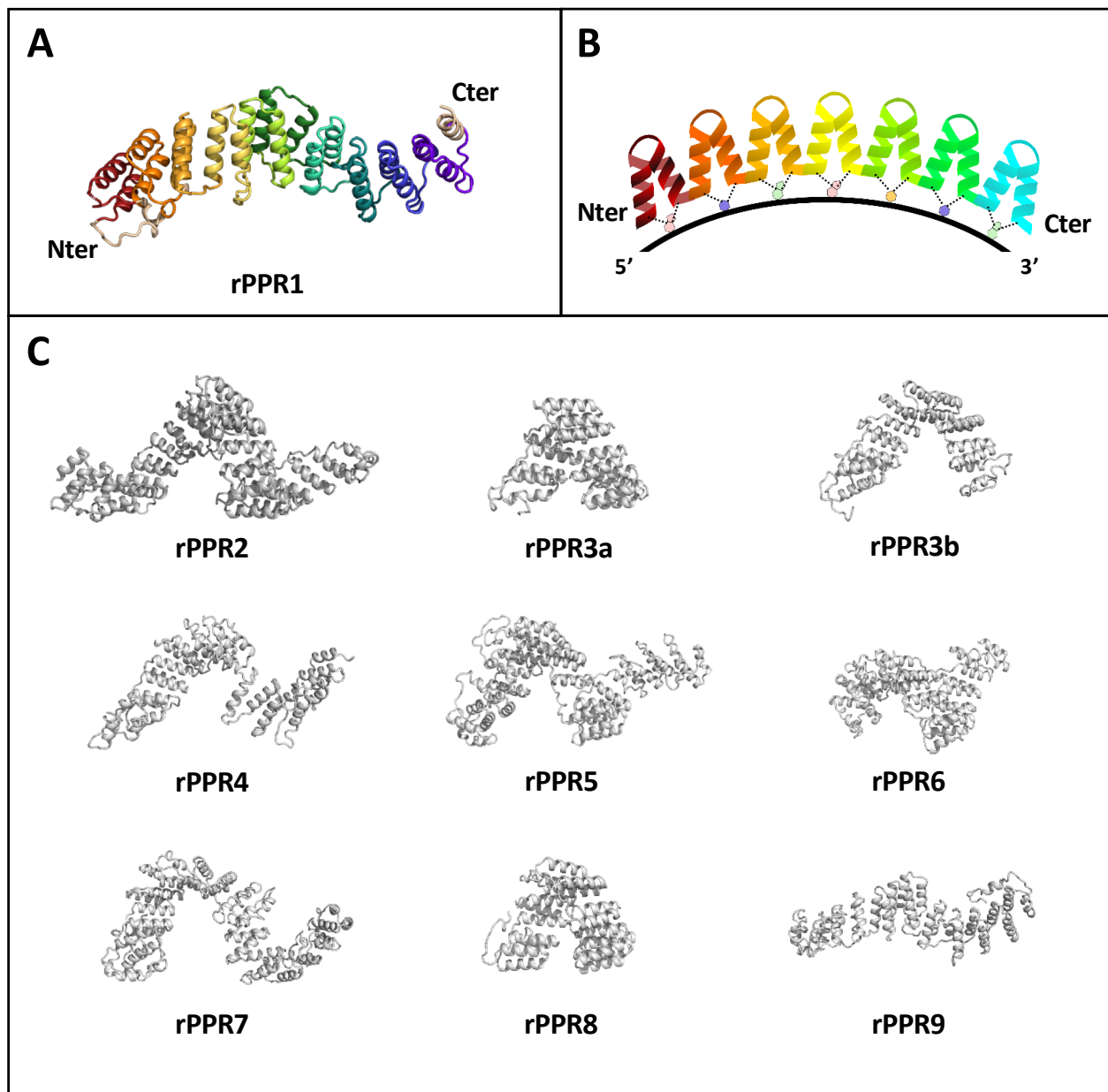
Supplementary Figure 6 Tridimensional structure prediction of the 10 rPPR proteins identified in Arabidopsis mitoribosome. **(A)** rPPR1 structure prediction. Each of the 10 PPR motifs are depicted in a different colors. **(B)** Schematic representation of a PPR structure and mechanism of transcript recognition. Each PPR repeat, 35 amino acids long, folds into a helix-turn-helix structure. Each repeat is represented in a different color. The succession of PPR motifs forms a superhelix. Each PPR motif specifically recognizes one nucleotide according to a recognition code connecting amino acids 5 and 35 of each repeat and the recognized nucleotide[10]. **(C)** Structure predictions of all the rPPR identified. All models were build with Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2)[65] using protein sequences without their mitochondrial targeting sequences and molecular representations were prepared with PyMol) and oriented from Nter to Cter from left to right.
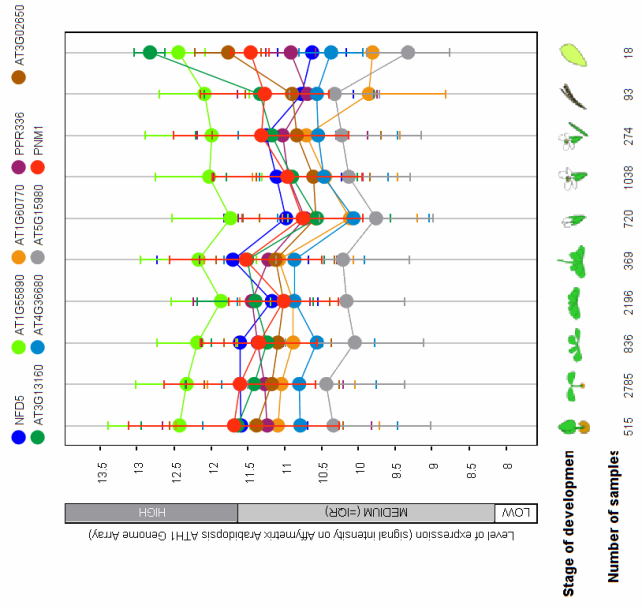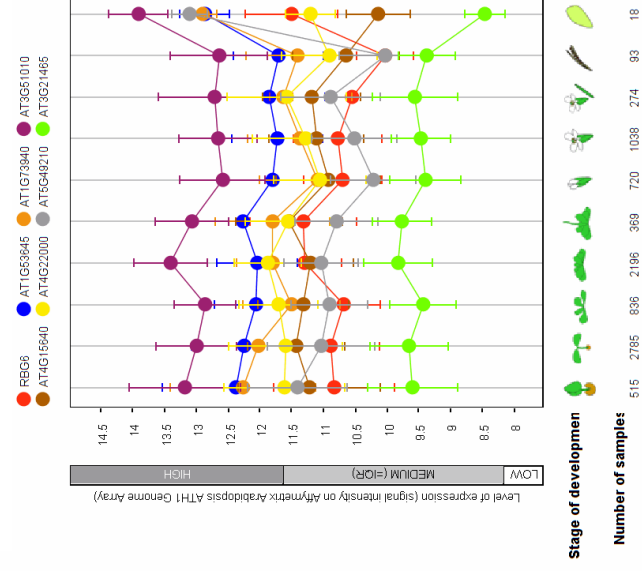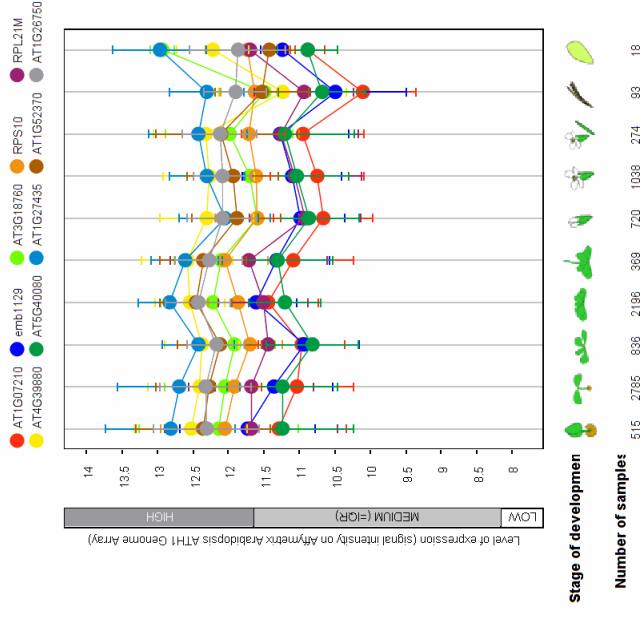
Supplementary Figure 7 Expression patterns during plant development of plant specific r-proteins compared to canonical mitochondrial r-proteins. Genevestigator Arabidopsis expression level analysis in different developmental stages, for mitochondrial r-proteins and the additional proteins. Dots represent mean expression values and error bars standard deviations across n>3 replicate experiments. All rPPR genes are represented except rPPR5 and all except At5g26800 and At1G47278 are represented for non-PPR additional proteins because these three gene are absent from the dataset used. Ten random mitoribosomal proteins were chosen as reference. Mitochondrial r-proteins genes all follow the same pattern of expression during the different developmental stages. Levels of expression range from 10 to 13, which is considered as high-medium to high levels. This range of expression is similar for both rPPRs and non-PPR additional protein genes. At3g21465 is expressed at the lowest level compared to canonical r-proteins. Data set: Platform: AT_AFFY_ATH1: Affymetrix Arabidopsis ATH1 Genome Array.
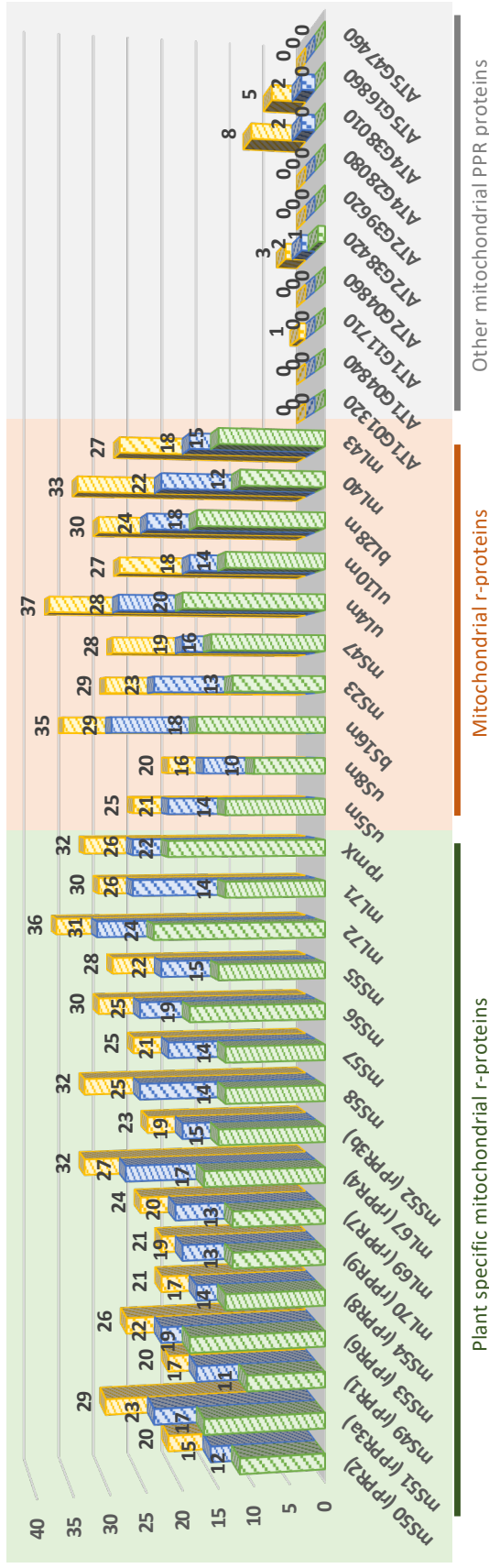
Supplementary Figure 8 Novel plant specific r-proteins and canonical r-proteins share similar co-expression patterns. Histograms represent the numbers of mitochondrial canonical r-proteins genes co-expressed with each plant specific r-protein gene among the top 100 co-expressed genes (green), the top 200 co-expressed genes (blue) and the top 300 co-expressed genes (yellow). For mL102, mS83 (At3g21465), mS86 and mS87 no co-expression data is available. These co-expression patterns are very similar to those of 10 exemplary chosen non-plant specific mitoribosomal proteins genes (on the right). As a control, 10 PPR proteins also found in immunoprecipitation experiments but not significantly enriched were investigated. These non-ribosomal PPRs are not co-expressed with mitoribosome proteins. Data were acquired from the publicly available data on ATTED-II using the Ath-m Platform ID[66].

A

B
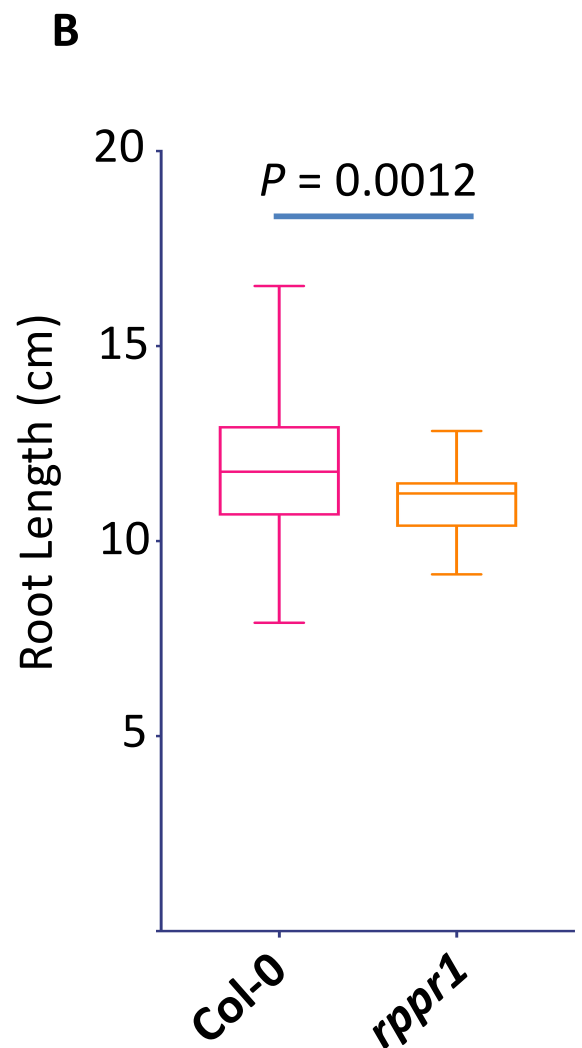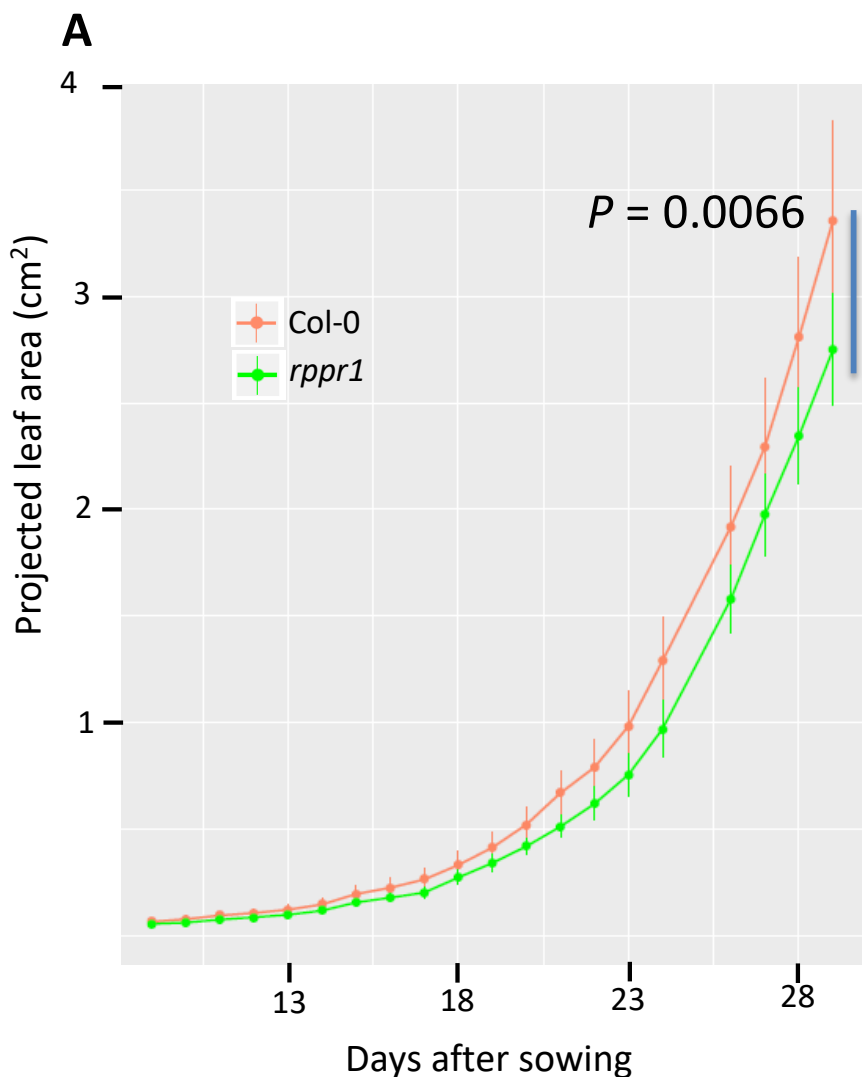
Supplementary Figure 9 *rppr1* mutants have a global retarded growth rate compared to wild type plants. **(A)** Projected rosette areas of wild-type (brown) and *rppr1* plants (green) measured over a period of 21 days in short day conditions using the Phenoscope phenotyping device. Mean values (n= 10) and standard deviations are shown for each time point. **(B)** Box plot showing average root lengths in wild type and *rppr1* plants. Measures (Col-0 n=49, *rppr1* n=50) were made on vertical Petri dishes after 14 days of growth. Minima, maxima, 25th, 50th and 75th percentiles are indicated. Statistical analyses were performed with Prism 7 software, using a Student's t-test (unpaired, two-tailed).
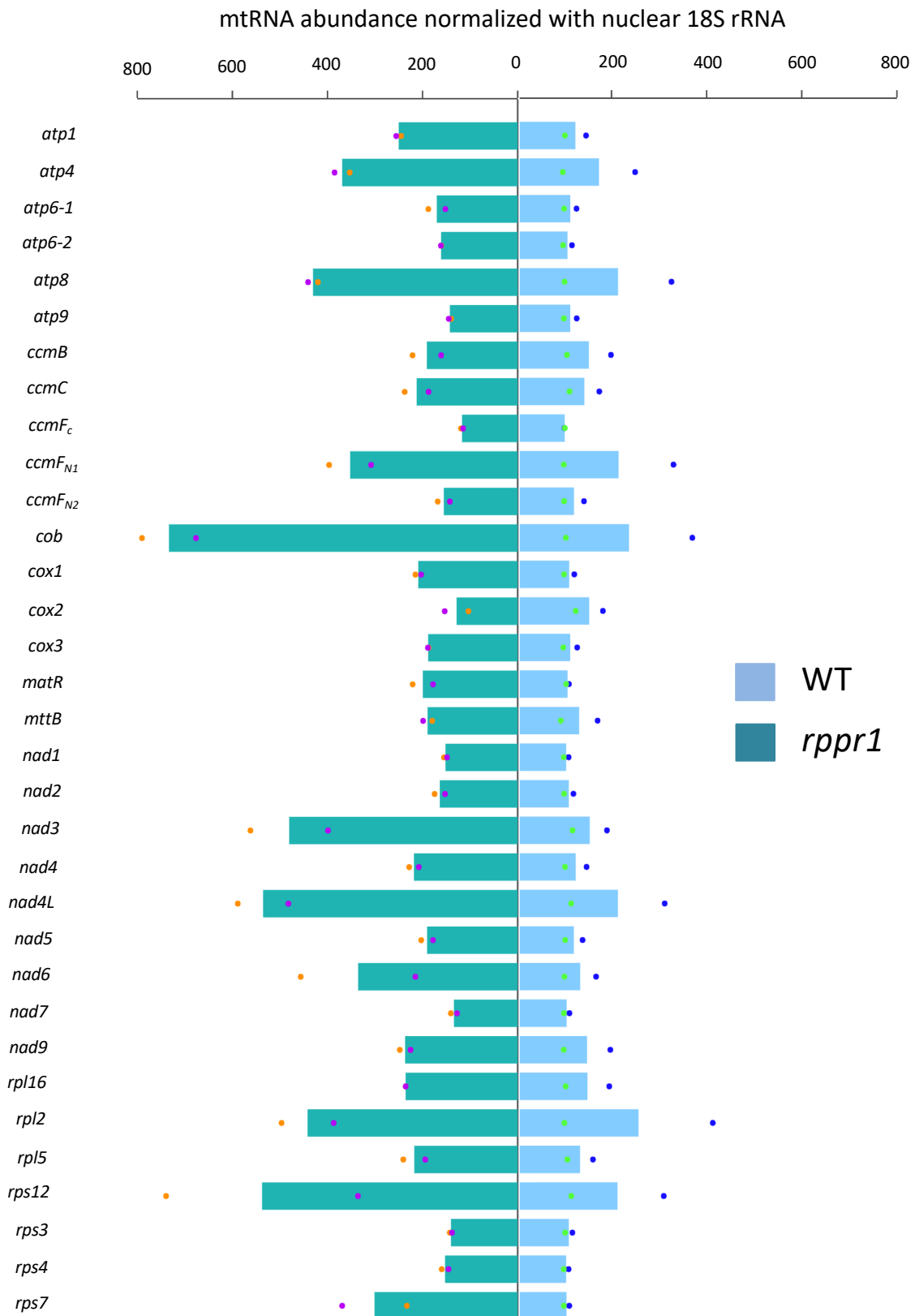
Supplementary Figure 10 Steady state levels of all mitochondrial encoded mRNAs measured by RT-quantitative PCR in wild type (blue bars) and *rppr1* mutants (green bars). Mean values are presented as histograms with dots indicating values for two biological replicates. Mitochondrial mRNA abundances were normalized with Arabidopsis nuclear 18S rRNA.
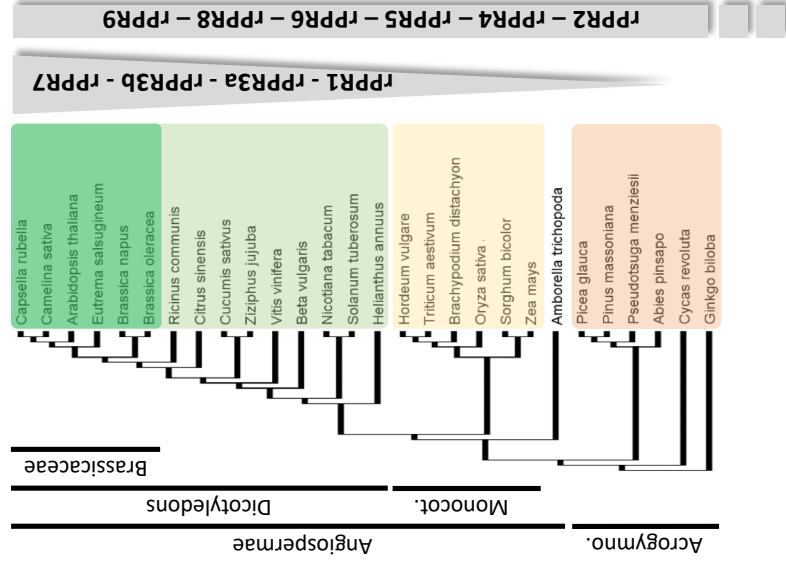
Supplementary Figure 11 rPPR proteins conservation across Viridiplanteae. To investigate rPPR conservation in Viridiplanteae, tBLASTn using NCBI public data for all rPPR proteins were performed. 28 different species were investigated using the Arabidopsis sequence as a starting point. rPPR1, rPPR3a, rPPR3b and rPPR7, all being part of the PPR336 family, seem to have originated in Acrogymnospermae as these proteins share a same orthologues in this group. rPPR3a and b are the result of a gene duplication specific to Brassicaceae, only one copy is found in the rest of Angiospermae. rPPR2, rPPR4, rPPR5, rPPR6, rPPR8 and rPPR9 are all conserved in these groups, suggesting that they arose in more basal groups (e.g Bryophyta, Polypodiopsida, Marchantiophyta). The most conserved are rPPR1, rPPR2 and rPPR5 with respectively a Pairwise % Positive (BLSM62) of 62%, 61.9% and 65% across the investigated species. The least conserved is rPPR6 with only 47%.

	Small subunit (EMDB-4408)	Large subunit (EMDB-4409)
Data collection and processing		
Magnification	59,000 X	59,000 X
Voltage (kV)	300	300
Electron exposure (e–/Å^2)	60	60
Defocus range (µm)	-0.6 to -4.5	-0.6 to -4.5
Pixel size (Å)	2.256	2.256
Initial particle images (no.)	~318,000	~318,000
Final particle images (no.)	~19,000 (~6%)	~24,000 (~7.5%)
Map resolution (Å)	~21	~16
FSC threshold		

Supplementary Table 8 Cryo-EM data collection, refinement and validation statistics

Mitochondrial r-proteins mutants

In the manuscript by Waltz et al. we describe the mutants obtained for all the rPPR proteins, except for rPPR6 for which no mutant was identified. However, during my PhD work, I also started to characterize mutants for the other plant specific r-proteins that are not PPR proteins.

For **mS84,** which is described as a IF2-like protein, no homozygous mutant could be obtained for its gene, suggesting an essential function. The two **mS83** proteins are annotated as adenylyl cyclase proteins. Two genes encode mS83 proteins in Arabidopsis. For At4g15640, homozygous mutants could be obtained that do not show any recognizable macroscopic phenotype. Similarly, for At3g21465 homozygous mutants could be obtained and no macroscopic phenotype was observed. This suggests that the two gene products are interchangeable as part of the mitoribosome and that a double mutant should be obtained for future functional studies. **mS85** is annotated as a glycine-rich RBP6 protein. Similarly to mS84, no homozygous mutant could be obtained for its gene, suggesting an essential function.

I was also able to obtain mutants for mS47 and mS22 which are mitochondrial r-proteins shared with the other mitoribosomes. For **mS47** no homozygous mutant was obtained, confirming the lethal phenotype previously observed for another mutant allele of this gene (Gipson et al., 2017). Finally, for **mS22**, no homozygous mutant was obtained suggesting here as well an essential function.

The determination of the precise molecular functions of all these novel proteins using mutants obtained or characterized during my PhD work will be the subject of the laboratory future work (see discussion below). For the function of essential genes, knock down mutants will be generated and analyzed at the molecular level. This will further contribute to reveal specificities of the plant mitochondria translation apparatus.

Participation to other projects of the laboratory

PRORP proteins in evolution

In the laboratory, the functional study of PRORP (PROtein Only RNase P) enzymes has been a major research focus of the past years. These proteins catalyze RNase P activity, which is a crucial step of pre-tRNAs maturation. It consists in the endonucleolytic cleavage of the additional 5' sequence (or 5' leader) of tRNA primary transcripts. The RNase P activity is found almost in all branches of life and is essential in both prokaryotes and eukaryotes, where it can be either carried out by a ribonucleoprotein complex (the ancestral form of RNase P) or by PRORP proteins. PRORP proteins belong to the family of PPR proteins, and are therefore specific to eukaryotes. In 2008 protein-only RNase P was first identified in human mitochondria and was shown to function as a heterotrimer complex containing a PRORP orthologue, with the three proteins required to perform RNase P activity (Holzmann et al., 2008). In 2010 my laboratory identified in Arabidopsis three different PRORP proteins, PRORP1 performs RNase P activity in both mitochondria and chloroplast while PRORP2 and 3 both perform RNase P activity in the nucleus (Gobert et al., 2010). Contrary to human mitochondria, all these PRORP proteins were shown to function alone, thus showing that PRORP is the catalytic enzyme of proteinaceous RNase P. This also showed for the first time that an organism (Arabidopsis) can function without the ancestral ribonucleoprotein form of RNase P (Gutmann et al., 2012b) and that ribonucleoprotein RNase P was not universally conserved, contrary to previous assumptions (Altman, 2007). Finally, mechanistic studies of PRORP enzymes suggested that PRORP and RNA based RNase P share similar modes of substrate recognition (Pinker et al., 2013, 2017) thus making a very interesting case of convergent evolution.

Still, important questions remained, i.e. on the reason why ribonucleoprotein RNase P was retained in some organisms and/or compartments and why it was replaced by PRORP in other cases. To address these questions, it is required to have a much deeper knowledge of RNase P functional and mechanistic diversity in the different phyla of life (Lechner et al., 2015). For this my laboratory started to study PRORP enzymes in a variety of model organisms representing the diversity of eukaryotes. A first study on the model green algae *Chlamydomonas reinhardtii* showed that a single PRORP has taken over RNase P activities in all mitochondria, chloroplast and the nucleus of the algae (Bonnard et al., 2016). During my PhD thesis, I have also contributed to this work by starting to look at PRORP enzymes in the Apicomplexa *Plasmodium falciparum* and the nematode *Romanomermis culicivorax*.
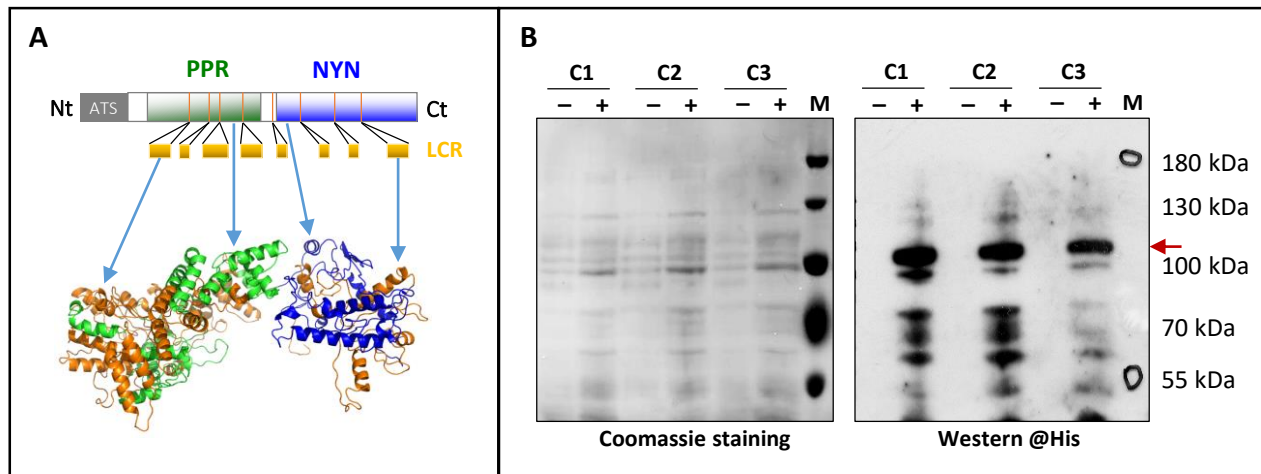
Figure 27: PfPRORP

A Schematic representation of the *Plasmodium falciparum* PRORP protein (PfPRORP) main domains as well as a prediction of its 3D structure obtained with Phyre2. This protein contains an apicoplast targeting sequence (ATS) as well as the domains defining PRORP proteins: a PPR domain and a NYN catalytic domain. In addition, PfPRORP is characterized by the occurrence of a number of low complexity regions (LCR) typical of the genus Plasmodium.

B Induction assays performed overnight at 17°C identified the best condition for the expression of PfPRORP. The protein is expressed only when induced (**+**) but several degradation products are observed. The red arrow indicate the signals likely corresponding to PfPRORP that has a calculated molecular weight of 110 kDa.

Apicoplastic *Plasmodium falciparum* PRORP (PfPRORP)

During my Master's internship and the first months of my PhD, I was mainly involved in the study of Plasmodium PRORP. Among eukaryotes, no PRORP protein of the SAR (Stramenopiles, Alveolata, and Rhizaria) supergroup of eukaryotes has been characterized to date. Therefore I was involved in the characterization of a putative PRORP protein, predicted to perform the RNase P activity in the apicoplast of *Plasmodium falciparum*, the parasite responsible for malaria (Kalanon and McFadden, 2010). The PRORP-like protein identified in Plasmodium has features characteristic of Plasmodium proteins such as the presence of numerous low complexity regions (LCR) (Frugier et al., 2010). Moreover the *P. falciparum* putative PRORP shares very low sequence identity with other eukaryote PRORP (e.g. only 15% sequence identity in amino acids with Arabidopsis PRORP1) and is much bigger, characterized by a molecular weight of 138 kDa (compared to 70 kDa for PRORP1) (Fig 27).

In brief, I first cloned the PfPRORP gene to produce the protein in bacteria. The original cDNA had a low GC content (21.2%) and seemingly had a codon-usage unsuitable for prokaryote expression, therefore cloning and protein expression were highly hampered. Another version of the cDNA was produced with an optimized codon usage for protein expression in prokaryote systems. Using this version of the gene I was able to clone it and expressed it in bacteria. Unfortunately, I did not managed to obtain samples suitable for activity assays. At this stage the protein could not be obtained in sufficient quantity, purity and solubility for further biochemical characterizations. Future work will involve other expression systems e.g. in eukaryote cells. The characterization of PfPRORP would also be very interesting because of its high divergence as compared to human PRORP. It would thus make a suitable target for new antimalarial strategies.
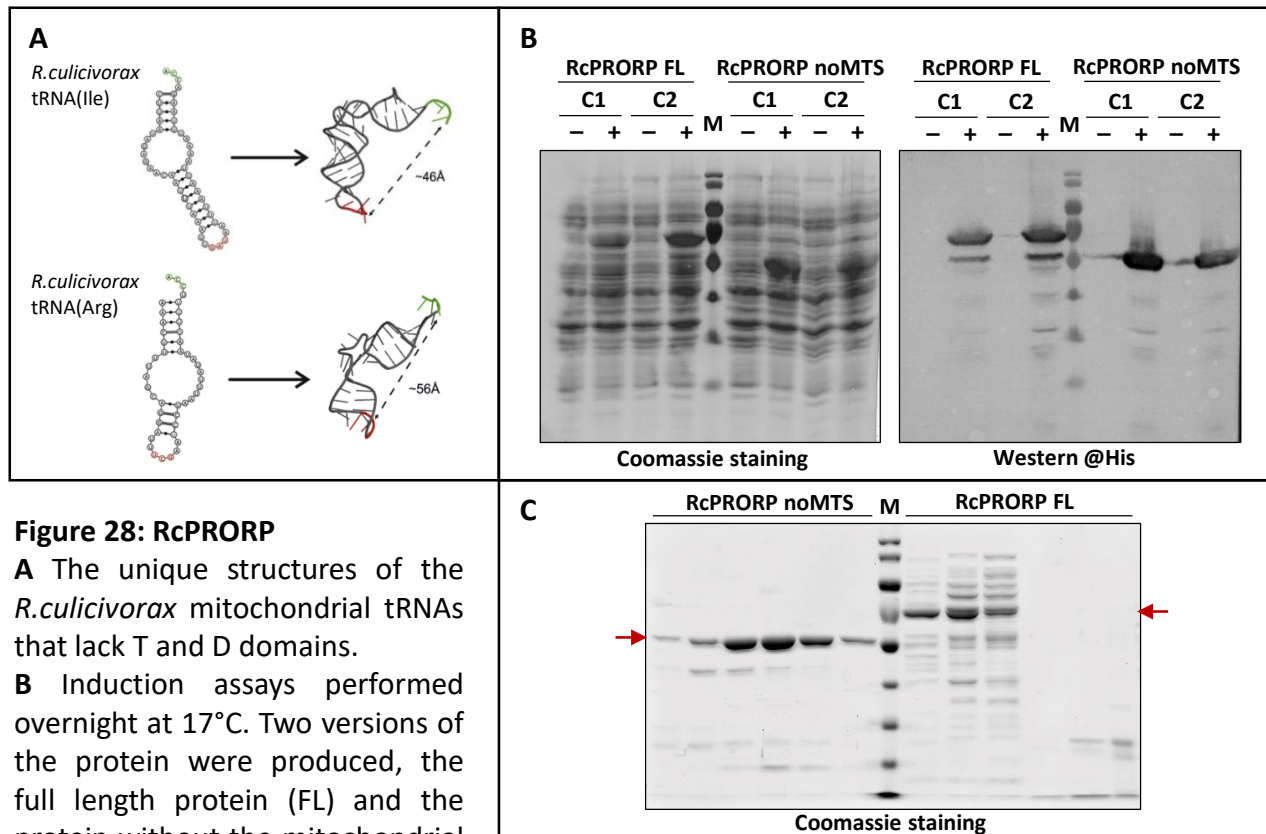
Figure 28: RcPRORP

A The unique structures of the *R.culicivorax* mitochondrial tRNAs that lack T and D domains.

B Induction assays performed overnight at 17°C. Two versions of the protein were produced, the full length protein (FL) and the protein without the mitochondrial targeting sequence (noMTS).

Both proteins are well expressed (**+**) with limited degradation in the different clone analyzed (**CX**). **C** Both proteins were purified and the resulting elution fractions were further purified by gel-filtration. The different elution fractions containing the highest levels of RcPRORP are presented here. The red arrows indicate bands corresponding to the proteins of interest calculated molecular weights.

Mitochondrial *Romanomermis culicivorax* PRORP (RcPRORP)

Additionally, I was also involved in the study of the PRORP protein of the worn *Romanomermis culicivorax*. In animal mitochondria, many of the mitochondria-encoded tRNAs have non-canonical structures, e.g missing the D or T loop, but in *R.culicivorax* all mitochondrial tRNAs are non-canonical (Fig 28 A) (Jühling et al., 2012). Moreover, this organisms was shown to possess the smallest tRNAs described to date (Jühling et al., 2018; Wende et al., 2014). Therefore we were wondering how the *R.culicivorax* PRORP could deal with such bizarre tRNAs, especially because our work showed that PRORP PPR motifs bind the D and T loops of tRNAs (Pinker et al., 2017). Hence, similarly to PfPRORP, I cloned, expressed and purified RcPRORP and tested its activity on tRNAs. Two different versions of RcPRORP were produced, the full-length protein and the protein devoid of mitochondrial targeting sequence. Contrary to PfPRORP, RcPRORP is similar to canonical PRORPs (e.g AtPRORP1 or HsPRORP) therefore the protein expression was much easier (Fig 28 B). The purification process was efficient, but as seen on Fig 28 C, the noMTS version was obtained in more quantities and the samples were much purer. Still, the recombinant protein was unable to process pre-tRNAs *in vitro* (data not shown). Two reasons could explain this result i) we performed the tests with canonical tRNAs, hence the RcPRORP might not be able to perform its activity on canonical tRNAs or ii) it is more likely that RcPRORP functions as a complex, similar to the heterotrimeric complex found in human mitochondria, therefore it might not be able to perform its activity without its partners. The study of this enzyme, in collaboration with Prof. Mario Mörl (University of Leipzig) will thus require to look for potential partners *in vivo*. This should reveal if yet unidentified proteins were specifically recruited to bind the non-canonical structures of Romanomermis mitochondrial tRNAs.
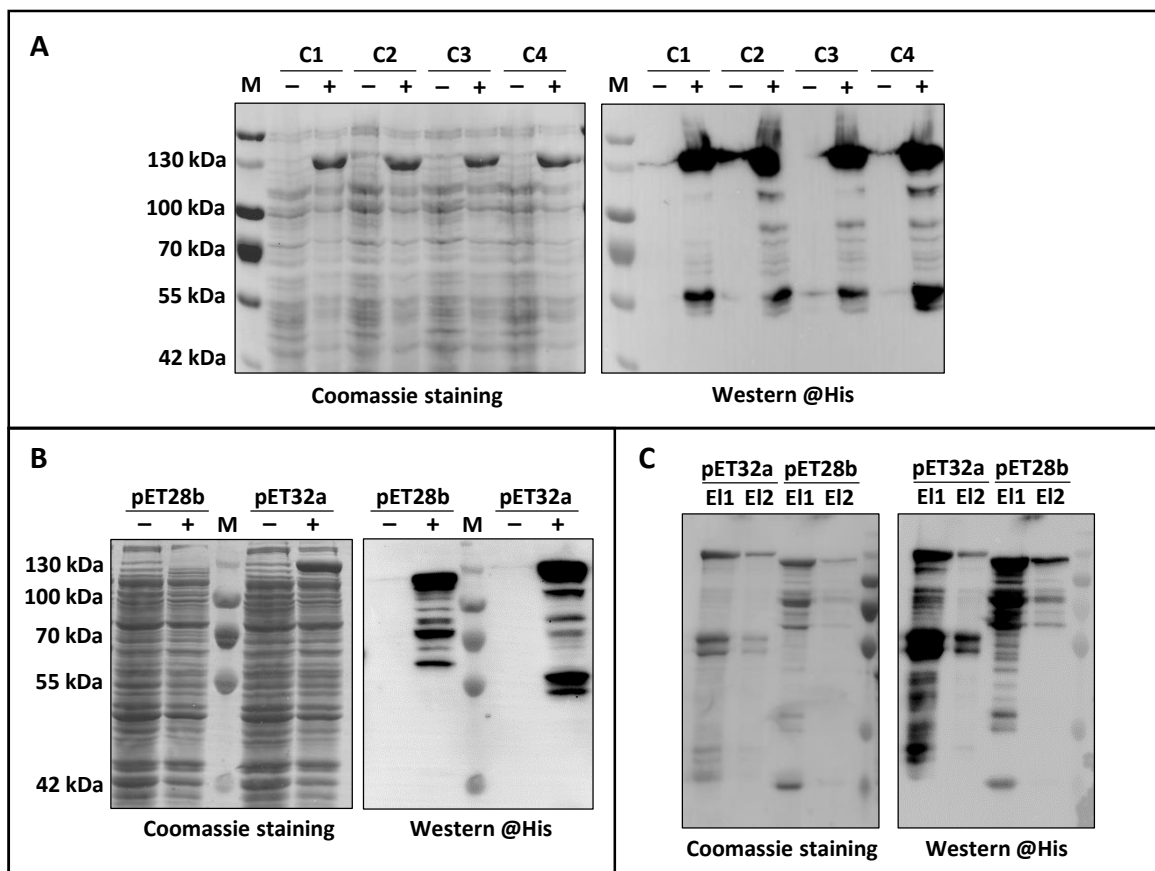
Figure 29: MNU2 expression and purification
A Induction assays performed overnight at 17°C with the pET32a construct. Some degradation products are observed, especially above 55kDa. Recombinant MNU2 is detected in induced samples (+) at an apparent molecular weight of 130 kDa.
B Induction assays performed as in **A** but with the pET28b construct as well. The protein is expressed at lower levels as compared with the pET32a construct.
C Final elutions, after concentration, of protein purification performed in denaturing conditions. Both constructs were tested. **El1** corresponds to the first elution step, **El2** corresponds to the second elution step.

Characterization of a novel mitochondrial nuclease: MNU2

In the framework of the study of PRORP proteins, the laboratory has initiated a project to identify Arabidopsis PRORP proteins interacting partners. Indeed, even though the three different AtPRORPs are able to catalyze RNase P activity alone *in vitro*, they might associate with additional proteins to form tRNA maturation complexes, or other functional complexes, *in vivo*. In this context, we decided to investigate PRORP1 protein interaction network in Arabidopsis mitochondria.

For this, as a first step, a blue native PAGE analysis revealed that PRORP1 indeed occurs in at least one protein complex *in vivo*. In order to identify interaction partners, a co-IP approach using plants expressing an HA tagged version of PRORP1 was implemented. It revealed that PRORP1 co-purifies with different proteins involved in mitochondrial gene expression processes. In particular, PRORP1 was able to immuno-precipitate MNU2, another NYN domain mitochondrial nuclease. MNU2 (as well as MNU1) was previously described as a protein involved in 5' maturation of mRNAs in plant mitochondria (Stoll and Binder, 2016). As a next step, direct interaction between PRORP1 and identified proteins was monitored by a yeast-two-hybrid-like genetic system based on the "split ubiquitin" approach. Direct interaction was only observed between PRORP1 and MNU2. Other candidates identified by co-IP did not show direct interaction with PRORP1. This direct interaction was mapped down to the PRORP1 PPPY motif, particularly conserved in Streptophyta, which is a protein / protein interaction motif used for direct interaction with MNU2 through its WW domain. Altogether, these results, presented in the following manuscript, reveal the existence of an RNA 5' maturation complex in Arabidopsis mitochondria and suggest that PRORP proteins cooperate with other gene expression regulators for RNA maturation *in vivo*.

At this stage, the exact molecular function of the PRORP1/MNU2 complex is unknown. Furthermore, the nature of MNU2 activity as an endonuclease or an exonuclease as well as its target specificity are not known. During this project I thus tried to express and purify recombinant forms of MNU2 in order to test its activity *in vitro*. A plasmid with Δ74-MNU2 in the pET32a vector was obtained from Dr. Stefan Binder. This construct allows to express the protein without its MTS and in fusion with a thioredoxin-His tag at its N-terminal end. Additionally, I cloned Δ74-MNU2 in pET28b to express it in fusion with a 6xHis tag at its C-terminal end.

Induction assays were first performed using the pET32a:MNU2 construct. As shown in Fig 29 A, the protein was efficiently expressed. Inductions tests were also performed in parallel to pET32a:MNU2 with the newly obtained pET28b:MNU2, which is the expression vector routinely

used in the laboratory. As seen in the Fig 29 B, the protein was also expressed but not to the level obtained with the pET32a:MNU2 construct. As a next step, I tried to purify the protein using the two different constructs. A pilot experiment indicated that the protein was not efficiently purified in native conditions, therefore the purification was also performed under denaturant conditions. Fig 29 C presents the final purified fractions after concentration of the samples, for the purification performed in denaturing conditions. As seen on the figure, only a small quantity of the protein was retrieved and the fractions were heavily contaminated with other proteins, most likely corresponding to MNU2 degradation products. The results obtained in native condition are not shown, as no protein could be retrieved, most likely because the affinity tag was masked in the protein structure and the protein did not bind the column. Hence at this stage the protein could not be obtained to perform activity assays. Moreover, even if the purification of the protein was achieved in denaturant conditions, it would be mandatory to confirm its structural integrity. Bacterial expression systems might not be optimal for the expression of MNU2, hence other expression systems will be tested in the future e.g. in eukaryote cells. The expression the catalytic domain only could be tested as well to characterize the activity of the enzyme as an endo- or exoribonuclease, which is one of the main objectives of the purification of MNU2.

Further functional studies using MNU2 mutants as well as PRORP1 down-regulation mutants will reveal to which degree MNU2 cooperates with PRORP1 for tRNA and / or mRNA maturation *in vivo*.

Determination of protein-only RNase P interactome in Arabidopsis mitochondria identifies a complex between PRORP1 and another NYN domain nuclease

Ayoub Bouchoucha[1] and Florent Waltz[1], Géraldine Bonnard[1], Mathilde Arrivé[1+], Philippe Hammann[2+], Lauriane Kuhn[2+], Cédric Schelcher[1+], Hélène Zuber[1+], Anthony Gobert[1*] and Philippe Giegé[1*]

[1]UPR 2357 du CNRS, IBMP, Université de Strasbourg, 12 rue du général Zimmer, F-67084 Strasbourg, France

[2]Plateforme protéomique Strasbourg Esplanade, FRC1589 du CNRS, Université de Strasbourg, IBMC, 15 rue René Descartes, Strasbourg, F-67084, France.

+by alphabetical order

*For correspondence (email: anthony.gobert@ibmp-cnrs.unistra.fr or giege@unistra.fr)

In revision in Plant journal

Determination of protein-only RNase P interactome in Arabidopsis mitochondria identifies a complex between PRORP1 and another NYN domain nuclease

Ayoub Bouchoucha[1] and Florent Waltz[1], Géraldine Bonnard[1], Mathilde Arrivé[1+], Philippe Hammann[2+], Lauriane Kuhn[2+], Cédric Schelcher[1+], Hélène Zuber[1+], Anthony Gobert[1*] and Philippe Giegé[1*]

[1]UPR 2357 du CNRS, IBMP, Université de Strasbourg, 12 rue du général Zimmer, F-67084 Strasbourg, France

[2]Plateforme protéomique Strasbourg Esplanade, FRC1589 du CNRS, Université de Strasbourg, IBMC, 15 rue René Descartes, Strasbourg, F-67084, France.

[+]by alphabetical order

*For correspondence (email: anthony.gobert@ibmp-cnrs.unistra.fr or giege@unistra.fr)

SUMMARY

The essential endonuclease activity that removes 5' leader sequences from transfer RNA precursors is called RNase P. While ribonucleoprotein RNase P enzymes containing a ribozyme are found in all domains of life, another type of RNase P called "PRORP", for "PROtein-only RNase P", only composed of protein occurs in a wide variety of eukaryotes, in organelles and the nucleus. Here, in order to find how PRORP functions integrate with other cell processes, we exemplarily explore the protein interaction network of PRORP1 in Arabidopsis mitochondria. Although PRORP proteins function as single subunit enzymes *in vitro*, we find that PRORP1 occurs in a number of protein complexes and is present in polysome fractions. The analysis of immuno-precipitated protein complexes identifies proteins involved in mitochondrial gene expression processes. In particular, direct interaction is established between PRORP1 and MNU2 a mitochondrial nuclease found here to be involved in tRNA biogenesis. A specific domain of MNU2 and a conserved signature of PRORP1 are found to be directly accountable for this protein interaction. Altogether, results reveal the existence of a novel RNA maturation complex in Arabidopsis mitochondria and suggest that PRORP proteins cooperate with other gene expression regulators for RNA maturation *in vivo*.

Keywords: RNase P, RNA maturation, mitochondrial nucleases, pentatricopeptide repeats

INTRODUCTION

Similar to all other RNA molecules, transfer RNAs (tRNAs) undergo many maturation processes in order to become functional. One of the crucial steps of tRNA biogenesis is performed by an activity termed RNase P that removes 5' leader sequences from tRNA precursors. This simple endonuclease function found in all domains of life, is essential to obtain usable tRNAs and thus critical for translation (Altman, 2007). RNase P activity has attracted considerable attention since over thirty years, in particular because the first characterized RNase P enzyme, in *Escherichia coli*, was found to be a ribonucleoprotein (RNP) particle containing a ribozyme (Guerrier-Takada *et al.*, 1983). Similar RNP RNases P were found in other Bacteria, in Archaea and in Eukarya, in particular in human nuclei and in both yeast nucleus and mitochondria (Hartmann and Hartmann, 2003). Despite major differences in the RNA subunit structure and in the RNP protein content (Hernandez-Cid *et al.*, 2012; Jarrous, 2017), all these RNPs are characterized by the incidence of a conserved catalytic RNA. This led to the assumption that RNase P would universally occur as RNPs and that RNase P represents one of the rare conserved vestige of a prebiotic RNA world.

More recently, this view was contradicted and the interest in RNase P renewed with the identification of a second type of RNase P, only composed of protein in eukaryotes (Holzmann *et al.*, 2008; Gobert *et al.*, 2010; Pinker *et al.*, 2013). This other type of RNase P called PRORP for 'PROtein-only RNase P" is found in four out of five eukaryote super-groups, in organelles and / or the nucleus (Lechner *et al.*, 2015). While some eukaryotes such as human have retained PRORP specifically for mitochondrial RNase P activity (Holzmann *et al.*, 2008) and an RNP is present in the nucleus, other eukaryotes use PRORP in both organelles and the nucleus and RNP RNase P is entirely absent. For instance, in Arabidopsis and Trypanosoma, multiple PRORPs perform specialized RNase P activities in organelles and the nucleus (Gutmann *et al.*, 2012; Täschner *et al.*, 2012), while in Chlamydomonas, a single triple localized PRORP is responsible for RNase P activity in the nucleus, mitochondria and chloroplasts (Bonnard *et al.*, 2016). PRORP proteins belong to the large family of pentatricopeptide repeat (PPR) proteins, a eukaryote specific family of RNA binders involved in numerous gene expression processes (Giegé, 2013; Hammani *et al.*, 2014). PRORP proteins are Λ shaped proteins, with an N-terminal PPR domain believed to confer substrate specificity to the enzyme, making one arm of the Λ. The other arm is made by a C-terminal catalytic domain belonging to the NYN (N4BP1, YacP-like Nuclease) family (Anantharaman and Aravind, 2006). The apex of the Λ is a structural zinc-binding domain that connects the two main domains and appears to confer flexibility to the enzyme (Schelcher *et al.*, 2016; Pinker *et al.*, 2017). Other PRORP features are specific to certain phyla. For instance, in Streptophyta, a G rich insertion as well as a PPPY motif are highly conserved (Lechner *et al.*, 2015). The comparison of PRORP and RNP RNase P structures and modes of action by several research groups has revealed that the two types of enzymes use a fundamentally similar catalytic mechanism and seem to share a similar RNA binding strategy (Pinker *et al.*, 2017; Gobert *et al.*, 2013; Klemm *et al.*, 2017; Walczyk *et al.*, 2016; Mao *et al.*, 2016; Chen *et al.*, 2016), which is a remarkable case of convergent evolution. Still, while a rapidly growing number of studies investigate PRORP mode of action, reviewed by Schelcher et al. (Schelcher *et al.*, 2016), little is known on their functional

diversity, i.e. on their transcriptome wide substrate spectra and on the integration of PRORP functions with other cellular processes.

In this context, we exemplarily investigate PRORP1 protein interaction network in Arabidopsis mitochondria. We find that PRORP1 occurs in different protein complexes. In particular, we find that PRORP1 PPPY motif is a protein / protein interaction motif used for direct interaction with the mitochondrial nuclease MNU2 involved in mitochondrial RNA 5' maturation (Stoll and Binder, 2016). We also show that MNU2 is involved in tRNA accumulation in Arabidopsis mitochondria. This suggests that the two nucleases could cooperate *in vivo* for tRNA 5' processing.

RESULTS

PRORP1 occurs in protein complexes in Arabidopsis mitochondria

Eukaryotic RNA binding proteins act almost exclusively in complexes with few to hundreds of interacting partners (Smirnov *et al.*, 2017; Dreyfuss *et al.*, 2002). Some of these complexes are well described like the spliceosome or the RISC complex, but in most cases complexes and protein interaction networks are not characterized. In the case of PPR proteins including PRORP, interacting partners are largely unknown, although some PPR proteins were found in polysome fractions (Uyttewaal *et al.*, 2008; Hammani *et al.*, 2011), PNM1 was also found to interact with two nuclear proteins, NAP1 and the transcription factor TCP8 (Hammani *et al.*, 2011) and different PPR proteins interact with each other or with MORF proteins as part of the RNA editing machinery in plant organelles (Boussardon *et al.*, 2012; Takenaka *et al.*, 2012; Hartel *et al.*, 2013). Very little is known on PRORP protein interactors. In human, PRORP requires TRMT10C and SDR5C1 (formerly known as MRPP1 and MRPP2) for RNase P activity (Holzmann *et al.*, 2008), possibly to help binding the non-canonical fold of human mitochondrial tRNAs, as discussed by Salinas-Giegé *et al.*, 2015, and human PRORP is part of mitochondrial RNA granules (Jourdain *et al.*, 2013; Antonicka *et al.*, 2013). Still, the integration of PRORP activity with other cell functions remains largely unexplored. As a pilot study, we searched for functional complexes involving PRORP1 in Arabidopsis mitochondria. *Ab initio*, it could be expected that PRORP might occur in putative tRNA maturation complexes involving e.g. RNase Z, CCA nucleotidyl transferase or tRNA modification enzymes.

Thus, in order to purify and characterize complexes containing PRORP1, an Arabidopsis *prorp1* knock-out line was complemented with a construct expressing a C-terminal fusion of PRORP1 with an HA affinity tag placed under the control of *PRORP1* endogenous promoter. Mitochondria were extracted from the *PRORP1-HA* line and the occurrence of PRORP1 in complexes was first analysed by Blue Native PAGE. Mitochondrial complexes were separated according to size on a first dimensional Blue Native gel and subunits of the respective complexes were resolved in the second dimension by SDS PAGE. Immuno-detection analysis with HA tag specific antibodies revealed fours signals corresponding to PRORP1-HA (Figure 1). The major signal, that represents 60% of the overall PRORP1-HA signal, most likely corresponds to a monomeric form of PRORP1. However, the other signals reveal complexes of *circa* 150, 300 and 500 kDa (Figure 1B). Only one signal corresponding to

the 500 kDa complex is detected on the first dimension, probably because the HA epitope was masked in the other native complexes. Altogether, this showed that *in vivo*, PRORP1 occurs in different mitochondrial complexes.

Characterization of PRORP1 mitochondrial protein interaction network

Then, in order to characterize the protein complexes involving PRORP1, mitochondria purified from the *PRORP1-HA* line were lysed and complexes solubilized from pure mitochondria were purified by affinity to the HA-tag. Immuno-detection analysis established that PRORP1-HA could indeed be pulled down by this approach (Figure S1). Eight independent PRORP1-HA immuno-affinity experiments were performed as well as six control experiments where HA immuno-affinity purifications were performed with wild type col-0 Arabidopsis plants. Proteins from both PRORP1-HA and control samples were trypsin digested and identified by quantitative proteomics using nano LC-ESI-MS/MS. A statistical analysis using the msmsTests R package was performed to identify proteins that are significantly over-represented in PRORP1 immunoprecipitations as compared to control experiments and thus likely represent PRORP1 interacting partners. Among the over 1000 proteins identified in the proteomic analysis (Table S1), 67 proteins with adjusted p-values under 0.05 were considered *bona fide* PRORP1 interaction partners (Figure 2). The retained proteins are mainly involved in gene expression processes and predicted to be localized to plant mitochondria. They include a series of putative RNA and / or DNA binding proteins, 12 PPR proteins, 15 ribosomal proteins, other translation related proteins as well as proteins of unknown function (Figure 2). This led to the hypothesis that PRORP1 might be associated to other gene expression regulators and in some way to the translation apparatus in Arabidopsis mitochondria.

In order to test this hypothesis, mitochondrial complexes were separated by sucrose density gradients. Eight fractions representing the entire gradients were collected. Equivalent amounts of proteins from each fraction were reacted with HA-tag antibodies. The 60-kDa signal of PRORP1-HA was detected in fractions at the bottom of the gradients (Figure S2). Fractions were also reacted with antibodies specific for the mitochondrial ribosomal protein RPS1 and NAD9 from respiratory complex I. NAD9 signal was only detected in the two top fractions of the gradients, suggesting that they contain complexes of sizes up to 2000 kDa. In contrast, RPS1 was detected in the same bottom fractions of the gradient as PRORP1 as well as in top fractions. These bottom fractions likely contain polysomes while the top fractions likely correspond to free ribosomes (Uyttewaal *et al.*, 2008). This suggested that PRORP1 might be associated with polysomes. To prove this, samples were treated with puromycin that specifically destabilizes ribosomes (Lu and Draper, 1994). In these assays, PRORP1, similar to RPS1, was no longer detected in bottom fractions (Figure S2). This confirms that PRORP1 is indeed associated with polysomes. While direct interaction of PRORP1 with ribosomal subunits cannot be ruled out, it is most likely indirectly associated with ribosomes because PRORP1 is not detected in free ribosome fractions. However, the precise nature of this association with polysomes is unknown.

PRORP1 directly interacts with the mitochondrial nuclease MNU2

As a next step, a yeast two hybrid-like approach was used to try to identify proteins that might be in direct interaction with PRORP1. Sucrose gradient experiments described above had suggested that interactions with ribosomal proteins are likely indirect. Similarly, a previous study has suggested that PPR proteins and PRORP can interact indirectly via RNA. Indeed, PRORP and PPR proteins can cooperate for the maturation of a same transcript as shown for RFL2 and PRORP1 that are both required for the processing of *orf291* transcript in Arabidopsis mitochondria (Fujii *et al.*, 2016). Although the two proteins functionally interact, they do not appear to interact directly (Fujii *et al.*, 2016). In this light, the direct interaction of PRORP1 with PPR proteins identified here was not explored. Direct protein associations were rather investigated for the other putative RNA binding proteins identified in this analysis. Selected proteins were thus MNU2, a putative nuclease containing a NYN-like domain, shown to be involved in mRNA 5'maturation in Arabidopsis mitochondria (Stoll and Binder, 2016), a glycine rich RNA binding protein of unknown function, PNPase, a 3' to 5' exonuclease involved in RNA 3' maturation and decay in plant mitochondria (Holec *et al.*, 2006)(Hammani and Giege, 2014) and mTERF30 a putative RNA and / or DNA binding protein of unknown function (Kleine, 2012). The identification of these proteins in immuno-purifications of complexes suggested that RNase P activity held by PRORP1 might be connected to other 5' and / or 3' RNA maturation processes in Arabidopsis mitochondria.

Direct protein interaction was monitored with the "DUAL hunter" technology, a versatile yeast-two hybrid-like genetic system based on the "split ubiquitin" system, that enables to monitor interactions between any combination of soluble and / or membrane proteins. Briefly, PRORP1 was inserted in frame with the membrane protein Ost4p, the C-terminal half of ubiquitin and the transcription factor LexA-VP16 (PRORP1 Cub bait construct). Then, cDNAs encoding the proteins of interest were fused at the C-terminal of a mutated version of the N-terminal domain of ubiquitin (Nub X prey constructs). If bait and prey interact, Cub and Nub complement to form split-ubiquitin. Then, ubiquitin-specific proteases release LexA-VP16 that migrates to the nucleus and activates the transcription of reporter genes (Mockli *et al.*, 2007). The capacity of the PRORP1 Cub construct to activate reporter genes was investigated through the interaction with the wild type form of Nub (NubWT) that is able to interact with Cub and reconstitute functional ubiquitin without protein interaction with the bait. Then, the absence of auto-activation of the PRORP1 Cub construct was investigated through the interaction with (i) the mutated Nub (Nub) and (ii) the control prey Δp53 (Figure 3A). Then, PRORP1 Cub interaction with the MNU2, glycine rich RBP6, mTERF30 and PNPase was investigated. The activation of reporter genes *ADE2* and *HIS3* could only be observed with the MNU2 construct, thus showing that only MNU2 can physically interact with PRORP1 (Figure 3A).

PRORP1 PPPY motif is a protein interaction platform

Subsequently, the PRORP1 / MNU2 interaction was characterized in details, i.e. to identify domains or motifs of the two proteins responsible for protein interaction. Bioinformatic analyses of MNU2 sequence features revealed that it contains a mitochondrial targeting signal, a N-terminal NYN-like

putative nuclease domain, three N and C-terminal OST-like putative RNA binding domains and a central WW protein / protein interaction domain (Stoll and Binder, 2016; Anantharaman *et al.*, 2010; Macias *et al.*, 1996). MNU2 cDNA was thus divided in three constructs containing the separate NYN, WW and two OST domains respectively (Figure 3B) and cloned in frame with Nub. Interactions with PRORP1 were visualized though the activation of *ADE2* and *HIS3* and quantified by β-galactosidase assays through the activation of the *lacZ* reporter gene. This revealed that the WW construct interacts with PRORP1 at a level close to that of wild type MNU2, whereas interaction levels with the OST and NYN constructs decreased by 80% and 76% respectively as compared to wild type MNU2 (Figure 3C). This strongly suggests that the WW domain of MNU2 is responsible for the interaction with PRORP1. Functional and mechanistic analyses of WW domain containing proteins have revealed that WW domains specifically interact with proline rich motifs, most of the times characterized by a PPPY signature (Yagi *et al.*, 1999). Interestingly a proline rich domain is conserved in PRORP sequences, localized in PRORP connecting domain and facing toward the outside of PRORP Λ shape (Figure 3 and Figure 4). While a PxxY signature is fairly conserved in distantly related PRORP sequences (Figure 5) i.e. in Opisthokonta, Excavata, SAR and Archaeplastida, a MPPPYS motif is particularly well conserved in Streptophyta, including Arabidopsis (Figure 5). In order to test the importance of PRORP1 motif MPPPYS for the interaction with MNU2, point mutants of the PRORP1 Cub construct were generated, i.e. including mutations of the first proline, of the tyrosine and of the serine of the motif. Interaction assays of the point mutants with wild type MNU2 revealed that the serine mutant interacts with MNU2 at a level close to that of wild type PRORP1, while the proline and tyrosine mutants interaction levels decreased by 69% and 73% respectively. This is in accordance with previous analyses of the interaction of PEB2 PPPY motif with YAP WW domains that had shown that the first proline and the tyrosine of the motif are the most important residues for the interaction with WW domains (Yagi *et al.*, 1999). Altogether, results strongly suggest that the WW domain of MNU2 and the PPPY motif of PRORP1 are responsible for PRORP1 / MNU2 interaction in Arabidopsis mitochondria.

PRORP1 interaction with MNU2 is restricted to mitochondria

In order to identify the subcellular distribution of MNU2 interaction with PRORP1, MNU2 specific antibodies were raised. An MNU2 specific signal detected at an apparent molecular weight of about 100 kDa was not observed in extracts from *mnu2* knock out lines (Figure S3). While MNU2 was already shown to participate in the 5' maturation of some mitochondrial mRNAs (Stoll and Binder, 2016), it was not clear whether MNU2 localisation is restricted to mitochondria or whether it is a dual targeted protein similar to PRORP1. Western analysis performed on purified chloroplasts and mitochondria fractions unambiguously showed that MNU2 is only present in mitochondria (Figure S3), thus showing that the PRORP1 / MNU2 interaction is restricted to mitochondria. MNU2 specific antibodies were then reacted on blots representing Arabidopsis mitochondrial complexes separated on two dimensional Blue Native / SDS PAGE gels. This detected a signal corresponding to a complex of about 150 kDa (Figure 1C). This signal might correspond to the PRORP1 / MNU2 complex.

MNU2 is required for tRNA accumulation in Arabidopsis mitochondria

Since PRORP1 is involved in tRNA maturation, the involvement of its interacting partner, MNU2, in this process was also monitored. While MNU2 had already been described as a non-essential nuclease involved in the maturation of mRNA 5' ends in plant mitochondria (Stoll and Binder, 2016), its function for tRNA biogenesis had not been explored. For this, mitochondrial RNAs from *mnu1/mnu2* knock out lines and from control wild type plants were compared by RNA blots hybridizations (Figure 6). This revealed that the exemplarily chosen mitochondrial tRNA^{Ser} (GCU), tRNA^{Lys} (UUU) and tRNA^{Gly} (GCC) levels decreased by 22%, 41% and 42% respectively in three biological replicate experiments (i.e. with RNAs extracted from independent mitochondrial purifications). Wilcoxon tests were performed to verify that the average values for the biological replicates were significantly different ($p < 0.05$) between control and mutant samples. As a control, mitochondrial 5S rRNA levels were measured and did not vary in *mnu1/mnu2* mutants as compared to control plants. Although of moderate effect and non-essential (contrary to PRORP1 that is essential for tRNA maturation), the function of MNU2 does appear to be related to the accumulation of mitochondrial mature tRNAs *in vivo* (Figure 6).

DISCUSSION

With this study, the functional network of PRORP1 in Arabidopsis mitochondria begins to be unravelled. The presence of PRORP1 in polysome fractions while it is absent from free ribosome fractions suggests that it might be indirectly associated with the translation apparatus in plant mitochondria. This interaction could be mediated by RNA, i.e. PRORP1 similar to other PPR proteins (Hammani and Giegé, 2014) might be involved in the maturation of mRNAs already loaded on mitochondrial ribosomes. Beyond tRNA maturation, PRORP1 was already shown to be involved in the maturation of mitochondrial mRNAs at the level of tRNA-like structures as observed *in vivo* for *nad6* 3' end maturation (Gutmann *et al.*, 2012) and *orf291* processing (Fujii *et al.*, 2016). Such an association would suggest that the maturation of mRNAs and translation are tightly coupled processes in plant mitochondria.

The analysis of immuno-precipitated protein complexes has revealed a number of potential protein partners of PRORP1. With the exception of MNU2, other candidates did not show direct interaction with PRORP1. This suggests that similar to the translation apparatus, PRORP1 is indirectly associated via proteins and / or RNAs to gene expression regulators such as RBP6, mTERF30 or PNPase.

The characterisation of PRORP1 / MNU2 interaction has revealed that a specific domain of MNU2 and a proline rich motif of PRORP1 are directly accountable for protein interaction. A phylogenetic analysis of 388 PRORP sequences representing the diversity of PRORP across eukaryotes (Lechner *et al.*, 2015) has revealed that the proline rich motif is very well conserved in Streptophyta (Figure 5). However, a degenerate motif where the first proline and the tryptophan are relatively well conserved is also present in other distantly related eukaryote branches (Figure 5). This suggests that the proline rich motif is ancient and might have already been present in an ancestral PRORP, in an organism at the root of the major modern eukaryote groups. While the conserved PPPY

motifs in Streptophyta are most probably all involved in protein interactions, the functionality of the degenerate motifs in other eukaryote groups is questionable but cannot be ruled out. Similarly, the distribution of MNU2 orthologues across eukaryotes was examined. The analysis of genomic data available identified 109 MNU2-like sequences, specifically in Spermatophyta. All these proteins are predicted to be localised to mitochondria and / or plastids similar to PRORP1. Beyond seed plants, no other protein bearing all the hallmarks of MNU proteins (i.e. NYN, OST and WW domains) could be identified (Figure 5). This suggest that PRORP / MNU interactions are specific to seed plant organelles. The conservation of the proline rich domain in nuclear PRORP proteins, while MNU2-like proteins are strictly organellar, suggests that other WW domain proteins might interact with nuclear PRORPs in Embryophyta nuclei. Similarly, other yet unidentified WW domain proteins might interact with PRORP proteins in other eukaryote clades.

The major question remains to understand the functional reason for the interaction between PRORP1 and MNU2 in Arabidopsis mitochondria. A published report showed that the function of MNU proteins is required for the maturation of some Arabidopsis mitochondrial mRNAs (Stoll and Binder, 2016) and results obtained here suggest that it is might also be involved in the biogenesis of mitochondrial tRNAs. However, MNU2 mode of action is unknown. For instance, it is unclear whether MNU2 is an endo or an exonuclease. In the PRORP1 / MNU2 complex, a possible function of MNU2 could be to degrade tRNA 5' leader sequences after PRORP1 cleavage. Alternatively, MNU2 might directly participate in pre-tRNA and / or mRNA maturation. For example, it might trim long leader sequences, while PRORP1 would perform the final maturation of shortened leader sequences. This would be in agreement with the observations that pre-tRNAs are often transcribed with long leader sequences in plant mitochondria (Hammani and Giegé, 2014) while PRORP cleaves with higher efficiency pre-tRNAs with very short leader sequences *in vitro* (Howard *et al.*, 2016). Such a cooperativity of two nucleases for RNA maturation would be reminiscent of the concerted action of RNase II and PNPase for the 3' end maturation of mRNAs in plant mitochondria (Stoll and Binder, 2016; Perrin *et al.*, 2004).

The fast growing amount of data on PRORP proteins mode of action, i.e. the determination of their catalytic constants has surprisingly revealed that PRORP enzymes are not as good catalysts as ribonucleoprotein RNase P enzymes *in vitro* (Schelcher *et al.*, 2016). The identification of interaction partners such as MNU2 enables to propose that protein partner functions might enhance or regulate PRORP activity *in vivo*. In presence of additional factors, PRORP proteins might turn out to be better catalysts. Such hypotheses will have to be investigated through the functional and mechanistic characterisation of complexes involving PRORP proteins. In the longer term, complete interaction networks of PRORP enzymes in mitochondria, chloroplasts and the nucleus will reveal the full extent of their integration among gene expression and other cellular processes.

EXPARIMENTAL PROCEDURES

Plant material

Heterozygous Arabidopsis PRORP1 (At2g32230) knock-out mutants Col-0 *GK-385G09* (Gobert *et al.*, 2010) were transformed by agroinfiltration with a pGWB616 clone (Nakamura *et al.*, 2010) encoding full length PRORP1 fused at its C-terminal end with an HA epitope and placed under the control of PRORP1 endogenous promoter (Earley *et al.*, 2006). Plants homozygous for wild type *prorp1* mutation and complemented with HA tagged PRORP1 were selected for subsequent analyses. Knock out lines for *mnu1*, *mnu2* and the double knock out *mnu1/mnu2* are the ones used by Stoll et al. (Stoll and Binder, 2016).

Mitochondria purification and Blue Native analysis

Arabidopsis mitochondria were prepared from inflorescences by differential centrifugations and step density gradients as described previously (Giegé *et al.*, 1998). Mitochondrial complexes were resolved by Blue-Native PAGE. For this, 500 μg of mitochondrial proteins were resuspended in ACA750 buffer containing 750 mM amino di-caproic acid, 50 mM bis-Tris and 0.5 mM Na_2EDTA, pH 7.0. Protein complexes were solubilized with digitonin, 5/1 detergent/protein (w/w) for 30 min on ice, centrifuged at 100,000 g for 15 min at 4°C and 5% (v/v) Serva blue solution (750 mM ACA750 solution, 5% (w/v) Serva Blue G250) was added to the supernatant. Complexes were separated on 5 to 13% acrylamide gradient gels in 0.5 M amino di-caproic acid, 50 mM bis-Tris pH 7.0 buffer, with 50 mM bis-Tris pH 7.0 anode buffer and 50 mM Tricine, 15 mM bis-Tris, 0.02% (v/v) Serva Blue G250, pH 7.0 cathode buffer. Electrophoresis was carried out overnight at 5 mA. Gel lanes were cut out and denatured for 1 h at room temperature in 50 mM Tris-HCl pH 6.8, 1% (w/v) SDS and 1% (v/v) ß mercaptoethanol. For the second dimension, components of the various complexes were resolved by SDS-PAGE as described previously (Giegé *et al.*, 2003).

Polysome analysis

Polysome-enriched fractions were prepared as described previously (Uyttewaal *et al.*, 2008) with purified Arabidopsis mitochondria. For polysome destabilisation experiments, lysates were treated with 10 mM puromycin for 30 min on ice before sucrose gradient separation.

Immuno-precipitations of protein complexes

Immuno-precipitation experiments were performed with the µMACS™ system (Miltenyi Biotec). In brief, 1 mg of purified Arabidopsis mitochondria was solubilized in 1 ml buffer containing 50 mM Tris HCl pH 8.0, 150 mM NaCl, 1% Triton X-100, and protease inhibitors for 30 min on ice. The solubilized extract was incubated with 50 µL of microbeads conjugated with anti-HA antibodies for 30 min on ice and applied to µMACS columns placed in the magnetic field of a µMACS separator at 4°C. Columns were washed 4 times with 200 μl 50 mM Tris HCl pH 8.0, 150 mM NaCl, 1% Triton X-100 and once with 200 μl 20 mM Tris HCl pH 7.5. Complexes were eluted with 95°C preheated elution buffer containing 50 mM Tris HCl pH 6.8, 50 mM DTT, 1% (w/v) SDS, 1 mM EDTA, 0,005% (w/v) bromophenol blue and 10% (v/v) glycerol

Proteomic analysis of protein complexes

Mass spectrometry analyses of the eluted complexes were performed at the Strasbourg-Esplanade proteomic platform. In brief, proteins were trypsin digested, mass spectrometry analyses and quantitative proteomics were carried out by nano LC-ESI-MS/MS analysis on AB Sciex TripleTOF mass spectrometers and quantitative label-free analysis was performed through in-house bioinformatics pipelines. To identify significantly enriched proteins, a statistical analysis by spectral counts using the msmsTest R package was performed (Gregori *et al.*, 2013). The implemented negative binominal model, which is based on the solution provided by the edgeR package was used (Robinson *et al.*, 2010). p-value were then adjusted using the Benjamini & Hochberg method. Proteins that were over-represented in PROPR1 IP were visualized as a volcano plot that displays log2-fold-change and - log10-p.value on the x and y axes, respectively. Graph was plotted using the Plotly's R graphing library, with colour scale adjusted according to the adjusted p-value.

Yeast split ubiquitin assays

PRORP1 direct protein interactions were investigated in yeast with the DUAL hunter system, (DualsystemsBiotech®). Briefly, mature PRORP1 cDNA (without mitochondrial targeting signal) was cloned with *Sfi*I technology in pDHB1. The bait construct obtained was transformed in the yeast strain NMY51. The bait strain was co-transformed according to manufacturer's instructions with constructs representing mature forms of glycine-rich RBP6, mTERF30, PNPase and MNU2 (residues 72 to 924) cloned in prey vector pPR3N. Sequences of MNU2 corresponding to residues 72 to 539, 540 to 617 and 618 to 924 termed NYN, WW and OST were also cloned in pPR3N vector. Transformation in yeast was controlled by the growth on minus leucine and tryptophan media. Protein interaction was monitored by the expression of the reporter genes *ADE2*, *HIS3* and *lacZ*. The expression of *ADE2* and *HIS3* was visualized by the growth on minus adenine and histidine media supplemented with 10 mM 3-amino triazol. The expression of *lacZ* was followed by measuring at OD_{420} the accumulation of the product metabolized by β-galactosidase with 2.2 mM 2-nitrophenyl β-D-galactopyranoside (*o*-NPG, Sigma) as a substrate.

Protein expression and antibodies production

A cDNA corresponding to the C-terminal part of MNU2 (A539 to V924) was cloned in p0GWA to express recombinant protein fused to a C-terminal poly-histidine tag in *E. coli* BL21 cells. Proteins were purified in denaturing conditions by HisTrap affinity chromatography. Proteins were injected to rabbits to raise polyclonal antibodies. The serum was used at 1/ 20000 dilution for western analysis.

RNA blot analyses

RNA blot analyses of MNU2 knock-out plants were performed as described previously (Gutmann *et al.*, 2012). Blots were hybridised with radiolabelled gene specific oligonucleotides and signal revealed with a FLA-7000 PhosphorImager (Fujifilm) and quantified with ImageJ.

Bioinformatic analyses

Subcellular localisation predictions of protein identified in immuno-purification experiments were determined with TargetP (http://www.cbs.dtu.dk/services/TargetP/) and Predotar (http://urgi.versailles.inra.fr/predotar/predotar.html). Structure models were determined with the Phyre2 algorithm (Protein Homology/analogY Recognition Engine V2.0) in the intensive modelling mode (Kelley and Sternberg, 2009). Molecular docking and related figures were obtained with PyMOL (The PyMOL Molecular Graphics System, Version 1.5.0, Schrödinger, LLC). PRORP and MNU-like sequences were retrieved using the BLAST tool in NCBI and Phytozome1.2. Proteins were aligned using Muscle (Edgar, 2004). WebLogo 3 was used to highlight conserved residues (Crooks *et al.*, 2004).

SUPORTING INFORMATION

Figure S1. PRORP1 can be immuno-purified with extracts from Arabidopsis lines expressing PRORP1-HA.

Figure S2. PRORP1 is associated with polysomes in Arabidopsis mitochondria.

Figure S3. MNU2 is a mitochondrial protein.

Table S1. PRORP1 protein interaction network in Arabidopsis mitochondria.

REFERENCES

Altman, S. (2007) A view of RNase P. *Mol Biosyst*, **3**, 604–607.

Anantharaman, V. and Aravind, L. (2006) The NYN domains: novel predicted RNAses with a PIN domain-like fold. *RNA Biol*, **3**, 18–27.

Anantharaman, V., Zhang, D. and Aravind, L. (2010) OST-HTH: a novel predicted RNA-binding domain. *Biol Direct*, **5**, 13.

Antonicka, H., Sasarman, F., Nishimura, T., Paupe, V. and Shoubridge, E.A. (2013) The mitochondrial RNA-binding protein GRSF1 localizes to RNA granules and is required for posttranscriptional mitochondrial gene expression. *Cell Metab.*, **17**, 386–398.

Aragón, E., Goerner, N., Xi, Q., Gomes, T., Gao, S., Massagué, J. and Maclas, M.J. (2012) Structural basis for the versatile interactions of Smad7 with regulator WW domains in TGF-β pathways. *Structure*, **20**, 1726–1736.

Bonnard, G., Gobert, A., Pinker, F., Arrivé, M., Salinas, T. and Giegé, P. (2016) A single gene encodes both organelles and nuclear RNase P enzymes in *Chlamydomonas reinhardtii. Plant J.*,**87**, 270-280.

Boussardon, C., Salone, V., Avon, A., Berthome, R., Hammani, K., Okuda, K., Shikanai, T., Small, I. and Lurin, C. (2012) Two interacting proteins are necessary for the editing of the NdhD-1 site in Arabidopsis plastids. *Plant Cell*, **24**, 3684–3694.

Chen, T.H., Tanimoto, A., Shkriabai, N., Kvaratskhelia, M., Wysocki, V. and Gopalan, V. (2016) Use of chemical modification and mass spectrometry to identify substrate-contacting sites in proteinaceous RNase P, a tRNA processing enzyme. *Nucleic Acids Res*. **44**, 5344-5355.

Crooks, G.E., Hon, G., Chandonia, J.M. and Brenner, S.E. (2004) WebLogo: a sequence logo generator. *Genome Res*, **14**, 1188–1190.

Dreyfuss, G., Kim, V.N. and Kataoka, N. (2002) Messenger-RNA-binding proteins and the messages they carry. *Nat Rev Mol Cell Biol*, **3**, 195–205.

Earley, K.W., Haag, J.R., Pontes, O., Opper, K., Juehne, T., Song, K. and Pikaard, C.S. (2006) Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J*, **45**, 616–629.

Edgar, R.C. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, **5**, 113.

Fujii, S., Suzuki, T., Giegé, P., Higashiyama, T., Koizuka, N. and Shikanai, T. (2016) The Restorer-of-fertility-like 2 pentatricopeptide repeat protein and RNase P are required for the processing of mitochondrial orf291 RNA in Arabidopsis. *Plant J*, **86**, 504-513.

Giegé, P. (2013) Pentatricopeptide repeat proteins: a set of modular RNA-specific binders massively used for organelle gene expression. *RNA Biol*, **10**, 1417–1418.

Giegé, P., Konthur, Z., Walter, G. and Brennicke, A. (1998) An ordered Arabidopsis thaliana mitochondrial cDNA library on high- density filters allows rapid systematic analysis of plant gene expression: a pilot study. *Plant J*, **15**, 721–6.

Giegé, P., Sweetlove, L. and Leaver, C.J. (2003) Identification of mitochondrial protein complexes in Arabidopsis using two-dimensional Blue-Native polyacrylamide gel electrophoresis. *Plant Mol. Biol. Report.*, **21**, 133–144.

Gobert, A., Gutmann, B., Taschner, A., Gößringer, M., Holzmann, J., Hartmann, R.K., Rossmanith, W. and Giegé, P. (2010) A single Arabidopsis organellar protein has RNase P activity. *Nat Struct Molec Biol*, **17**, 740–744.

Gobert, A., Pinker, F., Fuchsbauer, O., Gutmann, B., Boutin, R., Roblin, P., Sauter, C. and Giegé, P. (2013) Structural insights into protein-only RNase P complexed with tRNA. *Nat Commun*, **4**, 1353.

Gregori, J., Villarreal, L., Sánchez, A., Baselga, J. and Villanueva, J. (2013) An effect size filter improves the reproducibility in spectral counting-based comparative proteomics. *J. Proteomics*, **95**, 55–65.

Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. and Altman, S. (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell*, **35**, 849–857.

Gutmann, B., Gobert, A. and Giegé, P. (2012) PRORP proteins support RNase P activity in both organelles and the nucleus in Arabidopsis. *Genes Dev.*, **26**, 1022–1027.

Hammani, K., Bonnard, G., Bouchoucha, A., Gobert, A., Pinker, F., Salinas, T. and Giegé, P. (2014) Helical repeats modular proteins are major players for organelle gene expression. *Biochimie*, **100**, 141–150.

Hammani, K. and Giegé, P. (2014) RNA metabolism in plant mitochondria. *Trends Plant Sci*, **19**, 380–389.

Hammani, K., Gobert, A., Hleibieh, K., Choulier, L., Small, I. and Giegé, P. (2011) An Arabidopsis dual-localized pentatricopeptide repeat protein interacts with nuclear proteins involved in gene

expression regulation. *Plant Cell*, **23**, 730–40.

Hartel, B., Zehrmann, A., Verbitskiy, D., Merwe, J.A. van der, Brennicke, A. and Takenaka, M. (2013) MEF10 is required for RNA editing at nad2-842 in mitochondria of *Arabidopsis thaliana* and interacts with MORF8. *Plant Mol Biol*, **81**, 337–346.

Hartmann, E. and Hartmann, R.K. (2003) The enigma of ribonuclease P evolution. *Trends Genet*, **19**, 561–569.

Hernandez-Cid, A., Aguirre-Sampieri, S., Diaz-Vilchis, A. and Torres-Larios, A. (2012) Ribonucleases P/MRP and the expanding ribonucleoprotein world. *IUBMB Life*, **64**, 521–528.

Holec, S., Lange, H., Kuhn, K., Alioua, M., Borner, T. and Gagliardi, D. (2006) Relaxed transcription in Arabidopsis mitochondria is counterbalanced by RNA stability control mediated by polyadenylation and polynucleotide phosphorylase. *Mol Cell Biol*, **26**, 2869–2876.

Holzmann, J., Frank, P., Loffler, E., Bennett, K.L., Gerner, C. and Rossmanith, W. (2008) RNase P without RNA: identification and functional reconstitution of the human mitochondrial tRNA processing enzyme. *Cell*, **135**, 462–474.

Howard, M.J., Karasik, A., Klemm, B.P., Mei, C., Shanmuganathan, A., Fierke, C.A. and Koutmos, M. (2016) Differential substrate recognition by isozymes of plant protein-only Ribonuclease P. *RNA*, **22**, 782–792.

Howard, M.J., Lim, W.H., Fierke, C.A. and Koutmos, M. (2012) Mitochondrial ribonuclease P structure provides insight into the evolution of catalytic strategies for precursor-tRNA 5' processing. *Proc Natl Acad Sci U S A*, **109**, 16149–16154.

Jarrous, N. (2017) Roles of RNase P and Its Subunits. *Trends Genet.*, **33**, 594–603.

Jourdain, A.A., Koppen, M., Wydro, M., Rodley, C.D., Lightowlers, R.N., Chrzanowska-Lightowlers, Z.M. and Martinou, J.C. (2013) GRSF1 regulates RNA processing in mitochondrial RNA granules. *Cell Metab.*, **17**, 399–410.

Kelley, L.A. and Sternberg, M.J.E. (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Protoc.*, **4**, 363–371.

Kleine, T. (2012) *Arabidopsis thaliana* mTERF proteins: evolution and functional classification. *Front Plant Sci*, **3**, 233.

Klemm, B.P., Karasik, A., Kaitany, K.J., et al. (2017) Molecular recognition of pre-tRNA by Arabidopsis protein-only Ribonuclease P. *RNA*, rna.061457.117.

Lechner, M., Rossmanith, W., Hartmann, R.K., Tholken, C., Gutmann, B., Giegé, P. and Gobert, A. (2015) Distribution of Ribonucleoprotein and Protein-Only RNase P in Eukarya. *Mol Biol Evol*, **32**, 3186–3193.

Lu, M. and Draper, D.E. (1994) Bases defining an ammonium and magnesium ion-dependent tertiary structure within the large subunit ribosomal RNA. *J Mol Biol*, **244**, 572–585.

Macias, M.J., Hyvonen, M., Baraldi, E., Schultz, J., Sudol, M., Saraste, M. and Oschkinat, H. (1996) Structure of the WW domain of a kinase-associated protein complexed with a proline-rich peptide. *Nature*, **382**, 646–649.

Mao, G., Chen, T.H., Srivastava, A.S., Kosek, D., Biswas, P.K., Gopalan, V. and Kirsebom, L.A. (2016) Cleavage of model substrates by *Arabidopsis thaliana* PRORP1 reveals new insights into its substrate requirements. *PLoS One*, **11**, e0160246.

Mockli, N., Deplazes, A., Hassa, P.O., Zhang, Z., Peter, M., Hottiger, M.O., Stagljar, I. and Auerbach, D. (2007) Yeast split-ubiquitin-based cytosolic screening system to detect interactions between transcriptionally active proteins. *Biotechniques*, **42**, 725–730.

Nakamura, S., Mano, S., Tanaka, Y., et al. (2010) Gateway binary vectors with the bialaphos resistance gene, bar, as a selection marker for plant transformation. *Biosci Biotechnol Biochem*, **74**, 1315–1319.

Perrin, R., Meyer, E.H., Zaepfel, M., Kim, Y.J., Mache, R., Grienenberger, J.M., Gualberto, J.M. and Gagliardi, D. (2004) Two exoribonucleases act sequentially to process mature 3'-ends of atp9 mRNAs in Arabidopsis mitochondria. *J Biol Chem*, **279**, 25440–25446.

Pinker, F., Bonnard, G., Gobert, A., Gutmann, B., Hammani, K., Sauter, C., Gegenheimer, P.A. and Giegé, P. (2013) PPR proteins shed a new light on RNase P biology. *RNA Biol*, **10**, 1457–1468.

Pinker, F., Schelcher, C., Fernandez-Millan, P., Gobert, A., Birck, C., Thureau, A., Roblin, P., Giegé, P. and Sauter, X.C. (2017) Biophysical analysis of Arabidopsis protein-only RNase P alone and in complex with tRNA provides a refined model of tRNA binding. *J. Biol. Chem.*, **292**, 13904–13913.

Robinson, M.D., McCarthy, D.J. and Smyth, G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–40.

Salinas-Giegé, T., Giegé, R. and Giegé, P. (2015) tRNA biology in mitochondria. *IJMS*, **16**, 4518–

4559.

Schelcher, C., Sauter, C. and Giegé, P. (2016) Mechanistic and structural studies of protein-only RNase P compared to ribonucleoproteins reveal the two Faces of the same enzymatic activity. *Biomolecules*, **6**, E30.

Smirnov, A., Schneider, C., Hör, J. and Vogel, J. (2017) Discovery of new RNA classes and global RNA-binding proteins. *Curr. Opin. Microbiol.*, **39**, 1–9.

Stoll, B. and Binder, S. (2016) Two NYN domain containing putative nucleases are involved in transcript maturation in Arabidopsis mitochondria. *Plant J*, **85**, 278–288.

Takenaka, M., Zehrmann, A., Verbitskiy, D., Kugelmann, M., Hartel, B. and Brennicke, A. (2012) Multiple organellar RNA editing factor (MORF) family proteins are required for RNA editing in mitochondria and plastids of plants. *Proc Natl Acad Sci U S A*, **109**, 5104–5109.

Täschner, A., Weber, C., Buzet, A., Hartmann, R.K., Hartig, A. and Rossmanith, W. (2012) Nuclear RNase P of *Trypanosoma brucei*: A Single Protein in Place of the Multicomponent RNA-Protein Complex. *Cell Rep.*, **2**, 19–25.

Uyttewaal, M., Mireau, H., Rurek, M., Hammani, K., Arnal, N., Quadrado, M. and Giegé, P. (2008) PPR336 is associated with polysomes in plant mitochondria. *J Mol Biol*, **375**, 626–636.

Walczyk, D., Gößringer, M., Rossmanith, W., Zatsepin, T.S., Oretskaya, T.S. and Hartmann, R.K. (2016) Analysis of the Cleavage Mechanism by Protein-Only RNase P Using Precursor tRNA Substrates with Modifications at the Cleavage Site. *J. Mol. Biol.*, **428**, 4917–4928.

Yagi, R., Chen, L.F., Shigesada, K., Murakami, Y. and Ito, Y. (1999) A WW domain-containing yes-associated protein (YAP) is a novel transcriptional co-activator. *EMBO J*, **18**, 2551–2562.
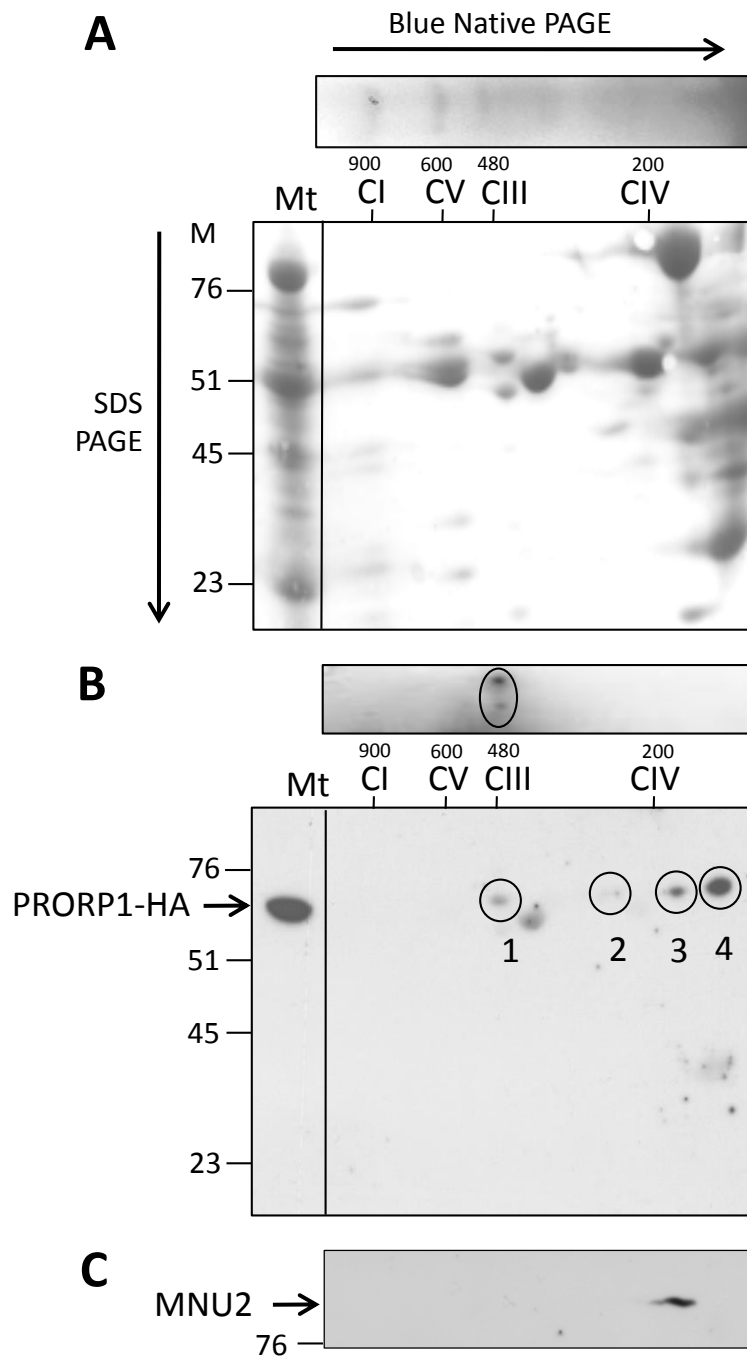
Figure 1. Blue Native analysis identifies PRORP1 in mitochondrial protein complexes. (A) Stained membrane representing Arabidopsis mitochondrial complexes separated in a first dimension by Blue Native PAGE and as a second dimension by SDS PAGE. "Mt" is a control track where total mitochondrial proteins were separated by SDS PAGE. CI to CV indicate respiratory chain complexes I to V assigned according to literature (Giegé et al., 2003) and their respective sizes in kDa. M indicate molecular weight markers in kDa. (B) Western analysis of the membrane reacted with HA-specific antibodies, revealing the 60 kDa signal of PRORP1-HA in the control track and four signals on the 2D BN / SDS PAGE membrane. "1-3" represent three complexes involving PRORP1-HA, while "4" shows a signal seemingly corresponding to PRORP1-HA alone. (C) Western analysis of a two dimensional membrane reacted with MNU2-specific antibodies, revealing the 100 kDa signal of MNU2.

A

| Accession	Name	IP PRORP1	IP col0	LogFC	p.value	adjp	rank
AT2G32230	PRORP1	265.6	0	11.1	1.2E-14	1,7E-11	1
RNA binding proteins							
AT5G09840	**MNU2**	8	0	6.0	5.6E-05	8.5E-03	7
AT5G14580	**PNPase**	7.5	0	5.9	2.9E-04	1.5E-02	28
AT1G61980	**mTERF30**	6.5	0	5.7	3.9E-04	1.6E-02	34
AT1G18630	**glycine-rich RBP6**	3.1	0	4.7	4.8E-04	1.7E-02	39
AT3G22310	mitochondrial RNA helicase 1	26.5	1.3	4.2	1.3E-03	3.6E-02	54
PPR proteins							
AT2G37230	PPR	24.1	0.5	5.3	4.5E-05	8.5E-03	5
AT3G02650	PPR	8	0	6.0	4.5E-05	8.5E-03	6
AT1G55890	PPR	9.2	0	6.2	1.1E-04	9.9E-03	15
AT3G49240	PPR	15.4	0	7.0	1.2E-04	9.9E-03	17
AT3G13160	PPR	14.5	0.5	4.5	2.5E-04	1.4E-02	26
AT3G13150	PPR	6	0	5.6	2.8E-04	1.5E-02	27
AT4G36680	PPR	8.4	0	6.1	3.2E-04	1.5E-02	29
AT2G15690	PPR	9.8	0	6.3	4.1E-04	1.7E-02	35
AT5G15980	PPR	6.5	0	5.7	5.2E-04	1.8E-02	42
AT1G80270	PPR	32.4	1.5	4.3	6.2E-04	2.0E-02	44
AT3G15590	PPR	17.8	1	4.0	1.6E-03	4.1E-02	56
AT4G01990	PPR	4	0	5.0	2.1E-03	4.6E-02	65
Ribosomal proteins							
AT3G59650	RPL51	7.5	0	5.9	7.8E-06	5.6E-03	2
AT3G49080	RPS5	8	0	6.0	4.2E-05	8.5E-03	4
AT3G18740	RPL7	8.4	0.2	4.9	5.8E-05	8.5E-03	8
AT3G18240	RPS24	6.5	0	5.7	6.3E-05	8.5E-03	10
AT5G66860	RPL25	8.6	0	6.1	6.8E-05	8.5E-03	11
ATMG01270	RPS7	3.6	0	4.9	9.8E-05	9.9E-03	13
AT2G47610	RPL7	11.1	0.2	5.3	2.0E-04	1.2E-02	23
ATMG00560	RPL2	7.2	0	5.9	2.0E-04	1.2E-02	24
AT1G16870	RPS28	6.8	0.2	4.6	3.3E-04	1.5E-02	31
AT3G09200	RPL10	11.5	0.3	4.7	4.8E-04	1.7E-02	40
AT5G46160	RPL14	7.9	0.2	4.8	4.9E-04	1.7E-02	41
ATMG00080	RPL16	2.5	0	4.4	9.0E-04	2.6E-02	49
AT1G64880	RPS5	5.5	0.3	3.6	1.1E-03	3.1E-02	53
AT2G42710	RPL1	15.4	1.3	3.4	1.9E-03	4.2E-02	64
AT3G24830	RPL13	7.1	0.2	4.6	2.1E-03	4.6E-02	66
Other translation related proteins							
AT4G11160	Initiation factor 2	12.2	0.2	5.4	2.8E-05	8.5E-03	3
AT3G58140	Phe-tRNA synthetase	10.6	0	6.4	1.9E-04	1.2E-02	22
AT2G45030	Elongation factor EFG/EF2	5.5	0	5.5	1.8E-03	4.2E-02	60
Other proteins							
AT3G15660	glutaredoxin 4 (GRX4)	6	0	5.6	6.0E-05	8.5E-03	9
AT1G30680	toprim domain protein	8.1	0	6.0	1.1E-04	9.9E-03	16
AT1G53645	hydroxyproline-rich glycoprot	9.8	0	6.3	1.3E-04	1.0E-02	18
AT2G43360	BIOTIN AUXOTROPH 2	10.1	0	6.4	1.5E-04	1.1E-02	19
AT4G37910	HSP70	74.9	9.7	2.9	1.8E-04	1.2E-02	20
AT5G62270	mucin-related protein	5.9	0	5.6	1.8E-04	1.2E-02	21
AT4G20360	"RAB GTPase homolog E1B	19.1	1.2	3.9	2.3E-04	1.4E-02	25
AT2G37250	adenosine kinase (ADK)	5.8	0	5.6	3.2E-04	1.5E-02	30
AT3G07770	HSP89	16	0.2	5.8	3.3E-04	1.5E-02	32
AT5G26860	"lon protease 1 (LON1)	6.9	0	5.8	3.4E-04	1.5E-02	33
AT1G06130	glyoxalase 2-4 (GLX2-4)	4.5	0	5.2	4.3E-04	1.7E-02	36
AT5G62530	aldehyde dehydrogenase 12A1	5.6	0	5.5	4.5E-04	1.7E-02	37
AT3G08530	Clathrin, heavy chain	18.8	0	7.2	4.6E-04	1.7E-02	38
15 other proteins							
Unknown proteins							
AT1G26750	unknown protein	6.9	0	5.8	7.1E-05	8.5E-03	12
AT5G24165	unknown protein	5.5	0	5.5	1.1E-04	9.9E-03	14
AT5G49210	unknown protein	3	0	4.6	1.7E-03	4.2E-02	58

B
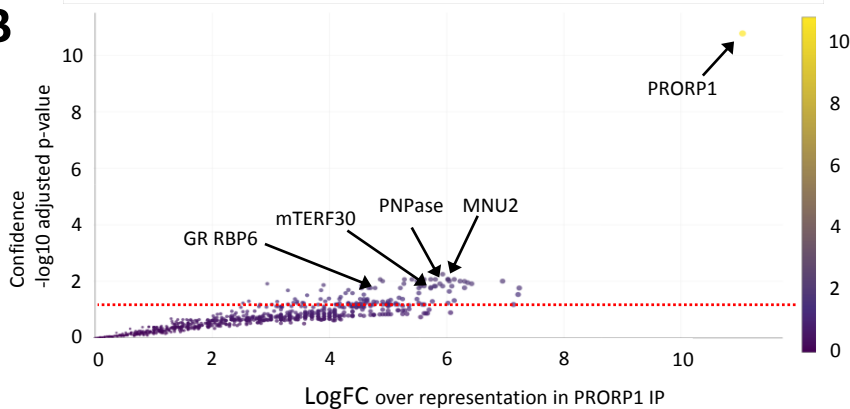
Figure 2. PRORP1 protein interaction network in Arabidopsis mitochondria. (A) Proteins over represented in 8 independent PRORP1 immunoprecipitation (IP) fractions were identified by a statistical analysis on spectral counts using the msmsTest package. 67 proteins with adjusted p-values "adjp" under 0.05 and thus considered as *bona fide* interacting partners are listed here. Proteins were ranked according to decreasing p-values "rank" and sorted by functional categories. "IP PRORP1" shows the average number of spectra for the respective proteins in PRORP1 IPs, "IP col0" shows average numbers of spectra in control experiments where IPs were performed with wild-type col0 plants and "LogFC" indicates the fold change over representation of proteins in PRORP1 IPs expressed in log2 scale. (B) The PRORP1 interacting partners are displayed as a volcano plot according to their over representation in PRORP1 IPs (x-axis, LogFC) and to the statistical confidence in their enriched identification as shown by decreasing p-values (y-axis, -log10(adjp)). The red dotted line indicates the significativity threshold of 1.3 (–log10(0.05)). PRORP1 as well as the four proteins tested here for direct interaction are indicated by arrows.
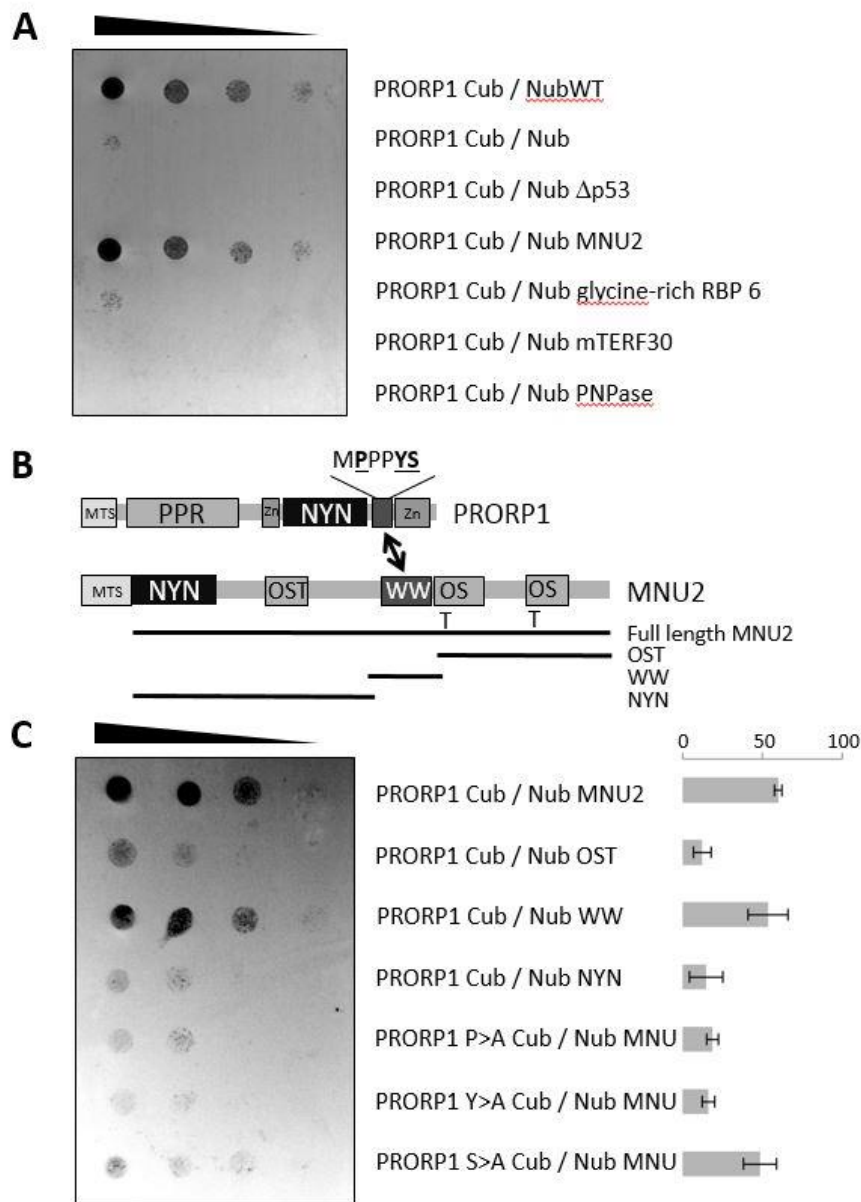
Figure 3. PRORP1 direct interaction with MNU2 is evidenced by yeast split-ubiquitin assays. (A) The upper panel shows interaction assays of the bait construct PRORP1 Cub with positive control NubWT and negative controls Nub and Dp53. The lower panel shows interaction assays of PRORP1 with MNU2, glycine rich RBP6, mTERF30 and PNPase. Drops representing four five-fold serial dilutions of yeast cultures of double transformants were deposited on plates with minus leucine, tryptophan, adenine and histidine media supplemented with 10 mM 3-amino triazol. Black triangles indicate decreasing cell concentrations in the yeast drops. (B) Schematic representation of PRORP1 and MNU2, highlighting their representative domains. Positions of the MPPPYS motif of PRORP1 that were mutated to alanine are underlined and black bars represent the individual sub-constructs of MNU2 termed "NYN", "WW" and "OST" that were used in interaction assays. (C) The four first lanes show interaction assays of PRORP1 Cub with MNU2 mutants, while the lower lanes show interactions of full length MNU2 with PRORP1 point mutants. Histograms indicate interaction quantifications measuring the accumulation of a b-galactosidase reaction product measured in nanomole / min. Error bars represent standard deviations for experiments performed with three independent yeast double transformants.
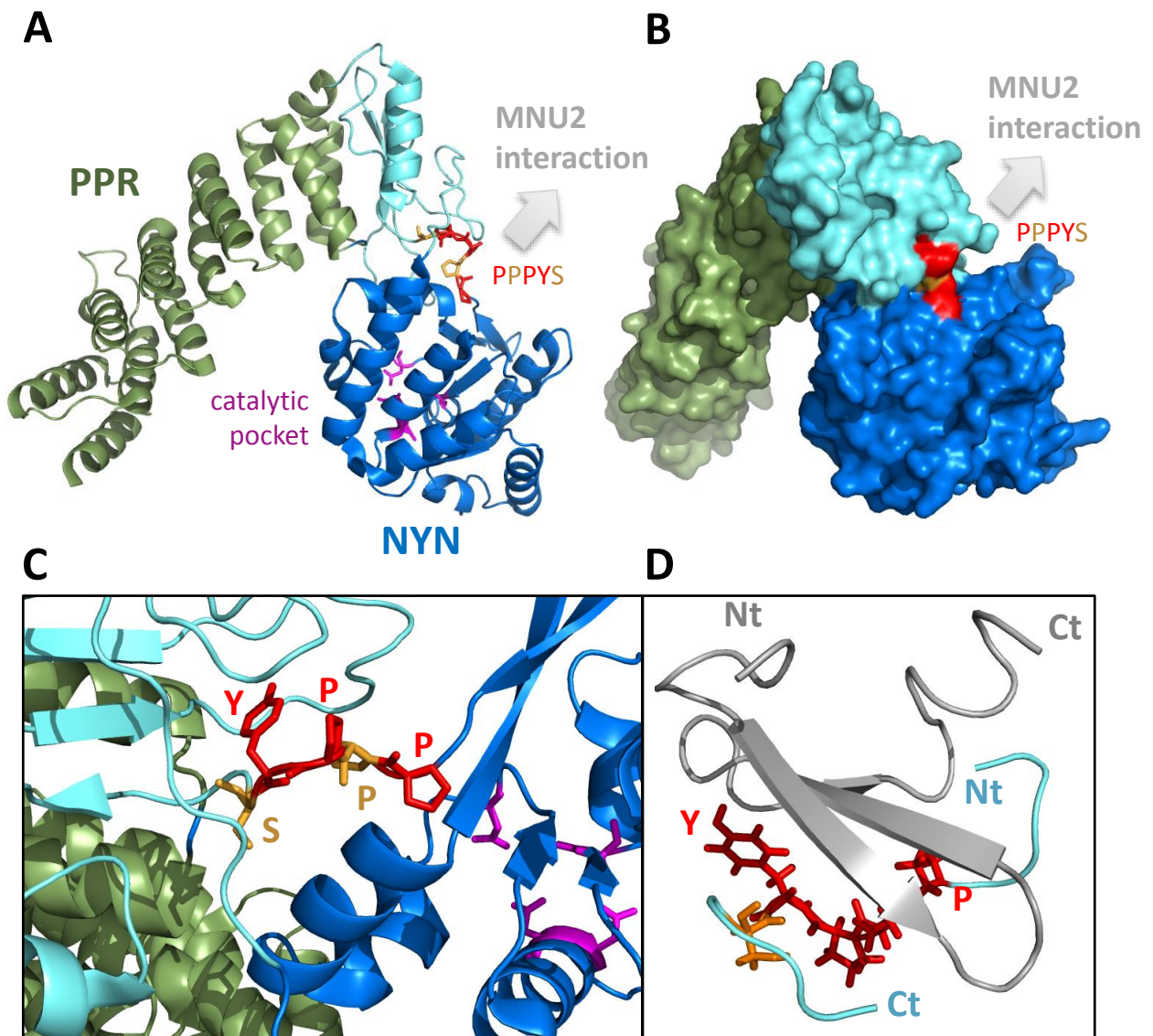
Figure 4. Structural model of the interaction between PRORP1 motif PPPY and MNU2 WW domain. (A) Crystal structure of Arabidopsis PRORP1 (Howard *et al.*, 2012), indicating the location of the PPPY motif shown in red, highlighting its accessibility for protein interaction. (B) Structural envelope of Arabidopsis crystal structure shown in A. (C) Close-up of the PRORP1 PPPY motif showing the orientation of individual residues, in particular the accessibility of the first proline and tyrosine that were shown to be important for protein / protein interaction, and inaccessibility of the second proline and serine (orange) not required for interaction. (D) The NMR structure of the WW domain of Smurf2 (grey) in interaction with the PPPY motif (red) of Smad7 (PDB, 2LTZ) (Aragón *et al.*, 2012) shows the orientation of the WW domain interaction with the proline rich motif. Extrapolation to PRORP1 / MNU2 interaction enables to propose that the catalytic domain of MNU2 situated in Nt of the WW domain might be oriented toward the catalytic pocket of PRORP1.
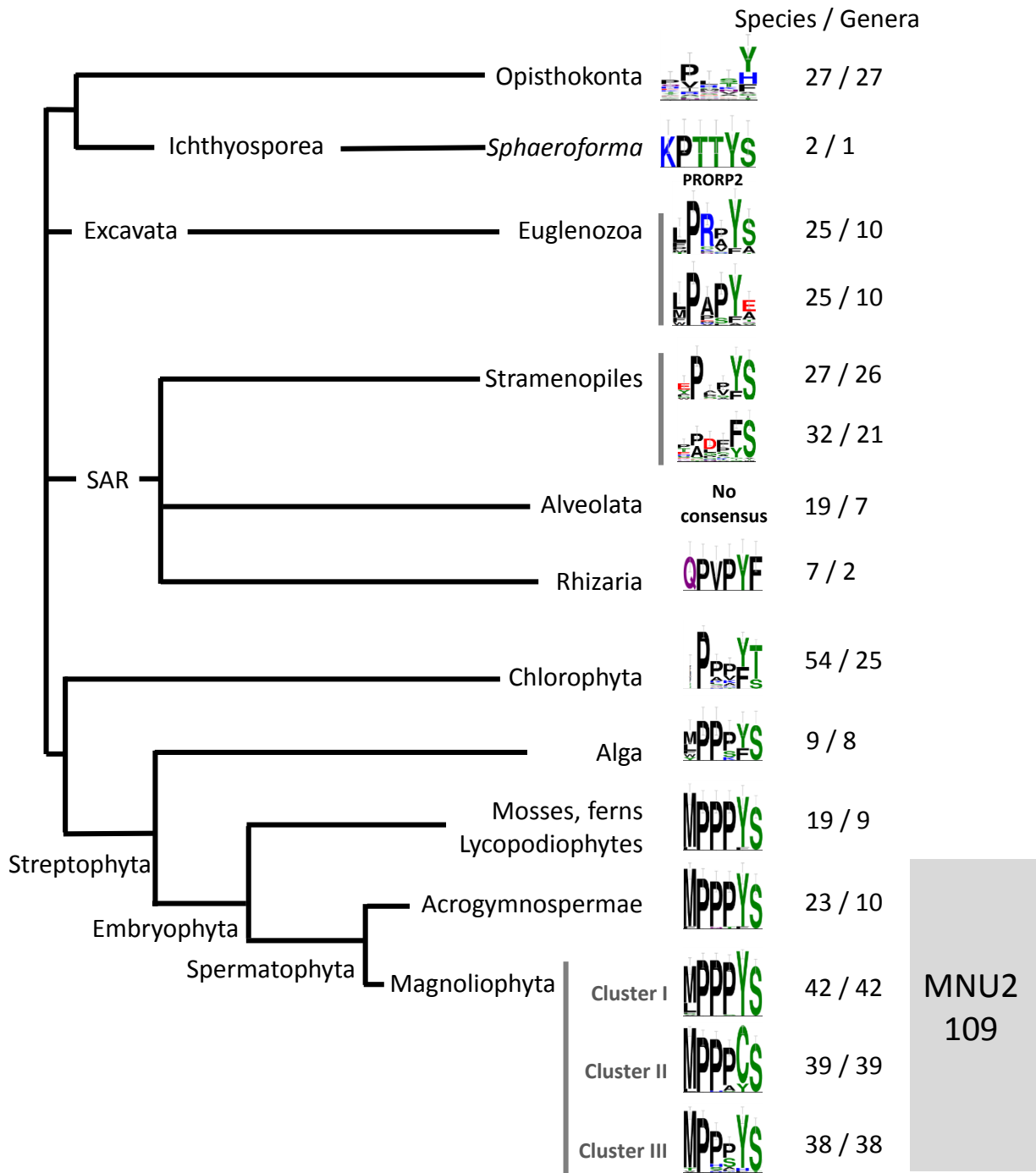
Figure 5. Phylogenetic distribution of PRORP1 PPPY motif and MNU2-like sequences. Phylogenetic distribution of PRORP sequences from representative eukaryote groups drawn according to (Lechner *et al.*, 2015). Logos represent the conservation of the proline rich motif in the respective eukaryote groups. Logos were generated with 388 PRORP sequences. Beyond Ichthyosporea where little sequence information is available (2 species), the other eukaryote groups are represented by PRORP sequences from 7 to 54 species, representing 2 to 42 different genera. The grey box indicates the occurrence of *bona fide* MNU2 orthologue sequences identified in 109 species, specifically among Archaeplastida, in the subgroup of Spermatophyta
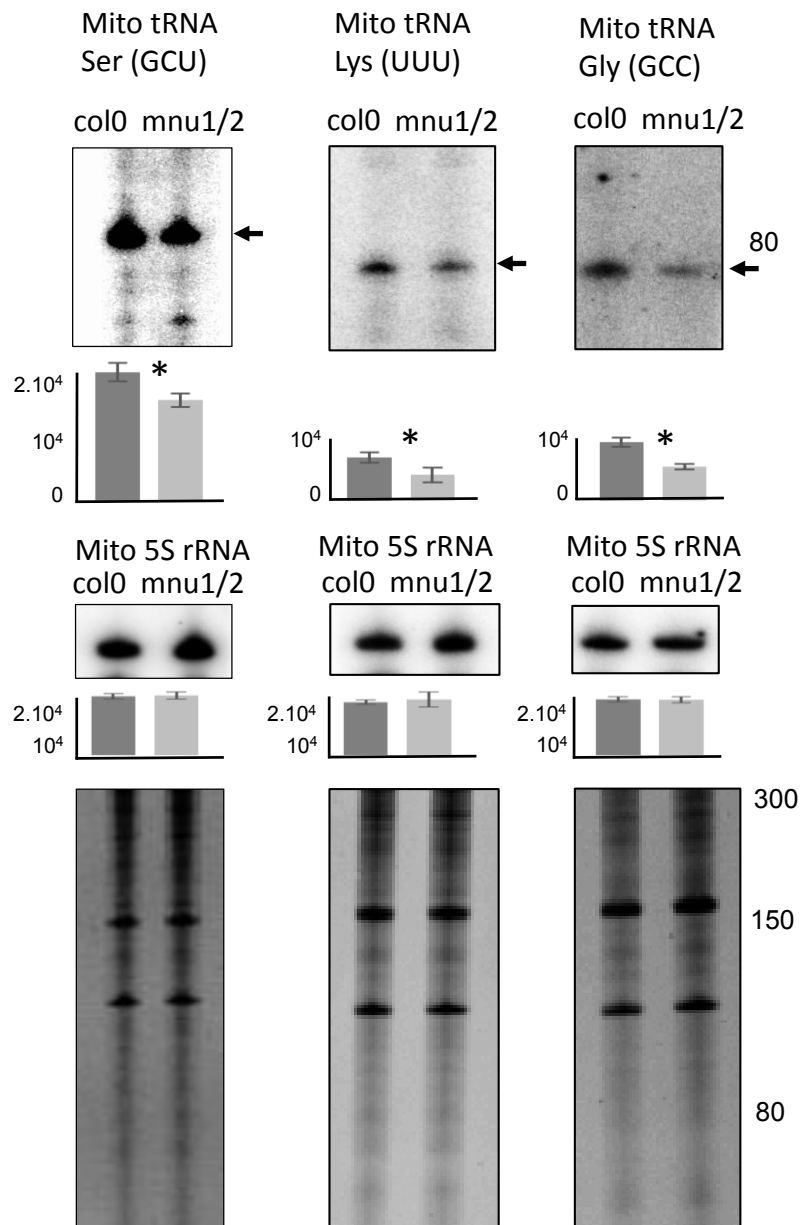
Figure 6. MNU2 is involved in mitochondrial tRNA accumulation. Total mitochondrial RNA from control (col0) and double knock out lines (mnu1/2) were separated on denaturing acrylamide gels, blotted and analysed by hybridisations using specific probes representing the Arabidopsis mitochondrial tRNA[Ser] (GCU), tRNA[Lys] (UUU) and tRNA[Gly] (GCC). Arrows show the main signals corresponding to the sizes of mature tRNAs (91, 76 and 76 nucleotides respectively). Control hybridisations were performed with 5S rRNAs and quantified. The average quantifications of signals (PhosphorImager signal intensities qantified by ImageJ) in three replicate experiments are shown as histograms with dark grey bars showing col0 samples and light grey bars, mutant samples. Stained blots used in the respective hybridizations are shown in the bottom panels. Molecular weight markers are indicated in ribonucleotides.
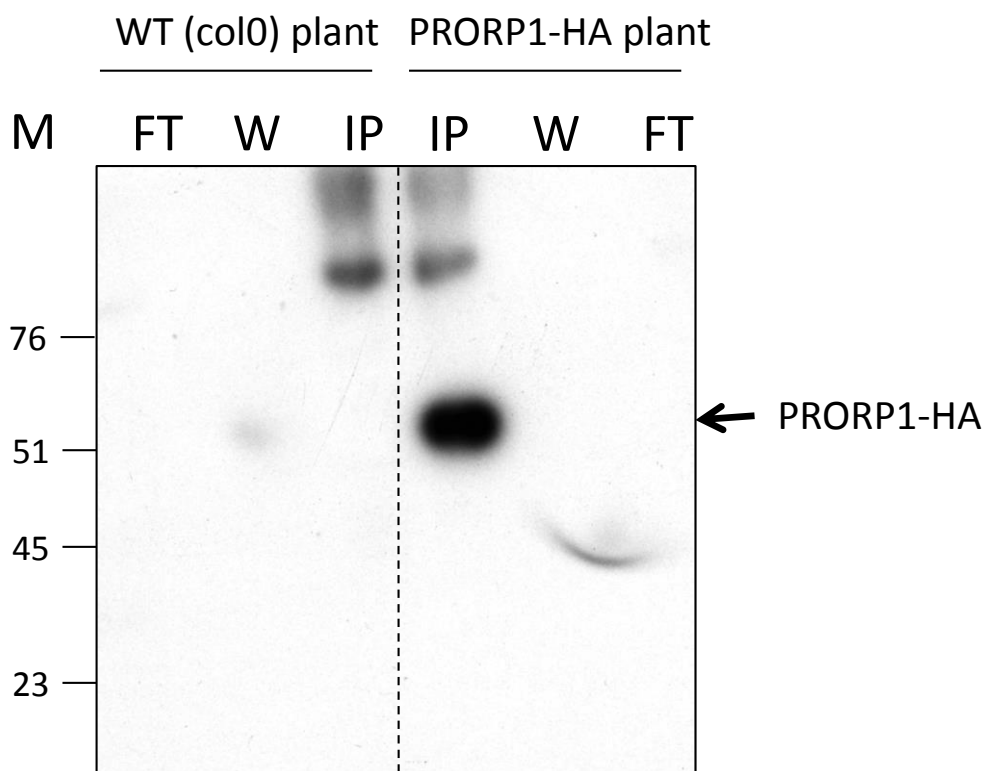
Figure S1. PRORP1 can be immuno-purified with extracts from Arabidopsis lines expressing PRORP1-HA. Immuno-purification experiments were performed with 1 mg of mitochondria purified from control wild type Arabidopsis plants (WT) and from an Arabidopsis lines expressing PRORP1-HA. "FT" represents the flow through of the affinity purification, "W" wash buffer and "IP" the eluted immuno-purified fraction. Fractions were separated by SDS PAGE and analysed by western blot using HA tag specific antibodies. M represents molecular weight markers indicated in kDa.
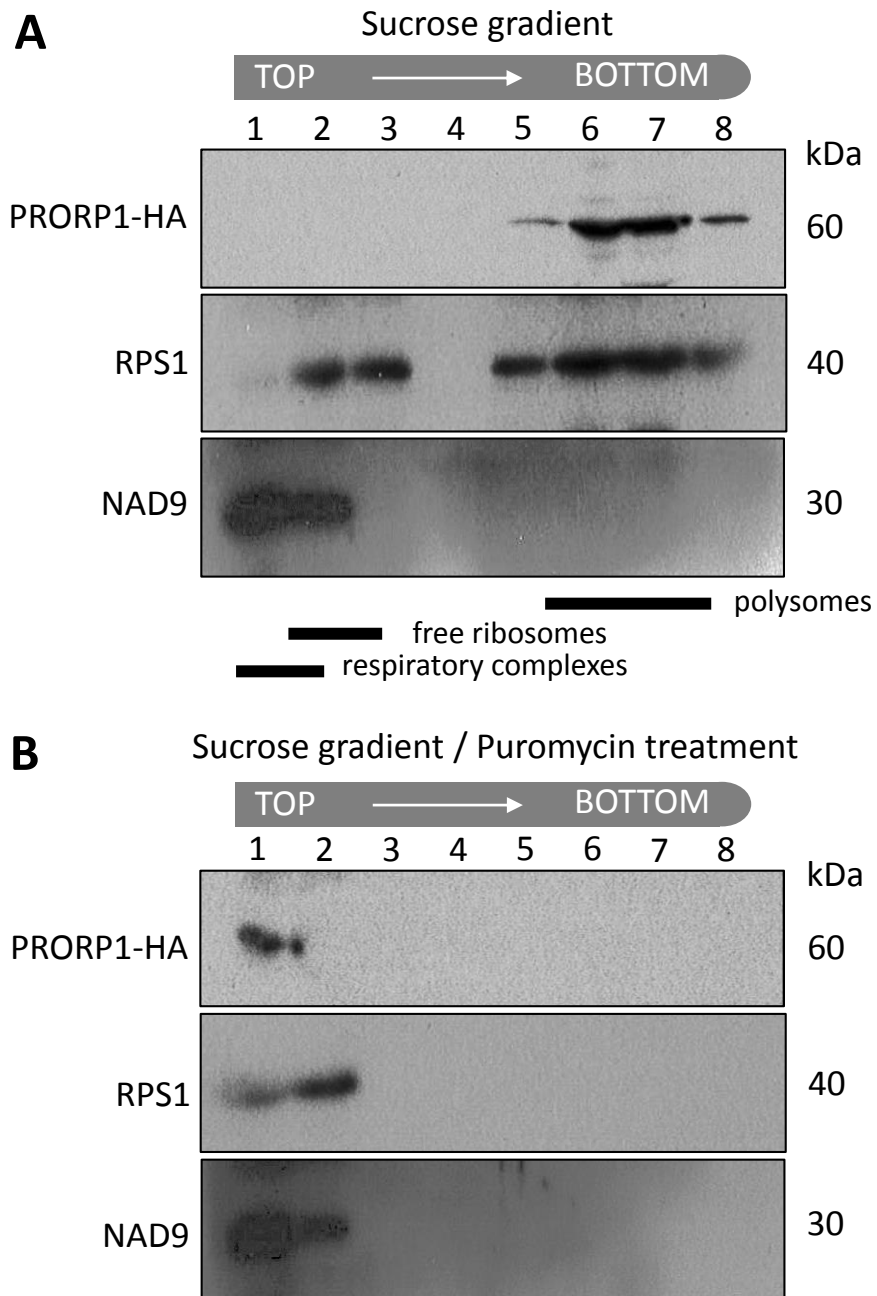
Figure S2. PRORP1 is associated with polysomes in Arabidopsis mitochondria. (A) Mitochondrial complexes from Arabidopsis lines expressing PRORP1-HA were separated on sucrose gradients. "1 to 8" are fractions representing the entire gradient containing complexes of increasing molecular weights. Fractions were separated by SDS PAGE and analysed on western blots reacted with HA-tag specific antibodies, with antibodies specific from the mitoribosome subunit RPS1 and respiratory complex I. (B) The same experiment was performed, preceded by a puromycin treatment that specifically destabilises ribosomes. Molecular weight markers indicated in kDa are shown on the right.
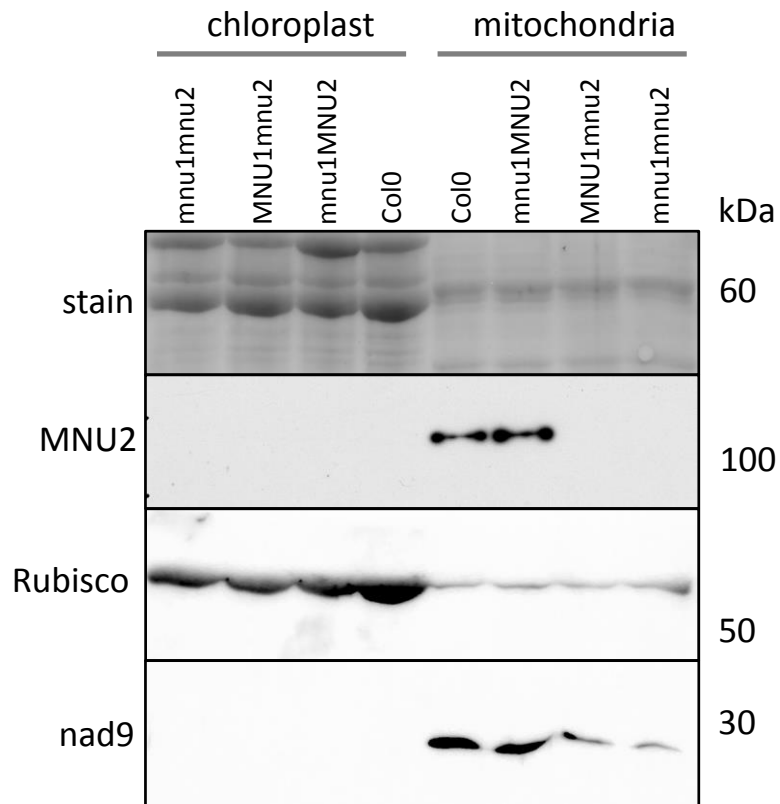
Figure S3. MNU2 is a mitochondrial protein. Immuno-detection of MNU2 on purified chloroplast and mitochondria fractions extracted from wild type Col0 plants, from mnu1mnu2 double knock out mutants and from single mnu1 or mnu2 knock out mutants. Signals are only observed in mitochondrial extracts from col0 and mnu1 knock out plants. Purity of fractions was assessed with antibodies for the chloroplast Rubisco large subunit (Rubisco) and the mitochondrial complex I subunit 9 (nad9).

Discussion

SSU proteins

Bact.	Mamm.	Yeast	Trypa.	Ara
bS1	bS1m	bS1	–	–
uS2	uS2m	uS2	–	uS2m
uS3	uS3m	uS3	uS3m	uS3m
uS4	–	uS4	–	uS4m
uS5	uS5m	uS5	uS5m	uS5m
bS6	bS6m	bS6	bS6m	bS6m
uS7	uS7m	uS7	–	uS7m
uS8	–	uS8	uS8m	uS8m
uS9	uS9m	uS9	uS9m	uS9m
uS10	uS10m	uS10	uS10m	uS10m
uS11	uS11m	uS11	uS11m	uS11m
uS12	uS12m	uS12	uS12m	uS12m
uS13	–	uS13	–	uS13m
uS14	uS14m	uS14	uS14m	uS14m
uS15	uS15m	uS15	uS15m	uS15m
bS16	bS16m	bS16	bS16m	bS16m
uS17	uS17m	uS17	uS17m	uS17m
bS18	bS18m	bS18	bS18m	bS18m
uS19	–	uS19	uS19m	uS19m
bS20	–	–	–	–
bS21	bS21m	bS21	bS21m	bS21m
bTHX	–	–	–	bTHXm

LSU proteins

Bact.	Mamm.	Yeast	Trypa.	Ara
uL1	uL1m	uL1	–	uL1m
uL2	uL2m	uL2	–	uL2m
uL3	uL3m	uL3	uL3m	uL3m
uL4	uL4m	uL4	uL4m	uL4m
uL5	–	uL5	–	uL5m
uL6	–	uL6	–	uL6m
bL9	bL9m	bL9	bL9m	bL9m
uL10	uL10m	uL10	uL10m	uL10m
uL11	uL11m	uL11	uL11m	uL11m
bL12	bL12m	bL12	bL12m	bL12m
uL13	uL13m	uL13	uL13m	uL13m
uL14	uL14m	uL14	–	uL14m
uL15	uL15m	uL15	uL15m	uL15m
uL16	uL16m	uL16	uL16m	uL16m
uL18	uL18m	–	–	uL18m
bL19	bL19m	bL19	bL19m	bL19m
bL20	bL20m	–	bL20m	bL20m
bL21	bL21m	bL21	bL21m	bL21m
uL22	uL22m	uL22	uL22m	uL22m
uL23	uL23m	uL23	uL23m	uL23m
uL24	uL24m	uL24	uL24m	uL24m
bL25	–	–	–	bL25m
bL27	bL27m	bL27	bL27m	bL27m
bL28	bL28m	bL28	bL28m	bL28m
uL29	uL29m	uL29	uL29m	uL29m
uL30	uL30m	uL30	uL30m	uL30m
bL31	bL31m	bL31	bL31m	bL31m
bL32	bL32m	bL32	bL32m	–
bL33	bL33m	bL33	bL33m	bL33m
bL34	bL34m	bL34	–	–
bL35	bL35m	bL35	bL35m	–
bL36	bL36m	bL36	bL36m	bL36m

SSU proteins (continued)

Bact.	Mamm.	Yeast	Trypa.	Ara
–	mS22	–	mS22	mS22
–	mS23	mS23	mS23	mS23
–	mS25	–	–	–
–	mS26	mS26	mS26	–
–	mS27 *	–	–	–
–	mS29 (DAP3)	mS29	mS29	mS29
–	mS31	–	–	–
–	mS33	mS33	mS33	mS33
–	mS34	–	mS34	mS34
–	mS35	mS35	mS35	mS35
–	mS37	mS37	mS37	–
–	mS38	mS38	mS38	–
–	mS39 *	–	–	–
–	mS40	–	–	–
–	–	mS41	mS41	–
–	–	mS42	mS42	–
–	–	mS43	mS43	–
–	–	mS44	–	–
–	–	mS45	–	–
–	–	mS46	–	–
–	–	mS47	mS47	mS47
–	–	–	mS48	–
–	–	–	mS49	–
–	–	–	mS50	–
–	–	–	mS51 *	–
–	–	–	mS52	–
–	–	–	mS53	–
–	–	–	mS54 *	–
–	–	–	mS55 *	–
–	–	–	mS56	–
–	–	–	mS57	–
–	–	–	mS58	–
–	–	–	mS59	–
–	–	–	mS60	–
–	–	–	mS61	–
–	–	–	mS62 *	–
–	–	–	mS63	–
–	–	–	mS64	–
–	–	–	mS65	–
–	–	–	mS66	–
–	–	–	mS67	–
–	–	–	mS68	–
–	–	–	mS69	–
–	–	–	mS70	–
–	–	–	mS71	–
–	–	–	mS72	–
–	–	–	mS73	–
–	–	–	mS74	–
–	–	–	–	mS75
–	–	–	–	mS76 (rPPR1)
–	–	–	–	mS77 (rPPR2)
–	–	–	–	mS78 (rPPR3a)
–	–	–	–	mS79 (rPPR3b)
–	–	–	–	mS80 (rPPR6)
–	–	–	–	mS81 (rPPR8)
–	–	–	–	mS82
–	–	–	–	mS83
–	–	–	–	mS84
–	–	–	–	mS85
–	–	–	–	mS86
–	–	–	–	mS87
–	–	–	–	mrpX

LSU proteins (continued)

Bact.	Mamm.	Yeast	Trypa.	Ara
–	mL37	–	–	–
–	mL38	mL38	mL38	–
–	mL39	–	–	–
–	mL40	mL40	mL40	mL40
–	mL41	mL41	mL41	mL41
–	mL42	–	mL42	–
–	mL43	mL43	mL43	mL43
–	mL44	mL44	–	–
–	mL45	–	–	–
–	mL46	mL46	mL46	mL46
–	mL48	–	–	–
–	mL49	mL49	mL49	–
–	mL50	mL50	–	–
–	mL51	–	–	–
–	mL52	–	mL52	–
–	mL53	mL53	mL53	mL53
–	mL54	mL54	–	mL54
–	–	mL57	–	–
–	–	mL58	–	–
–	–	mL59	–	–
–	–	mL60	–	–
–	–	mL61	–	–
–	mL62 (ICT1)	–	–	–
–	mL63	–	mL63	–
–	mL64 (CRIF1)	–	mL64	–
–	mL65	–	–	–
–	mL66	–	–	–
–	–	MHR1	–	–
–	–	–	mL67	–
–	–	–	mL68	–
–	–	–	mL69	–
–	–	–	mL70	–
–	–	–	mL71 *	–
–	–	–	mL72 *	–
–	–	–	mL73	–
–	–	–	mL74	–
–	–	–	mL75 *	–
–	–	–	mL76	–
–	–	–	mL77	–
–	–	–	mL78	–
–	–	–	mL79	–
–	–	–	mL80	–
–	–	–	mL81	–
–	–	–	mL82	–
–	–	–	mL83	–
–	–	–	mL84	–
–	–	–	mL85	–
–	–	–	mL86	–
–	–	–	mL87	–
–	–	–	mL88	–
–	–	–	mL89	–
–	–	–	mL90	–
–	–	–	mL91	–
–	–	–	mL92	–
–	–	–	mL93	–
–	–	–	mL94	–
–	–	–	mL95	–
–	–	–	mL96	–
–	–	–	mL97	–
–	–	–	mL98	–
–	–	–	mL99	–
–	–	–	mL100	–
–	–	–	–	mL101 (rPPR4)
–	–	–	–	mL102 (rPPR5)
–	–	–	–	mL103 (rPPR7)
–	–	–	–	mL104 (rPPR9)
–	–	–	–	mL105
–	–	–	–	mL106

Table 5: Compared protein compositions of charaterized mitoribosomes

Protein composition of the mammalian (**Mamm.**), yeast, trypanosome (**Trypa.**) and Arabidopsis (**Ara**) mitoribosomes compared with the bacterial one (**Bact.**). Proteins highlighted in light blue are universal to prokaryotes and eukaryotes, the ones in dark blue are not found in eukaryote cytosolic ribosomes. The proteins in yellow are specific to mitochondria and shared by different organisms, the ones in red are specific to the respective organism.

Proteins marked with a * are PPR proteins.

(Amunts et al., 2015; Desai et al., 2017; Greber et al., 2015; Ramrath et al., 2018)

Discussion

Composition of the plant mitoribosome composition compared with other mitoribosomes

In the manuscript by Waltz et al, the full protein composition of the Arabidopsis mitoribosome as well as the specific features of its ribosomal RNAs are described. The complementary biochemical approaches identified that 81 proteins compose the Arabidopsis mitoribosome. Among them, 13 proteins are shared with yeast and/or animal and/or trypanosoma mitoribosomes but 19 plant specific mitoribosome proteins were also identified, among which 10 are PPR proteins. In addition, Arabidopsis mitoribosome rRNAs are much larger as compared to prokaryotes and other mitoribosomes described to date.

Specificities of Arabidopsis mitoribosome rRNAs

Arabidopsis mitochondria has larger SSU and LSU rRNAs, with the SSU rRNA being 20% larger and the LSU rRNA 9% larger than *E.coli* rRNAs, respectively, and the 5S rRNA is present. The most striking feature of the Arabidopsis mitoribosome is its large head expansion caused by the 370 nt additional domain of the 18S rRNA. The function of this additional domain is unknown, but it is most likely coated by several of the plant specific r-proteins newly identified. Spatially this extension is located near the entry of the mRNA channel, therefore its function could be related to the recruitment of mRNAs to the SSU, and/or it could participate in the initiation of translation. Additionally, several expansion segments are present in the 26S rRNA, similarly to the yeast 21S rRNA. A higher resolution structure of Arabidopsis mitoribosome will probably give clues to understand the function of these rRNA expansion segments and novel domain.

Specificities of Arabidopsis mitoribosome protein composition

Concerning the original set of r-proteins inherited from the bacterial host, no general tendency can be observed toward the conservation of the proteins, except for bS20 which is absent in the mitoribosomes of all species described to date (Table 5). Most r-proteins could theoretically be lost, as experiments on bacterial ribosomes showed that it was possible to individually delete 22 of the 54 r-protein deleted from the genome (Shoji et al., 2011) some without any effect. As a result, in yeast, 5 proteins are missing (2 in the SSU – 3 in the LSU), in mammals 9 proteins are missing (6 in the SSU – 3 in the LSU), in trypanosome 15 are missing (7 in the SSU – 8 in the LSU) and in Arabidopsis 5 are missing (2 in the SSU – 3 in the LSU) which is

comparable to yeast. Interestingly the bacteria specific bTHX is present in plants mitoribosome (and chlororibosome) but not in the other species.

Concerning the mitochondrial specific r-proteins, some are shared by all mitoribosomes, namely, mS23, mS29, mS33, mS35, mL40, mL41, mL43, mL46, mL53. These proteins most likely constitute a common core of proteins that was acquired during the early evolution of eukaryotes. Others are shared by all but one, this the case for mS47 for example that is not found in mammals. In this case this is surely due to a secondary loss of the protein, specific to this lineage. This is most likely also the case for mS22, mS26, mS34, mS37, mS38, mL38, mL49 and mL54 which are found in 3 out of 4 mitoribosomes. The other possibility being that these proteins were all acquired independently in each lineage, which is less likely.

Proteins specific to each lineage are also found. In trypanosoma, the acquisition of lineage specific proteins has been particularly massive, with a stunning 61 specific r-proteins (27 in the SSU and 34 in the LSU). This huge accumulation of proteins is most likely related to the loss of the original bacterial r-proteins as well as to the strong reduction of its rRNAs. In Arabidopsis I identified 19 additional proteins (13 in the SSU and 6 in the LSU), which were not identified in any other mitoribosome. Mammals have 10 specific r-proteins and 9 are found in yeast. Interestingly, in plants, a significant portion of these species specific r-proteins are PPR proteins (i.e. 10 PPR proteins), in accordance with the prevalence of PPR proteins in plant mitochondria gene expression processes. Still, 2 PPR proteins also occur in mammals mitoribosome and 7 in trypanosoma. In yeast, no PPR protein is found in the mitoribosome.

What about the plant specific r-proteins?

The functions of the 19 plant specific proteins identified in the Waltz et al manuscript are not yet elucidated. Interestingly the majority of those additional proteins are located in the mitoribosomes SSU, where this 370 nt additional rRNA domain is found. RNA is a fragile molecule, especially in the mitochondrial matrix where the pH is rather basic (Santo-Domingo and Demaurex, 2012). Hence several of those rPPRs could be involved in coating the additional rRNA domain to protect it from degradation and/or to stabilize its structure. Additionally, PPR proteins are known to be ssRNA binding proteins. However, in mammals, the mS39 protein does not bind to rRNA, it rather seems to act as a platform to thread mRNAs into the mitoribosome (Bieri et al., 2018). A similar role could be played by the plant rPPRs. It rises an important question: are these rPPRs all present at the same time in the mitoribosome, or are they interchangeable proteins, where each of them would mediate the recruitment of specific transcripts to the mitoribosome.
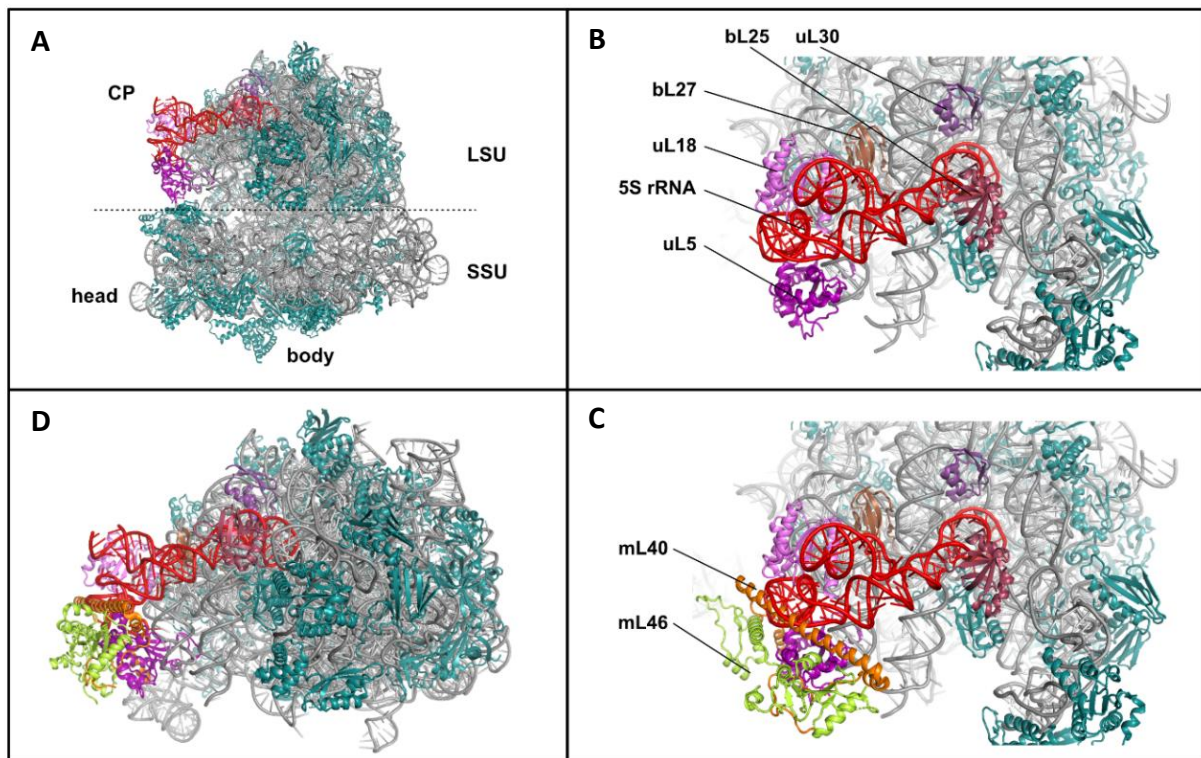
Figure 30: Arabidopsis mitochondrial central protuberance (CP) proteins compared with *E. coli* CP structure

A 3D model of the *E.coli* ribosome (4YBB). rRNA are represented in gray and r-proteins in blue. The CP components, 5S rRNA and its r-proteins, are represented in different color shades.

B Enlargement of the CP. All the components highlighted are present in the Arabidopsis mitoribosome.

C Same representation as **B** but yeast mL40 and mL46 are also represented. mL46 does not clash with any components of the structure. In this model a minor part of mL40 clashes with the 5S rRNA absent from yeast mitoribosome. In Arabidopsis mitoribosome, the CP must have a different architecture as mL40, mL46 and the 5S rRNA all occur.

D Overall view of the LSU with mL40 and mL46 stabilizing the CP by locking uL5 and the 5S rRNA.

From the results obtained, as well as the comparison of the mitochondrial r-proteins abundance with plant specific mitochondrial r-proteins, no definitive conclusion can be drawn. In all cases, the number of plant rPPRs is not sufficient to have a specific rPPR involved in e.g. the recruitment of an individual mRNA to the mitoribosome, as the number of encoded mRNAs exceeds by far the number of rPPRs identified. In trypanosoma the seven PPRs identified mainly mediate protein-protein interactions while only some of them have peripheral interactions with rRNAs. In this case it is suspected that these proteins were originally recruited to bind single stranded rRNA segments, but that they conserved their positions in the structure while the rRNAs were reduced.

One interesting case in the mitochondrial ribosomes divergence is the LSU central protuberance (CP). In prokaryotes and in the cytosolic ribosomes of eukaryotes, the CP host the 5S rRNA. In mammals the 5S rRNA was replaced by a tRNA, termed CP-tRNA, in yeast the 5S rRNA was lost and an extension of the 21S rRNA structurally compensate its loss (Desai et al., 2017; Greber et al., 2015). In trypanosoma the 5S was also lost, but in this case the CP is entirely shaped by proteins (Ramrath et al., 2018). The situation in plant is different with the occurrence of 5S rRNA. In the other organisms, this modification of the CP was accompanied by the loss of some of the original bacterial r-proteins and the acquisition of novel mitochondria specific r-proteins. In *E.coli* the 5S rRNA is mainly stabilized by the proteins uL5, uL18, bL25, bL27 and uL30 (Fig 30). Among those five proteins, bL27 and uL30 are conserved in all mitoribosomes described to date, but bL25 is absent from all but plants, uL5 is absent from trypanosoma and mammals and uL18 is absent from yeast and trypanosoma. Hence, in Arabidopsis the original core of CP proteins is conserved as well as the 5S rRNA. In plant, mL40 and mL46 are present, which are mitochondria specific CP proteins found in yeast, mammals and trypanosoma. As shown in Fig 30 C these proteins could theoretically be present even with the 5S rRNA present. These proteins might constitute another layer of protection for the CP by sheltering the uL5 protein.

While the functions of the additional non-PPR plant specific proteins remains elusive, functional analysis has been initiated for a few proteins, in particular mS83, mS84 as well as mL105 as described below. Among the plant specific r-proteins, mS83 represents an interesting case. Two different genes, At4g15640 and At3g21465, encode almost the same proteins (73.3% amino acids identity). These two proteins were annotated as "adenylyl cyclase", which is the enzyme responsible for the conversion of ATP to cAMP and pyrophosphate. I further looked into the features of these proteins, because nothing would point toward an adenylyl cyclase function as part of plant mitoribosomes. An analysis using InterPro revealed that the proteins were in fact
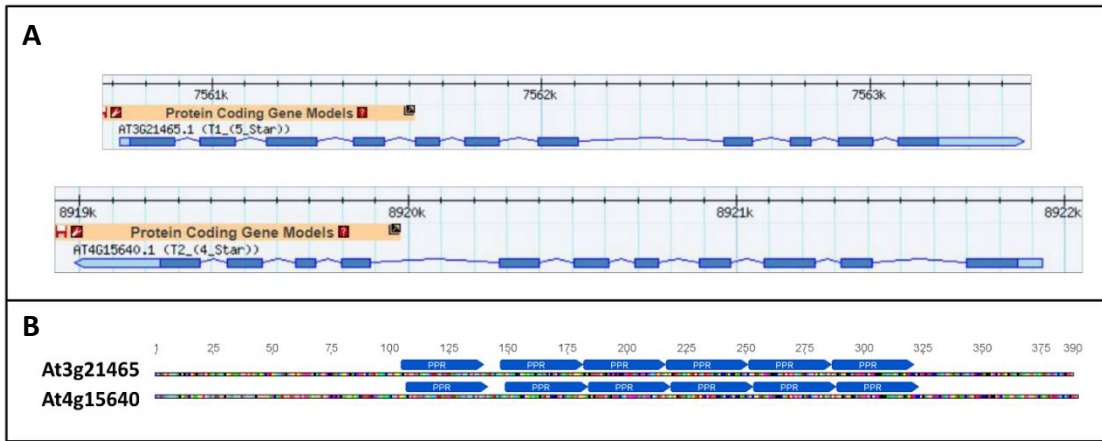
Figure 31: mS83, "adenylyl cyclases" proteins
A Gene organization of At3g21465 and At4g15640, both encoding for a mS83 r-protein. Both genes have 10 introns and 11 exons.
B Schematic representation of the two mS83 proteins. They each harbor six PPR-like motifs and share 73.3% amino acids identity.
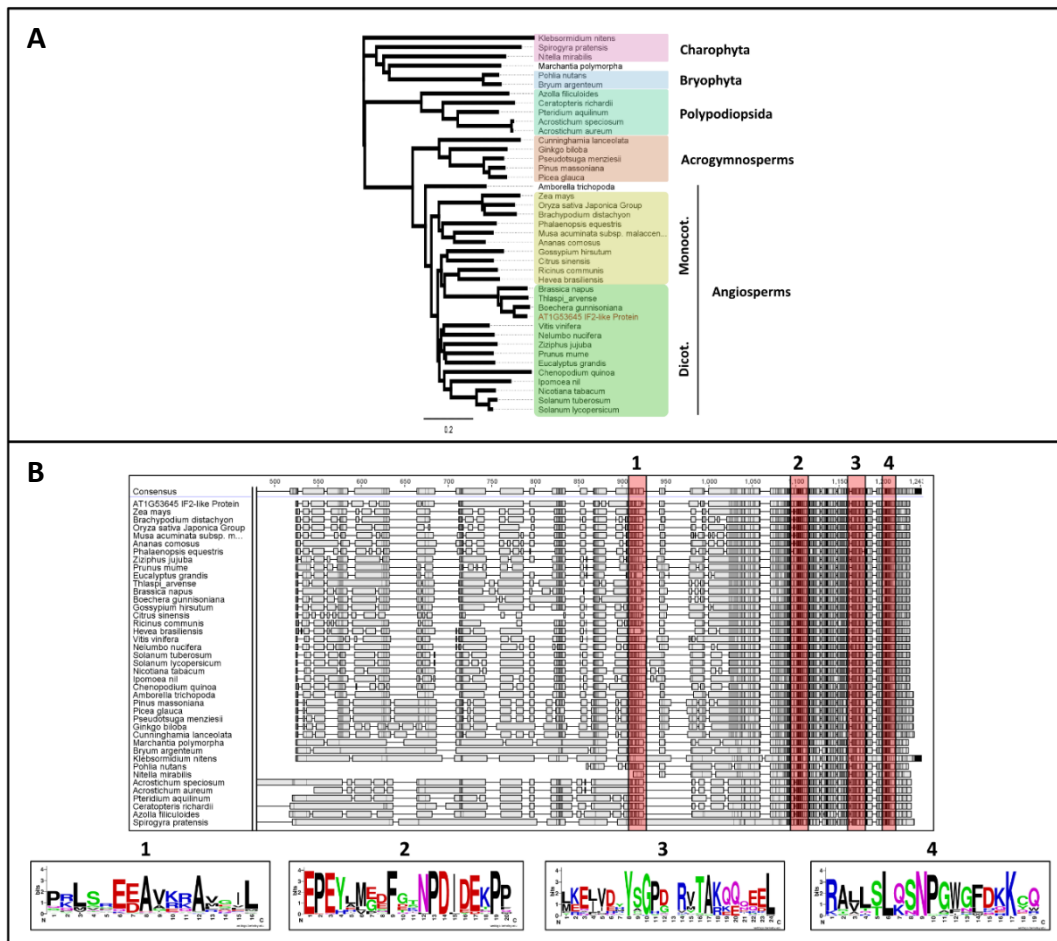
Figure 32: mS84, an "IF2-like" protein
A Phylogenic repartition of the **mS84** r-protein. The protein is highly conserved in the green lineage, down to bryophyta and is even found in charophyta (green algeae)
B Protein sequence alignment for all the organisms used in the phylogeny presented in **A**. For all the proteins, the C-terminal part is particularly conserved. **1**, **2**, **3** and **4** are highly conserved motifs of unknown functions for which WebLogo sequence representations were generated.

part of the "Tetratricopeptide-like helical domain superfamily" (ID:IPR011990). Therefore a TPRpred analysis was performed and revealed that both contained 6 PPR-like motifs, thus making them putative PPR-like proteins (Fig 31). However, the two genes both have ten introns, which is very unusual for PPR proteins (Lurin et al., 2004), for instance among the rPPRs, two genes have one intron and the rest have none. Hence these proteins might represent a novel class of PPR-like proteins. The reason why these proteins were initially annotated as "adenylyl cyclase" when they obviously are mainly composed of PPR-like motifs remains uncertain and their function unknown. Still, mutant analysis suggest that the two mS83 proteins might have redundant functions.

In the case of mS84, this protein was one of the non rPPRs proteins that was suspected at an early stage of the project to be also part of the mitoribosome as it was consistently found as one of the most enriched protein in both IPs and purified ribosomes. The protein is encoded by At1g53645 and is described as "Uncharacterized protein" or "hydroxyproline-rich glycoprotein family protein" in Arabidopsis databases. Blast analyses revealed that it is sometimes annotated as "IF2-like" in other species from the green lineage even though no specific protein domain corresponding to initiation factors could be identified. The protein is rather long compared to other r-proteins (523 aa as compared to 200 aa on average). The phylogeny of this protein was performed and it revealed that is was highly conserved in the green lineage, even down to bryophytes and green algae, with several domains at its C-terminus being strongly conserved (Fig 32). The function of the protein is yet unknown but its high conservation and the lethality of mS84 mutants indicate an essential function, most probably acquired early in the evolution of plant mitoribosomes. mS84 essential function and location as part of the SSU might indicate an important plant specific activity for translation initiation.

To which degree is the plant mitoribosome bound to membrane?

The biochemical characterization of plant mitoribosomes performed here has suggested that they are attached to membranes, i.e. because solubilization with non-ionic detergent was compulsory to obtain soluble monosomes. In mammals and yeast, the mitoribosomes were shown to be attached to the IMM by the LSU, where the peptide exit tunnel is located. In both cases, the attachment seems to be mediated by the r-protein mL45 and in yeast an rRNA fragment, the expansion segment of helix 96, is also involved in membrane attachment (Fig 16) (Ott et al., 2016; Pfeffer et al., 2015). Again, in both mammals and yeast, the nascent protein is thought to be co-translationally inserted into the inner membrane by the insertase Oxa1, thus constituting the central component of IMM protein insertion machinery. In Arabidopsis, no homolog of mL45 was

identified in the final list of mitoribosome core proteins. Interestingly mL45 is also absent from trypanosoma, suggesting that in both cases the attachment to the membrane is mediated differently. Alternatively, the attachment of mitoribosomes to the membrane might only be transient / or partial in plant and trypanosoma, as the number of non-membrane proteins encoded in their mt-genomes is greater compared to mammals and yeast. Interestingly, an analysis of sequence similarities revealed that mL45 has similarities with At5g27395. This protein is described as a TIM44-related protein, in accordance with the fact that mL45 is a TIM-related protein. This protein is not found in purified ribosomes but is found in co-IP. At5g27395 did not pass the drastic statistical criteria used here in co-IPs, but its values were just under the threshold. In addition, mL105 one of the two non-rPPR plant specific proteins identified in the LSU of the Arabidopsis mitoribosome, is annotated and has similarities with a "protein translocase". It is thus tempting to speculate that At5g27395 and /or mL105 might be involved in the attachment of plant mitoribosomes to the inner mitochondrial membrane. Future functional analyses will reveal if this is indeed the case. Furthermore, the investigation of plant Oxa1 protein (encoded by At5g62050) protein partners, e.g. by a co-IP approach could also be performed and reveal the mechanism by which mitochondrial encoded proteins are inserted in membranes in plants.

Functions of the PPR336/rPPR1 and 336L proteins

As described in the Waltz et al manuscript, PPR336, renamed rPPR1, is part of the Arabidopsis mitoribosome and its deletion globally impedes mitochondrial translation. rPPR1 has a close homolog, PPR336L, with which the double mutant *ppr336/ppr336L* was constructed. The double mutant has a clear delayed growth, contrary to the two single mutants which only have a slight phenotype, as shown by an in depth growth monitoring performed at the Phenoscope platform (IJPB, INRA, Versailles). Therefore it was expected that the two proteins would have redundant functions. However, results presented here clearly showed that PPR336L is not found in the mitoribosome.

Nevertheless, the molecular analysis of the *ppr336L* mutant by ribosome profiling (not shown here) suggested that PPR336L function is required for the translation of some mRNAs, i.e. nad2, nad4, nad7 and rps7 mRNAs. As a next step, mitochondrial mRNAs splicing efficiency was also analyzed by the laboratory of Hakim Mireau (IJPB, INRA, Versailles), for both *ppr336* and *ppr336L* mutants. Interestingly, it appeared that in *ppr336* the splicing of cox2 is affected whereas in *ppr336L* the splicing of the first intron of nad2 is affected. Hence both proteins could be involved in the splicing of two distinct transcripts. As PPR336L only appeared recently through gene

duplication in Brassicaceae, an explication could be that, the ancestral form of PPR336 was a protein supporting both functions in translation and splicing. But after gene duplication, PPR336 and PPR336L functions diverged to become more specialized, with PPR336 being the only one to be an integral component of the mitoribosome.

A similar case was observed for rPPR8 (At5g15980) which also has a close homolog resulting from a gene duplication specific to Brassicaceae. The gene was discovered by blast and was termed rPPR8L (At3g02490). The two proteins also share 70% identity in amino acids, and rPPR8L is also not found in the mitoribosome, similarly to PPR336L. Both mutants have no macroscopic phenotype. Double mutants were created but the selection process to obtain homozygous plants has not yet been performed. Here again the two proteins could have diverged to fulfill specific functions.

Perspectives of this work

As a perspective to this work, the short term goal will be to obtain a higher resolution structure of the plant mitoribosome, especially of the SSU which is the most divergent from the other mitoribosomes. For this, the structural analysis performed in collaboration with Dr. Yaser Hashem will be continued. The main problem that hampered the acquisition of a high-resolution structure, is the fact that mitoribosomes were packed in aggregates, thus only a few were individualized and accessible for 2D classification. This should be resolved by adjusting the final buffer conditions. Additionally, the work on crosslink-LC-MS/MS that was initiated with the Esplanade Proteomic Platform will be pursued. This technique is a structural mass spectrometry method that delivers tertiary structural information on proteins, but also on protein interaction networks in complexes, both *in vitro* and *in vivo*. In the case of mitoribosomes, it was used by the Ban lab as a complementary method to draw the interaction network of the mammalian mitoribosome proteins, hence confirming the results derived from the cryo-EM data (Greber et al., 2015). The structure and crosslink data will be used to confirm the result presented in the Waltz et al manuscript but most importantly will be essential to better understand the functions of the different plant specific mitochondrial r-proteins.

Another important perspective will be to understand how the plant mitoribosome is bound to the membrane. The plant mitoribosome purification results clearly showed an association with mitochondrial membrane fractions, but how this tethering is mediated remains to be identified. Is the plant mitoribosome always attached to the inner membrane, or not? A homolog of the protein mediating the association of the animal and yeast mitoribosome with the IMM, mL45, was found by sequence analyses, but this protein is only found in co-IP results and not in pure monosomes. Therefore direct association of mitoribosomes with membranes might be mediated by other protein(s).

Additionally, it is clear from the results, that all the proteins that were identified in the Waltz et al manuscript are mitochondrial proteins. Nevertheless these proteins could also be localized elsewhere in the cell. For example, it was shown that rPPR9 (PNM1) is also localized in the nucleus, the same goes for rPPR5 (At2g37230) that was shown to be localized to both chloroplast and mitochondria (Hammani et al., 2011; Wang et al., 2016). Given these information, I wanted to know if the other rPPRs identified could localize in other compartment of the cell. Additional localization could indicate that the protein might play a role in retrograde signaling, for example, rPPR9 was proposed to be involved in crosstalk between mitochondria and the nucleus.

Hence, the cloning of all the different rPPR genes in plasmids allowing their expression *in planta* has been initiated. The plasmid used allow the expression of the proteins with the fusion of a GFP-tag at their C-terminus and the genes are placed under the control of a 35S promoter. Currently, WT-plants were transformed with constructs relevant to study the localizations of rPPR1, rPPR2, rPPR3a, rPPR4, rPPR6 and rPPR7. Additionally GFP fusions with PPR336L and mS84 have been obtained and will be used for plant transformation. This analysis should give clues to understand how the plant mitochondrial translation apparatus is involved in functional cross-talks with other cell compartments.

Altogether, results obtained during my PhD work contribute to understand plant mitochondrial translation, and mitochondrial translation in general. Still, major questions remain, especially the mechanism by which plant mitochondrial translation is initiated remains completely mysterious. How is the mRNA recruited to the mitoribosomes and the AUG specified without the SD-sequence? It is likely that some of the plant specific r-proteins, especially rPPRs, are major players of this process. Other results obtained during this work suggest connections between the translation machinery and other gene expression processes. For instance, many factors related to gene expression were found in the rPPR1-HA co-immuno-precipitation. Among them, MNU2 and the PNPase were even found among the significantly enriched proteins. This suggests that the maturation and/or degradation of transcripts could be performed co-translationally in plant mitochondria.

Concluding remarks

The rapidly increasing number of studies on mitochondrial ribosomes has revealed that mitoribosomes diverged significantly in the respective eukaryote groups. Plants make no exception. The plant mitoribosome is also significantly different from the other mitoribosomes described to date. The evolutive drift of mitochondria between the different groups of eukaryotes, allowed the evolution of highly divergent mitoribosomes that nevertheless all perform the same core function, i.e. the translation of hydrophobic respiratory proteins. The comparison of mitochondrial translation systems thus constitutes a very nice model to study the evolution of systems toward a very specific function, as does the evolution of the respiratory chain (Sloan et al., 2018b). This is particularly interesting because, even though the ribosome is a crucial component of gene expression which is usually particularly conversed, the evolution has been tinkering a lot with mitoribosomes. Despite the huge recent interest for mitoribosomes, only a few of them have been characterized in details. Therefore, to get a bigger picture of the evolutive tinkering that operated on mitoribosomes, it will be necessary to characterize mitochondrial ribosomes from species representing unexplored eukaryote groups. For example, it will be relevant to characterize mitoribosomes in the apicomplexa *P.falciparum* where the mitochondrial genome is extremely reduced and where rRNAs are fragmented, which is also the case in *C.reinhardtii*. The worm *R.culicivorax* has the smallest tRNAs described to date, how does its mitochondrial ribosomes deal with them? On the other side, the jakobid *R.americana* has mitochondria that seemingly closely resemble the original endosymbiont, therefore its mitoribosomes should be more bacterial-like as compared to the other eukaryotes. Finally in plants, as described in the Waltz et al. manuscript, the specific features observed in Arabidopsis, i.e. the large 18S rRNA additional domain and several r-proteins, are only predicted to occur in Angiosperms. Therefore mitoribosomes should also be different in other major groups of the green lineage such as ferns, mosses or gymnosperms.

My work showed that plant mitoribosomes are undoubtedly significantly divergent from both prokaryote ribosomes and from other mitoribosomes. Still, major questions remain and many important processes have to be understood to fully comprehend mitochondrial translation. For instance, in plants but also in most other eukaryotes, the questions of translation initiation and regulation need to be addressed. Likewise, how mitoribosomes evolved will be fascinating to investigate. In this context, the upcoming years will definitely be full of exciting discoveries that will reveal how evolution played with mitoribosomes to optimize mitochondrial protein synthesis.

Materials and Methods
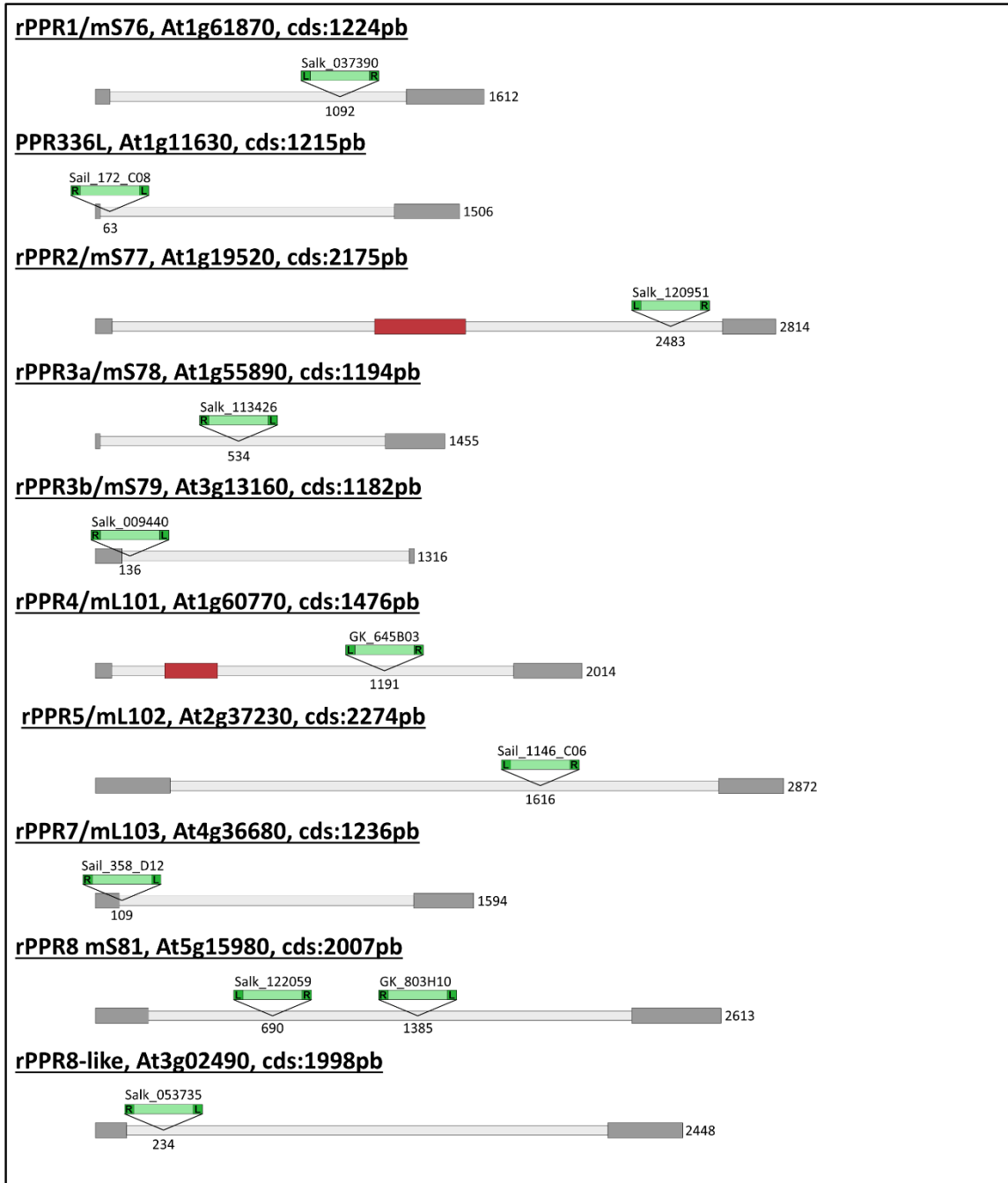
Figure 33: Schematic representation of the mutant identified in this study
Schematic representation of protein-coding genes used in this study. UTRs are represented by gray boxes, introns by red boxes, cds are displayed in light gray, T-DNA are shown in green shades.

Materials and Methods

Materials

Plant lines

Arabidopsis thaliana Col-0 plants were obtained from the INRA Stock Centre in Versailles (http://dbsgap.versailles.inra.fr/portail/). Plants were grown on soil under long-day conditions (16 h of light and 8 h of dark). Mutant plants were obtained from SALK T-DNA, SAIL T-DNA, GABI-Kat and WiscDsLox insertion lines collections (O'Malley et al., 2015). T-DNA insertion lines were selected and retained if the insertion was located in an exon of the gene and the insertion were confirmed by PCR amplification and sequencing. For all the mutants, genotyping primers were designed using the T-DNA Primer Design software available on the SIGnAL website (http://signal.salk.edu/tdnaprimers.2.html)

PPR mutants

For each gene of interest several mutants were ordered, if possible. The insertions lines listed here are the ones that were ultimately retained and described in the publication by Waltz et al. A graphical view of the genes and the insertion sites are presented in Fig 33.

The **At1g61870 (rPPR1 or mS76)** T-DNA insertion mutants are Salk_037390 and Salk_139562. For **At1g11630 (336L)** the T-DNA line is Sail_172_C08 (Uyttewaal et al., 2008).

One mutant for **At1g19520 (rPPR2 or mS77)** was previously characterized as nfd5 (Portereiko, 2006). Another line where the insertion is located in exon2 of the gene, Salk_120951, confirmed the lethal phenotype observed in the original publication.

For **At1g55890 (rPPR3a or mS78)**, the T-DNA line is Salk_113426, which has no visible phenotype.

For **At3g13160 (rPPR3b or mS79)**, the T-DNA line is Salk_009440, which is unable to produce seeds.

For **At1g60770 (rPPR4 or mL101)**, the T-DNA line is GK_645B03, which is characterized by a delayed growth.

For **At2g37230 (rPPR5 or mL102)**, the T-DNA line is Sail_1146_C06, which is characterized by a delayed growth.

For **At4g36680 (rPPR7 or mL103),** the T-DNA line is Sail_358_D12, which is characterized by a severe growth delay and dwarfism.

For **At5g15980 (rPPR8 or mS81)**, the T-DNA lines are Salk_122059 and GK_803H10, which has no

mS85, At1g18630, cds:468pb

Salk_030270

1851
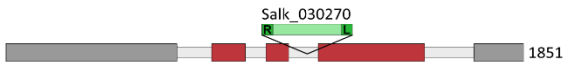
mS84, At1g53645, cds:

Sail_171_D03

2741

mS83, At4g15640, cds:1173pb

WiscDsLox485

3027
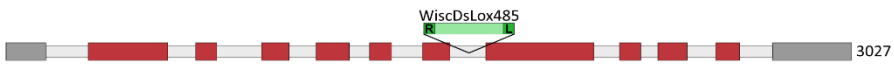
mS83, At3g21465, cds:1167pb

Salk_099373

2785

mS22, At1g64600, cds:1614

Salk_061674

2924

mS47, At4g31810, cds:1230pb

Sail_873_C07

3643

uL15m, At5g64670, cds:846pb

Sail_1291_E02

1827

Figure 33: Schematic representation of the mutant identified in this study - continued

visible phenotype. For **At3g02490 (rPPR8-like)**, the T-DNA line is Salk_053735, which has no visible phenotype. Double mutants using Salk_053735 and either Salk_122059 or GK_803H10 were created by crossing.

For **At5g60960 (rPPR9/PNM1 or mL104),** mutants were previously characterized and are lethal (Hammani et al., 2011).

For the rPPR1-HA plant line, At1g61870 was cloned under the control of its endogenous promoter (1 kb upstream of the initiation codon) into the pGWB1 binary vector, in fusion with a single C-terminal HA-tag. The ***rppr1/At1g11630*** knock out line was transformed by floral dip via *Agrobacterium tumefaciens* GV3101 (Clough and Bent, 1998). Plants were selected for hygromycin resistance and rPPR1-HA expression, as well as macroscopic phenotype complementation.

Non-PPR mutants

Additionally, mutants for the other plant-specific r-proteins of the mitoribosome were ordered and investigated.

For **At1g18630 (GR-RBP6 or mS85)**, the T-DNA line is Salk_030270.

For **At1g53645 (IF2-Like or mS84)**, the T-DNA line is Sail_171_D03.

For the "Adenylyl cyclases" proteins found in mitoribosome and termed **mS83**:

For **At4g15640** the T-DNA line is WiscDsLox485

For **At3g21465** the T-DNA line is Salk_099373.

For **At1g64600 (mS22)**, the T-DNA line is Salk_061674.

For **At4g31810 (mS47),** the T-DNA line is Sail_873_C07, no homozygous plant were obtained and already published data for another allele, Salk_002356, showed that the insertion is lethal (Gipson et al., 2017).

A mrpl15HA plant line was obtained from Dr. Hakim Mireau. The plant is knock-out for the **At5g64670** (Sail_1291_E02) gene encoding the r-protein uL15m, and complemented with an HA tagged uL15m under the control of the native promoter of the gene.

Cell line

An already available cell suspension created from *Arabidopsis thaliana* Col0 was used (Giegé et al., 1998).

Bacterial strains

Escherichia coli

TOP10 strain: Chemically Competent TOP10 cells were routinely used for plasmid amplification. The strain is characterized by mutations in the endA1 gene, inactivating intracellular endonuclease activity, and in recA eliminating homologous recombination, to increase the amount of plasmid DNA produced.

Genotype: *F- mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 Δ lacX74 recA1 araD139 Δ(araleu)7697 galU galK rpsL (StrR) endA1 nupG*

DB3.1 strain: DB3.1 is a HB101 derivative containing the gyrA462 allele which confers the strain resistance to the toxic effects of the ccdB gene, therefore used for propagating Gateway entry and destination vectors.

Genotype: *gyrA462 endA1 Δ(sr1-recA) mcrB mrr hsdS20 glnV44 (=supE44) ara14 galK2 lacY1 proA2 rpsL20 xyl5 leuB6 mtl1*

BL21 (DE3) strain: used for protein production. Chemically Competent BL21 cells, which contain the phage T7 RNA polymerase gene linked to the IPTG-inducible promoter, were used for expression of all plasmids containing the T7 promoter.

Genotype : *F– ompT gal dcm lon hsdSB(rB- mB-) λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])*

Rosetta2 (DE3) pLysS strain: Rosetta2 strains are BL21 derivatives designed to enhance the expression of eukaryotic proteins as it also provides seven tRNA genes encoding rare codons in *E.coli* (AUA, AGG, AGA, CUA, CCC, GGA, CGG) carried by on a compatible chloramphenicol-resistant plasmid.

Genotype: *F- ompT hsdSB(RB- mB-) gal dcm λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5]) pLysSRARE (CamR)*

Agrobacterium tumefaciens

GV3101 (pMP90) is a C58 strain derivative created by curing the pTiC58 Ti plasmid from a Rifampicin resistant C58 strain. Therefore GV3101 is a TiC58 free cured Rif resistant *Agrobacterium tumefaciens* strain, which is thus avirulent. This strain carries the pMP90 plasmid which is a T-DNA/T-region free helper plasmid, used in combination with binary vectors for plant

transformation. The strain is grown under Gentamycin and Rifampicin, a third antibiotic is used for the binary vector (Holsters et al., 1980).

Plasmids

GATEWAY cloning vectors

pDONR207   Entry Gateway® plasmid with the attP1 and attP2 sites and a gentamycin resistance marker and the classical *ccdB* gene.

pGWB1   Destination Gateway® plasmid with the attR1 and attR2 sites, used for *Agrobacterium tumefaciens* plant transformation. The plasmid confers resistance to spectinomycin in bacteria, and kanamycin and hygromycin in the plant. It allows the cloning of a gene under its own promoter, and without any tag.

pGWB14   From the same series of plasmid as pGWB1 but also the cloning of the gene of interest under the control of the 35S promoter and the C-ter fusion with a 3xHA tag

pGWB17   From the same series of plasmid as pGWB1 but also the cloning of the gene of interest under the control of the 35S promoter and the C-ter fusion with a 4xMyc tag

pB7FWG2   Destination Gateway® plasmid with the attR1 and attR2 sites, used for *Agrobacterium tumefaciens* plant transformation. The plasmid confers resistance to spectinomycin in bacteria, and basta in plants. It allows the cloning of a gene under the control of the 35S promoter, and in fusion with a C-ter GFP tag.

Restriction cloning vectors

pET28b   Bacterial protein expression plasmid that allows, by conventional cloning (digestion / ligation), to obtain recombinant proteins with a 6His tag at the N or C-termini or both. The gene is under the control of a T7 promoter; and the vector confers kanamycin resistance.

pART7   pART7 is an entry vector to clone a gene under the control of the 35S promoter, and in fusion with a C-ter GFP tag using restriction cloning. It also contains the octopine synthase (*ocs*) gene (Gleave, 1992).

pBART27   Modified version of pART27, conferring resistance to Basta in plants, instead of kanamycin initially. The p35S: gene: eYFP: *ocs* originally cloned in pART7 is integrated by digestion / ligation using the NotI restriction site. The vector is used for the transformation of plants by *Agrobacterium tumefaciens* (Gleave, 1992).

AGI	Line	Primer ID	Primer Sequence
			Genotyping
At1g61870	Salk_037390	Salk037 LP	GAAAATACCGATCTCGGGAAC
rPPR1/PPR336		Salk037 RP	AGAATCTCGCCGAGAAGAAAC
At1g11630	Sail_172_C08	Sail172 LP	GATCCAAAGAGTCCGGATCTC
PPR336L		Sail172 RP	GATCAAGCATATTCGCTCTGC
At1g19520	Salk_120951	Salk120 LP	GGCAGACACATCACAATTCAC
rPPR2		Salk120 RP	AAGTGCATTGCAAATTTGGTC
At1g55890	Salk_113426	Salk113 LP	TATTAACAGGCGCCATTGAAG
rPPR3a		Salk113 RP	TCAGGACGATAACCATGCTTC
At3g13160	Salk_009440	Salk009 LP	ATTGTAACAACTTCCCCGGAC
rPPR3b		Salk009 RP	CAGTTTCTCTCAGGCATTTCG
At1g60770	GK_645B03	GK645 LP	AATGCAAGTGACCCTATCACG
rPPR4		GK645 RP	TGTTGACAGAGAAAGCGGAAG
At2g37230	Sail_1146_C06	Sail1146 LP	TGTCCAGATCAGCTGTGTGTC
rPPR5		Sail1146 RP	TTGTTTGTGGAGATGAAAGGG
At4g36680	Sail_358_D12	Sail358 LP	TAAAATGCAGGTGCTTTGACC
rPPR7		Sail358 RP	CGTTGAAGGAAACAGCTGATC
	Salk_122059	Salk122 LP	AACAAACAGAAGCAGTGTGGG
At5g15980		Salk122 RP	CATTTCATCGCGATCTTTCTC
rPPR8	GK_803H10	GK803 LP	TTAAGGAGGCTTCTCCTGAGG
		GK803 RP	CCAACCATTTTCTCGAAACAC
At3g02490	Salk_053735	Salk053 LP	AAAAGGGAGTGGGTTTCAGTG
rPPR8L		Salk053 RP	ACACCAAAAATACGCAGCATC
At1g64600	Salk_061674	Salk061 LP	GGAACGCTGAGACGAAGTTC
mS22		Salk061 RP	ATCTCAATCAATGCAGCGTG
At4g31810	Sail_873_C07	Sail873 LP	TTTTGGGCTTGATACCGTTG
mS47		Sail873 RP	TGGAAGAGCCTAGATTTGCTG
At4g15640	WiscDsLox485	WDL485 LP	GTGGTTGTGGCTGTTGATTG
mS83		WDL485 RP	TGGTTTCTTTCAGACAGGGAG
At3g21465	Salk_099373	Salk099 LP	TAGCTGTGTGTGTTGCAGAGG
mS83		Salk099 RP	TATCGACCAAAGAATCGTTGG
At1g53645	Sail_171_D03	Sail171 LP	TCCAAAAACAGGCTTAAACCC
IF2-like or mS84		Sail171 RP	AAAGTAGGCCACCATCTCGTC
At1g18630	Salk_030270	Salk030 LP	ATTTGATTTGGATCCGACATG
GR-RBP6 or mS85		Salk030 RP	AATCCCATTGGTTTCTGAAGC
At5g64670	Sail_1291_E02	Sail1291 LP	TGCACTTGATTTTAAACAAATACG
bL15m		Sail1291 RP	AACACATTAACGTCGGTTTGC
Recombinant At1g61870	336HA	336HA Rev	CATCGTATGGGTATCCTTGCGG
		Sail LB2	GCTTCCTATTATATCTTCCCAAATTACCAATACA
T-DNA left border specific primers		Salk LBb1.3	ATTTTGCCGATTTCGGAA
		Gabi LB1a	ATATTGACCATCATACTCATTGC
		WiscDsLox LB	AACGTCCGCAATGTGTTATTAAGTTGTC
			qPCR
At5g64670		64670 LP qPCR	AATGGCCACTTCATTTCGAG
bL15m		64670 RP qPCR	TCTTCTCACTGATCCTCCTGCT
			p35S::GFP Cloning
At2g37230		CGATEF 37230	AAAAAGCAGGCTTACAATGGCTTTCATTTCCAGA
rPPR5		CGATER 37230	AGAAAGCTGGGTATAGAGAAACATTGTTTTGTTTC
At4g36680		CGATEF 36680	AAAAAGCAGGCTTACAATGGCTTCTTCTCGTATTTCTCTC
rPPR7		CGATER 36680	AGAAAGCTGGGTAAGCAGCAGCAGCCTCCT
At1g55890		CGATEF 55890	AAAAAGCAGGCTTACAATGTCGTCTCTATCTCGCGT
rPPR3a		CGATER 55890	AGAAAGCTGGGTACTCTTCCTGGGACGGC
At3g02650		CGATEF 02650	AAAAAGCAGGCTTACAATGTTGAGATCCTTTCTCTG
rPPR6		CGATER 02650	AGAAAGCTGGGTATGCTTCAGCAAGTAAGTTGCC
At1g60770		CGATEF 60770	AAAAAGCAGGCTTACAATGGCGATGCGACATTTGAG
rPPR4		CGATER 60770	AGAAAGCTGGGTACACATCCTGAGATACTTCATCGAG
At1g19520		CGATEF 19520	AAAAAGCAGGCTTACAATGAAATCTTTCCTCCTCTCTAG
rPPR2		CGATER 19520	AGAAAGCTGGGTACGAGTTAAATCGTCTCACTCT
At1g53645		CGATEF 53645	AAAAAGCAGGCTTACAATGAGGAAGTGCTATAGGAAGAA
IF2-like or mS84		CGATER 53645	AGAAAGCTGGGTATTTGTAGCTTTGTGACACCT
At5g15980		CGATEF 15980	AAAAAGCAGGCTTACAATGAGATATCAACAATGGCGAT
rPPR8		CGATER 15980	AGAAAGCTGGGTAAGCAGAAGCAGCCAAAGGTT
At1g61870		61870-EcoRI_Fw	TAAAGAATTCATGGCCGTTACTCTCTCGAATCC
rPPR1/PPR336		61870-KpnI_Rv	TAAAGGTACCTTTGCGGTAATGCGGCTTC
At1g11630		11630-EcoRI_Fw	TAAAGAATTCATGGCGTTCCTCTTCC
PPR336L		11630-KpnI_Rv	TAAAGGTACCTCTGAGGCAAAGGCAAT
		Adapter attB1	GGGGACAAGTTTGTACAAAAAAGCAGGCT
Used for the 2-step cloning with the GATEWAY		Adapter attB2	GGGGACCACTTTGTACAAGAAAGCTGGGT
		GFP Rv	GCTGAACTTGTGGCCGTTTA
Used to verify the p35S::GFP constructions		35S Fw	ACGCACAATCCCACTATCCTTC
			336HA line
At1g61870	To clone gene and	PPR336 P&G Fw	GACAATAAAATATGTGGTCCTAACATTACTG
rPPR1/PPR336	promoter	PPR336 P&G Rv	CTCCAAAACACATTAGTTCGTGGTA
	add attB 5'	PPR336 P&G attB Fw	GGGGACAAGTTTGTACAAAAAAGCAGGCTTATATGTGGTCCTAACATTACTGAATCT
	add HA +stop	PPR336 HA Stop Rv	CTACGCATAGTCAGGAACATCGTATGGGTATCCTTGCGGTAATGCGGCTTCGA
	add attB 3'	HA Stop attB Rv	GGGGACCACTTTGTACAAAGCTGGGTACTACGCATAGTCAGGAACATCGTA
		SeqL A	TCGCGTTAACGCTAGCATGGATCTC
Used to sequence pDONR207 entry vector		SeqL B	GTAACATCAGAGATTTTGAGACAC
			Cloning into pET28b for protein expression with a C-ter 6xHis
		RcPRORP NcoI F	AATCCATGGCAAGCTCTTTCCGCTT
RcPRORP		RcPRORP NcoI F NoMTS	AATCCATGGTCGAGAGCTTCATCCGTG
		RcPRORP XhoI R	AATCTCGAGCTTTTGCGAATACACAGC
PfPRORP		PfPRORP NcoI F	AATCCATGGTCCCTCTGGTTTACTCT
		PfPRORP XhoI R	AATCTCGAGAATTTTGGAGAAGTCAATGCA
At5g09840		MNU2 NcoI F	AATCCATGGTCTGGTGGGACTTCCTGAGC
MNU2		MNU2 NotI R	AATGCGGCCGCGACTTCTGAATCTTTCAC
		T7 Pro Fw	TAATACGACTCACTATAGGG
Used to sequence pET28b expression vector		T7 Ter Rv	GCTAGTTATTGCTCAGCGG

Table 6: List of primers

Primers / oligonucleotides

The olinucleotides used for the genotyping of plant mutant lines, qPCR, cloning and sequencing are listed in Table 6.

Antibodies

Primary

• The antibody raised in rabbit against the **PPR336/rPPR1** protein was already available in the lab and used at a 1/5000 dilution (Uyttewaal et al., 2008).

• The antibody raised in rabbit against the *A.thaliana* mitochondria encoded **uL16m** r-protein was obtained from the supplier Agrisera and used at a 1/5000 dilution (https://www.anticorps-enligne.fr/supplier/Agrisera/).

• The antibody raised in rabbit against the *A.thaliana* cytosolic **RPS6** cytosolic r-protein was obtained from Lyuba Ryabova and used at a 1/10,000 dilution.

• The antibody raised in rabbit against the wheat protein **NAD9** was already available in the lab and used at a 1/100,000 dilution (Gobert et al., 2010).

• The antibody raised in rabbit against the C-terminal part of the *A.thaliana* **MNU2** protein was produced in the lab during my PhD work and used at a 1/20,000 dilution.

• The antibody raised in mouse against the **HA peptide**, allowing to detect HA-tagged proteins, was obtained from the supplier Sigma-Aldrich and used at a 1/10,000 dilution.

Secondary

Goat anti-rabbit, or goat anti-mouse depending on the primary antibodies, conjugated with horseradish peroxidase (Amersham, UK) were used as secondary antibodies at a 1/10,000 dilution and visualized with enhanced chemi-luminescent reagents (Amersham, UK).

Methods

Nucleic acids analyses

Restriction cloning

In order to insert a DNA fragment of interest into a plasmid, the DNA fragment is first amplified using specific primers containing two different restriction sites, each present only once in the destination vector. Then the DNA fragment and the plasmid are digested using the two different restriction enzymes at the same time. The DNA fragments were digested by restriction enzymes according to the supplier's instructions (ThermoFisher FastDigest). The digested products are purified (NucleoSpin® Gel and PCR clean-up). Before ligation, the vector is dephosphorylated according to the supplier's instructions (ThermoFisher) by FastAP Thermosensitive Alkaline Phosphatase. Finally 2 nmol of dephosphorylated plasmid is mixed with 20 nmol of insert in a reaction volume of 10 μl in presence of T4 DNA ligase (3U per reaction) and 1X T4 DNA Ligase Reaction Buffer (supplied by New England Biolabs). The reaction is carried out for 16 h at 4 ° C. The ligation product is transformed in *E.coli* and positive bacteria are selected by PCR.

GATEWAY cloning

The Gateway® technology is a universal cloning method based on site-specific lambda bacteriophage recombination properties. The first step is to amplify the DNA of interest using specifically designated primers, containing the *attB1* and *attB2* sites respectively at 5' and 3' of the gene. Once the *attB* sites are added, the first Gateway® reaction can be carried out using the BP Clonase® II enzyme (ThermoFisher). This reaction facilitates the recombination of the *attB*-PCR product with the vector containing the *attP* sites. The product is an entry clone containing the gene flanked by the *attL* sites. During my work, I only used the pDONR207 plasmid as the entry vector. Reaction using the Gateway® LR Clonase® II enzyme mix then facilitates recombination of the entry clone possessing the *attL* sites with any destination vector containing the *attR* sites.

Heat shock bacterial transformation

50 μl of previously prepared thermo-competent cells are mixed with 100 ng to 1 μg of plasmid or ligation product, which is first incubated for 30 min on ice.

In the case of ***E.coli*** transformation: The mixture is subjected to a heat shock during 40 sec at 42°C allowing the entry of the plasmid into the bacteria. The mix is then placed on ice for 2 min. 400 μL of antibiotic-free LB medium is added to the mixture which will be incubated for 1 hour at 37 °C

under shaking to allow regeneration of the cell wall and the completion of a cell division cycle. The totality of the mix is then plated on selective medium.

In the case of **A.tumefaciens**: the heat choc is performed at 37°C for 5 min and the mix is placed on ice for 2 min. 400 µL of antibiotic-free LB medium is added to the mixture which will be incubated for 2 hours at 28 °C under shaking. The totality of the mix is then plated on selective medium.

DNA amplification by PCR

PCR (Polymerase Chain Reaction) is used to quickly obtain a large and exploitable amount of a specific segment of DNA. The principle is based on a succession of replication reactions of a double-stranded DNA template. Each reaction uses two oligonucleotide primers whose 3' ends are directed toward each other. The primers then define the sequence to be amplified. The reaction itself is carried out in a thermal cycler which allows the automation of the PCR reaction.

In the laboratory two enzymes are used:
• **GoTag® G2 Flexi DNA Polymerase**: it was used for routine DNA amplification: bacterial or plant genotyping (Promega supplier).

• **Phusion® High Fidelity DNA Polymerase**: it was used for its high fidelity to clone DNA molecules intended for protein expression for example. Its high fidelity is conferred by an exonuclease activity 3' to 5' of proofreading / correction (New England Biotechnologies).

Agarose gel electrophoresis

Agarose gel DNA analysis is used to separate DNA molecules based on their molecular weight. The DNA containing solution is supplemented with loading buffer and loaded onto an agarose gel (0.8 to 2% (w/v) in 0.5X TAE buffer) to which 0.5 µg/mL of ethidium bromide has been added. The electrophoresis is carried out in a 0.5X **TAE buffer** at a voltage of 100 V. Ethidium bromide is an intercalating agent used as a fluorescent tag, which allows to observe DNA molecules under UV.

- **TAE buffer :** 40 mM Tris-acetate pH 8, 1 mM EDTA

Phenol/chloroform nucleic acids extraction

One volume of pH 6.8 (for RNA and pH 8 for DNA) Tris-HCl saturated phenol / chloroform (50/50 v/v) is combined with an equal volume of nucleic acid sample. Chloroform mixed with phenol allows a clear separation between the aqueous and organic phases. The solution is homogenized using a vortex, which forces the phenol into the water layer thus forming an emulsion. The phases

are then separated by centrifugation for 5 min at 16,000 g. Proteins are denatured and arranged at the interphase. The nucleic acids remain in the aqueous phase where they are recovered.

Ethanol precipitation

For nucleic acids precipitation, starting from one volume of nucleic acid sample, 1/10th of 3M sodium acetate pH 5.2 is added, followed by three volumes of 100% ethanol. Glycogen can also be added, to a final 0.05-1 µg/µL concentration, which will significantly increase the recovery of nucleic acids. The mixture is homogenized and placed 20 min at -80°C, or overnight at -20°C, and the DNA is pelleted by 30 min centrifugation at 16,000 g at 4 ° C. The DNA pellet is then washed with 70% ethanol to remove the residual salts. The ethanol is removed and the nucleic acid pellet is dried before being resuspended in a suitable aqueous buffer.

Plasmid purification :

Plasmid purification from bacteria is based on the principle of alkaline lysis. In the lab it is performed using the NucleoSpin® Plasmid QuickPure kit (Macherey-Nagel). Bacteria are first pelleted and then resuspended in a TE buffer and containing RNases. The cells are lysed with a solution containing SDS and NaOH. A third buffer, containing sodium acetate, neutralizes the lysate. Hence, the plasmid DNA is renatured, while the genomic DNA is precipitated with cell debris and eliminated. The supernatant containing the plasmid DNA is loaded onto a silica column. Salts and other soluble components are washed away using a 70% ethanol. The plasmid DNA is finally eluted from the column with ultra-pure water.

Quantification :

To quantify nucleic acids a Thermo Scientific™ NanoDrop 2000c™ spectrophotometer was used.

Sanger DNA sequencing :

In order to ensure that the DNA templates do not contain any mutation, the clones are all verified by sequencing. Sequencing is conducted at the IBMP DNA Sequencing platform directed by Abdelmalek ALIOUA. Sequencing is based on the asymmetric amplification method of Sanger (Sanger et al., 1977). The nucleic acids are then separated and analyzed by capillary electrophoresis using an Applied Biosystems 3100 (Perkin Elmer) apparatus.

Rapid plant DNA extraction for genotyping

The equivalent of 100 µL of glass beads and 400 µL of **extraction buffer** are placed in a 2 mL screwable tube. Leaves of about 1 cm² are collected per plant to genotype and placed in the same

tube. For cell disruption, the Precellys® device was used using two cycles of 30 sec each at 5500 rpm. Large cell debris are pelleted at 16,000 g for 5 min at 4°C. 200 µL of supernatant is then transferred to a new Eppendorf tube and 150 µL of cold isopropanol is added. The mixture is vortexed and incubated for 5 min at room temperature. The DNA is pelleted at 16,000 g for 15 min at 4°C. The pellet is wash with 1 mL of 70% ethanol and the pelleted is finally dried and resuspended in the appropriate volume of ultra-pure water. For genotyping, all primers were designed so that only one PCR-program is sufficient for all genotyping experiments.

- **Extraction buffer:** 200 mM Tris-HCl pH 7.5, 250 mM NaCl, 25 mM EDTA

Plant total RNA extraction

For RNA extraction, TRIzol® Reagent was used. The samples (seedlings, flowers or leaves) are collected and immediately flash frozen in liquid nitrogen. The frozen samples are then grinded, either manually using a pestle and a mortar, or mechanically using a mixing device such as the Silamat. In a 1.5 mL Eppendorf tube, 1 mL of TRIzol® is added per 100 mg of grinded sample. The mixture is homogenized by vortex and the sample is then incubated 5 min at room temperature. 200 µl of chloroform per 1 mL of TRIzol® is added and the sample is vigorously vortexed. The sample is then centrifuged at 15,000 g for 15 min at 4°C. The upper aqueous phase contains the RNA exclusively and the DNA and proteins are located in the interphase and the bottom organic phase. The aqueous phase is transferred to a new tube and 1.5 volume of isopropanol is added. Similarly to DNA precipitation, glycogen can also be added to increase recovery. The sample is homogenized and incubated at room temperature for 10 min. The RNA is pelleted by centrifugation at 16,000 g for 15 min at 4°C. Supernatant is discarded and the RNA pellet is washed with 70% ethanol. The pellet is air-dried and resuspended in the adequate volume of ultra-pure water.

RT-qPCR

cDNA synthesis

Starting from purified RNA, the samples are first treated with DNase according to the supplier's instructions (DNase I Thermo-scientific®). The DNase treated samples are then subjected to retro-transcription to synthesize the cDNAs. The RT reaction is performed using the same initial amount of DNase treated RNAs in order to be comparable at the end (usually 1-2 µg of DNase treated RNAs). The RT reaction is performed using the SuperScript IV RT according to the supplier's instructions (Thermo-scientific®) and using a mix of both oligo dT primers and random hexamer primers.

qPCR

This technique is used to measure the relative level of expression of a given gene by measuring the amount of a target RNA via its cDNA (see above for total cDNA synthesis). The reactions are carried out in 384-well optical reaction plates. Each well can contain a reaction mixture of 10 µl containing 1X SYBR Green Master Mix (ThermoFisher), 250 nM of both sense and antisense primers and 1 µl of cDNA. The primers were designed using the Universal ProbeLibrary available on the Roche website (https://lifescience.roche.com/en_fr/brands/universal-probe-library.html#assay-design-center). All reactions are conducted in technical triplicate and are performed by the LightCycler® 480 Instrument II (Roche). The PCR begins with an initial denaturation step of 5 min at 95 ° C, followed by 40 cycles of 10 sec at 95°C, 15 sec at 60°C and 15 sec at 72°C. At each cycle, the amount of product is measured, allowing to measure the Ct (crossing threshold), which is the point at which the fluorescent signal is significantly greater than the background noise.

The results obtained are normalized using genes representing endogenous controls (Tip41, Act2 and GAPDH). The PCR efficiencies are calculated by the LinRegPCR program and only PCR efficiencies above 0.9 (1 being the maximum) are retained. Finally the relative DNA quantity is determined using an Excel table.

Polyacrylamide gel electrophoresis:

This procedure is used for the separation of RNA of small sizes or small quantities, or applications where higher resolution, which cannot be achieved with agarose gel electrophoresis, are required. Around 250 ng of RNA are mixed or resuspended in **loading buffer** and denatured 5 min at 70°C before being placed immediately on ice. The samples are loaded on the **gel** (of adequate concentration depending on the size of the nucleic acids of interest) and the nucleic acids are separated under constant amperage of 25 mA for 1 h. The migration takes place in **TBE buffer.**

- **Loading buffer :** 40% (v/v) formamide, 10 mM EDTA, 0.025% bromophenol blue, 0.025% xylene cyanol
- **X% polyacrylamide gel :** x% (v/v) acrylamide/bisacrylamide 19/1, 7 M urea, 1X TBE
- **TBE buffer :** 90 mM Tris-HCl pH 8, 2 mM EDTA, 90 mM borate

Proteins analysis

Total protein extraction :

Crude

For rapid analysis of total proteins, the samples (seedlings, flowers or leaves) are collected and immediately flash frozen in liquid nitrogen. The frozen samples are then grinded mechanically using a mixing device such as the Silamat. 200 µL of hot **protein extraction buffer** is added and strongly vortexed. The samples are then centrifuged 5 min at 16,000 g and the supernatant is retrieved. Proteins are then precipitated using acetone. 3 volumes of cold acetone is added and the samples are mixed before being placed 20 min at -20°C. Finally the samples are centrifuged 10 min at 10,000 g, 4°C. The protein pellets are washed with 80% acetone before being dried and finally resuspended in adequate volume of **protein extraction buffer.**

- **Protein extraction buffer :** 100 mM Tris-HCl pH 6.8, 1% glycerol, 2% SDS, 100 mM DTT, 4 M urea.

From TRIzol extraction

Proteins can also be purified during TRIzol extraction in parallel to RNA extraction. After phase separation the organic phase is retrieved. First DNA is precipitated by adding one volume of 100% ethanol and mixing. The mixture is incubated for 15min at room temperature, before being centrifuged 5 min at 16,000 g at 4°C. The supernatant is saved and proteins are precipitated by acetone as previously described.

Protein quantification

Bradford

Proteins are dosed by the Bradford method using Bio-Rad Protein Assay solution. In a total volume of 1 mL, 10 µL of sample to be assayed is added to 790 µL of water. 200 µl of reagent is added, the mixture is homogenized and after two minutes the OD_{595nm} is measured. The OD value is compared to a previously established standard range and the protein concentration value can be determined.

Nanodrop

To quantify proteins samples, a Thermo Scientific™ NanoDrop 2000c™ spectrophotometer was used.

Protein expression

Induction test

When expressing proteins from bacteria, prior to protein purification, induction tests are performed to confirm that the protein of interest is indeed expressed. For this, 3 ml of LB medium supplemented with the correct antibiotic as well as 1% glucose is inoculated with the candidate colony for expression. During this first incubation, glucose repress the expression of the T7 polymerase and the protein is not expressed. The culture is incubated at 37 ° C until OD reaches 0.7. The culture is split in two, and the two bacterial suspensions are centrifuged for 8 min at 4000 g. One of the bacterial pellets is resuspended in 3 mL of LB plus antiobiotic and glucose, which constitute the uninduced fraction. The other fraction is resuspended in 3 mL of LB plus antiobiotic containing 1 mM IPTG, allowing the expression of the T7 polymerase and the transcription of the gene of interest. The cultures are then incubated under stirring at 17 °C for 18 h. After incubation 100 μL of bacterial solution is sampled and centrifuged for 2 min at 16,000 g. The pellet is resuspended to a total volume of 20 μL using **loading buffer** and **Laemmli buffer.** The bacteria are thus lysed and the total proteins are analyzed by SDS-PAGE. The protein of interest is then visualized on the stained gel or immuno-detected by western blot analysis (see below).

Protein purification

Protein purified during this study were expressed in fusion with a 6xHis-tag. The purifications were performed in two steps, first by affinity to the 6His-tag and then by gel filtration.

First, the induced bacteria are disrupted using the "French Press". The resulting disrupting cell fraction is clarified by centrifugation (18.000 g, 15 min, 4°C) and the supernatant is passed through a 0.2 μm filter and the pH is adjusted to 7.8. The protein extract is then placed in the presence of 200 μL of Ni-NTA (Nickel-NitriloTricetic Acid resin) beads (Qiagen). The mixture is incubated 2 to 16 h at 4°C on a rotating wheel. It is then transferred to a column, previously equilibrated with 1 mL column of **LB$_{pET}$**. The column is then washed with 5 mL of **WB1$_{pET}$** followed by 5 mL of **WB2$_{pET}$** to remove non-specifically bound proteins. The proteins fixed on the Ni-NTA beads are then eluted with 2 mL of **EL1$_{pET}$** followed 2 mL of **EL2$_{pET}$**. The different fractions are analyzed by SDS-PAGE and the fractions containing the highest amount and/or purest protein of interest are selected.

The last purification step is based on a gel filtration allowing the separation of the molecules according to their size. The latter is performed on a Superdex 200 10/300 Increase column, and was performed using the automated Äkta pure system.

- **LB_{pET}** : 20 mM MOPS pH 7.8, 500 mM NaCl, 15% (v/v) glycerol, 0.2 mM DTT, 1/175 protease inhibitor, 50 mM imidazole

- **WB1_{pET}** : 20 mM MOPS pH 7.8, 150 mM NaCl, 15% (v/v) glycerol, 50 mM imidazole

- **WB2_{pET}** : 20 mM MOPS pH 7.8, 250 mM NaCl, 15% (v/v) glycerol, 75 mM imidazole

- **EL1_{pET}** : 20 mM MOPS pH 7.4, 250 mM NaCl, 15% (v/v) glycerol, 200 mM imidazole

- **EL2_{pET}** : 20 mM MOPS pH 7.4, 250 mM NaCl, 15% (v/v) glycerol, 500 mM imidazole

Protein co-immunoprecipitation

Immuno-precipitations were performed with protein extracts using the µMACS HA-tagged Protein Isolation Kit (Miltenyi Biotec). Mitochondria corresponding to 1 mg of proteins were lysed in **lysis buffer**, 30 min at 4°C on a rotating wheel. The lysate was clarified at 10,000 g, 10 min at 4°C and supernatant was kept and supplemented with 50 µl of anti-HA magnetic beads. The mix was incubated 30 min on a rotating wheel and loaded on the column. After loading the column was washed six times with 200 µl of **wash buffer**, and elution was performed with 120 µl of 90°C elution buffer (Miltenyi Biotec) with the magnetic beads. The resulting Co-IP proteins were analyzed by LC-MS/MS.

- **Lysis buffer:** 20 mM HEPES-KOH pH 7.6, 100 to 800 mM KCl, 30 mM $MgCl_2$, 1 mM DTT, 0.5 mg/ml heparin, 1% Triton X-100 supplemented with proteases inhibitors (Complete EDTA-free)

- **Wash buffer**: 20 mM HEPES-KOH pH 7.6, 100 to 800 mM KCl, 30 mM $MgCl_2$, 1 mM DTT, 0,1% Triton X-100 supplemented with proteases inhibitors (Complete EDTA-free)

Sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE)

Before performing SDS-PAGE analysis, the samples are prepared by resuspending the proteins pellets or protein solutions, to a total volume of 20 µL using **loading buffer** and **Laemmli buffer**. The samples are then denatured at 90°C for 5 minutes. The samples are loaded on the SDS-PAGE gel, composed of two different parts the **stacking** and **separation gel**, where proteins will be separated according to their molecular weight. The proteins are separated under constant amperage of 25 mA. The migration takes place in **Laemmli buffer**. At the end of the migration, the gel can be stained with Coomassie Brilliant Blue to visualize total proteins or transferred to a PVDF membrane for further analysis.

- **Loading buffer :** 100 mM Tris-HCl pH 6.8 ; 2% (w/v) SDS ; 10% (v/v) glycerol ; 3% (v/v) ß mercaptoethanol ; 0.01% (w/v) bromophenol blue

- **Laemmli buffer :** 25 mM Tris HCl pH 8.3 ; 200 mM glycine ; 0,1% (w/v) SDS

- **Stacking gel 5% :** 5% acrylamide/bisacrylamide 37.5/1 ; 0.125 M Tris-HCl pH 6.8 ; 0.1% (w/v) SDS ; 0.1 % (v/v) APS ; 0.01 % TEMED

- **X% separation gel** : x% acrylamide/bisacrylamide 37.5/1 ; 0.375 M Tris-HCl pH 8.8 ; 0.1% (w/v) SDS ; 0.1% (v/v) APS ; 0.01% TEMED

Protein transfer to a PVDF membrane under liquid condition

To have the proteins accessible to antibody detection, they are electro-transferred from the SDS gel onto a PVDF membrane (Immobilon-P, 0.45 µm, Millipore). First, the membrane is activated using methanol, then the gel and the membrane are pre-equilibrated in **transfer buffer** for 15 minutes. The protein transfer is carried out in a Mini Trans-Blot® Cell (Bio-Rad) apparatus in the presence of transfer buffer for 1 h at 4°C under 360 mA. The membrane can then be stained in a **membrane staining solution** and then washed with a **membrane bleach solution**.

- **Transfer buffer**: 15% MeOH, 20 mM Tris, 200 mM Glycine

- **Membrane staining solution:** 0.1% Coomassie Blue R-250, 50% methanol, 7% acetate

- **Membrane bleach solution**: 50% methanol, 7% acetate

Immunodetection of proteins (Western blot)

First the membrane is blocked to avoid unspecific binding of the antibodies. This is performed by immersing the membrane in a solution of TBS-Tween 0.2% (v/v) (TBS-T) and milk 5% (w/v) for one hour at room temperature. Blocking solution is discarded and the membrane is then incubated for 1 hour in a TBS-T solution, 5% (w/v) milk containing the primary antibody directed against the protein of interest. The primary antibody solution is discarded and the membrane is washed for 5 min in TBS-T, which is repeated 3 times. Finally, the membrane is incubated for 30 min with the secondary antibody coupled to peroxidase and diluted to 1/10,000 in TBS-T. The membrane is then washed three times for 5 min in TBS-T and placed in the presence of the peroxidase substrate (Lumi-Light Western Blotting Substrate, Roche). The substrate emits light when in contact with the secondary antibody. The protein of interest is revealed thanks to the emitted light that impresses a photographic film.

LC-MS/MS analysis

All the mass spectrometry analyses were performed at the Strasbourg-Esplanade proteomic platform (http://www-ibmc.u-strasbg.fr/proteo/Web/accueil.htm), by Lauriane Kuhn, Johana Chicher and Philippe Hamman. In brief, protein extracts were precipitated (cold 0.1 M ammonium acetate in 100% methanol) and digested with sequencing-grade trypsin. Each sample was analyzed

by nanoLC-ESI-MS/MS on a QExactive+ mass spectrometer coupled to an EASY-nanoLC-1000 (Thermo-Fisher Scientific), with a 160-min gradient. Data were searched against the TAIR *Arabidopsis thaliana* database with a decoy strategy (release TAIRv10, 27281 forward protein sequences). Peptides and proteins were identified with Mascot algorithm (version 2.5.1, Matrix Science) and data were further imported into Proline v1.4 software (http://proline.profiproteomics.fr/). The total number of MS/MS fragmentation spectra was used to quantify each protein (Spectral Count label-free relative quantification). In the case of co-IP, to identify significantly enriched proteins, a statistical analysis by the msmsTests R package using spectral counts was performed (Gregori J, Sanchez A, 2013). The whole MS dataset was first normalized by the total number of MS/MS spectra (column-wise normalization). The implemented negative binominal model, which is based on the solution provided by the edgeR package was used (Robinson et al., 2010). P-values were then adjusted using the Benjamini & Hochberg method. Proteins that were over-represented in IP were visualized as a volcano plot that displays log2-fold-change and - log10-p.value on the x and y axes, respectively. The graphic was plotted using the Plotly's R graphing library.

Mitochondrial complexes analysis by Blue Native PAGE (BN-PAGE)

BN-PAGE was developed for the separation of mitochondrial membrane proteins and complexes in the mass range of 10 kDa to 10 MDa. The complexes are separated in native conditions, therefore it allows to perform specific experiments such as in-gel activity assays or to identify physiological protein–protein interactions.

Mitochondrial complexes were resolved by Blue-Native PAGE. For this, the equivalent of 500 µg of mitochondrial proteins were resuspended in 75 µL of **ACA buffer** supplemented with 25 µL of 6% (w/v) DDM, mixed and incubated 5 min on ice. The samples were clarified by centrifugation at 18,000 g for 40 min at 4°C. Supernatant was saved and supplemented with 15 µL of **Coomassie Blue 5%.** Complexes were separated on a continuous 5 to 13% **acrylamide gel** (previously casted using a gradient forming device), in cathode and anode buffer. Electrophoresis is first carried out 1 h at 7 mA followed by 3 h at 15 mA. Gel lanes are cut out and used for further analysis. For second dimension analysis and first dimension transfer, the gel lanes are denatured for 1 h at room temperature in **denaturation buffer**. For the second dimension, components of the various complexes were resolved by SDS-PAGE as described above.

- **ACA Buffer:** 750 mM Amino di-Caproic Acid, 50 mM bis-Tris pH 7 and 0.5 mM EDTA
- **Coomassie Blue 5%:** 5% (w/v) G250 Coomassie Blue in ACA buffer

- **Cathode buffer:** 50 mM Tricine, 15 mM bis-Tris pH 7, 0.02% (w/v) Coomassie Blue (Serva Blue G)
- **Anode buffer:** 50 mM bis-Tris pH 7
- **Separation gel:** 250 mM Amino di-Caproic Acid, 25 mM bis-Tris pH 7
- **5%:** 5% acrylamide/bisacrylamide 37.5/1, 0.1% (v/v) APS, 0.01% TEMED
- **13%:** 13% acrylamide/bisacrylamide 37.5/1, 0.1% (v/v) APS, 0.01% TEMED, 10% (v/v) glycerol
- **Stacking gel:** 250 mM Amino di-Caproic Acid, 25 mM bis-Tris pH 7, 4% acrylamide/bisacrylamide 37.5/1, 0.1% (v/v) APS, 0.01% TEMED, 8.5% (v/v) glycerol
- **Denaturation buffer:** 50 mM Tris-HCl pH 6.8, 1% (v/v) SDS and 1% (v/v) ß mercaptoethanol

In-gel Complex I activity test:

The gel lane was first quickly washed in pure water before being incubated 10 min in **100 mM Tris-HCl pH 7.4 solution**. The buffer was discarded and replaced by the **reaction buffer** and the gel lanes were incubated 10 min under stirring. The reaction is stopped by adding a large amount of **STOP buffer** and the gel lanes are incubated overnight under stirring.

- **Reaction buffer:** 100 mM Tris-HCl pH 7.4, 0.2 mM NADH, 0.2% NBT (Nitro-Blue Tetrazolium)
- **STOP buffer:** 45% (v/v) methanol, 10% (v/v) acetate

Plants

Agro-transformation by floraldip

A 5 mL preculture composed of LB supplemented with appropriate antibiotics is inoculated with the Agrobacterium strain transformed with the construct of interest. The culture is incubated for 2 days at 28 ° C under stirring. 50 µL of preculture is then used to inoculate 100 mL of medium of identical composition to the previous medium. The culture is incubated for 1 day at 28°C under stirring. After incubation, the bacteria are sedimented by centrifugation at 4000 rpm for 10 min at room temperature. They are then resuspended in **floral-dip medium** and incubated for 1 hour in the latter to allow the activation of the virulence genes. The inflorescences of the plants to be agroinfiltrated are soaked in the solution during 30 sec. The plants are drained and blacked out for 24 hours, then returned to their original culture conditions.

- **Floral-dip medium:** 2.2 g/L MS medium (M0222 Duchefa), 5% sucrose, 0.025% Silvet L-77, 200 µM acetosyringone

Transient protein expression in *Nicotiana benthamiana* leaves

Transient protein expression in *Nicotiana benthamiana* leaves is usually used to assess protein localization using a GFP-fused version of the protein of interest. A 10 mL culture composed of LB supplemented with appropriate antibiotics is inoculated with the Agrobacterium strain transformed with the construct of interest. The culture is incubated 16-24h at 28°C under stirring. After incubation the Agrobacterium cells are pelleted and resuspended in an appropriate volume of **agroinfiltration buffer** in order to have an OD_{600} of 1. An Agrobacterium strain containing a plasmid with the suppressor of silencing p19 is also grown in parallel to the constructs of interest. This p19 strain is diluted to an OD_{600} of 0.5. Each strains of interest are mixed volume to volume with the p19 strain and the mixtures are left at room temperature for 1 h. Finally 6-10 leaves *Nicotiana benthamiana* are used for the agroinfiltration, which is performed by wounding the leaf and infiltrating the contract of interest + p19 mixture using a 1 mL syringe. The infiltrated leaves are harvested 2-4 days later and observed under a confocal microscope.

- **Agroinfiltration buffer:** 1 mM $MgCl_2$, 50 µM acetosyringone

Seeds sterilization

The equivalent of 100 µL of seeds is placed in a 1.5 mL eppendorf tube containing 500 µL of 70% ethanol and 10 µL of Tween 20. The mixture is incubated for 15 min on a rotating wheel at room temperature. The liquid is removed and the seeds are washed three times for 5 min with 100% ethanol. After the last wash, the seeds are dried and ready to be sowed. The whole procedure is performed under a laminar flow cabinet.

Arabidopsis cell culture

Arabidopsis Col0 cell culture is being maintained at the institute. Every week 4 mL of one-to-two weeks old culture is used to inoculate a new 100 mL of culture. The cell medium is composed of 4,41 g/L of Murashige and Skoog basal medium (Duchefa M0256) supplemented with 30 g/l sucrose, 500 µl/l NAA (2 mg/ml) and 25 µl/l Kinetin (2 mg/ml), pH 5.6. The cells are kept under stirring (200rpm) in the dark.

Arabidopsis mitochondria purification

To purify mitochondria, three different types of starting material were used: Arabidopsis flowers, dark grown Arabidopsis cells and cauliflower. Arabidopsis flowers were mainly used to study mitochondria of mutant plants (e.g *rPPR1*, rPPR1HA, *336L*…) and perform co-IP. Mitochondria from cell culture were used to perform ribosome purification and then structural analysis, and

finally cauliflower was mainly used for protocol optimization, as the yield of mitochondria is extremely high compared to Arabidopsis (flowers or cells).

The buffer conditions are different for the three different starting materiel, but the overall procedure is similar. First the material is harvested: 30 g of Arabidopsis flowers (6-8 weeks old plants), 1 L of one week cell suspension (corresponding to 100 g of dried cells) or 1 kg of cauliflower.

Then the material is ground in **extraction buffer**. Arabidopsis flowers are ground in a warring blender, Arabidopsis cells are ground by hand using a mortar and a pestle and a juice extractor is used for cauliflower.

In all cases the lysate was filtered and clarified by centrifugation at 1,500 g, 10 min at 4°C. Supernatant was kept and centrifuged at 18,000 g, 15 min at 4°C. Organelle pellet was re-suspended in **wash buffer** and the precedent centrifugations were repeated once. The resulting organelle pellet (termed crude mitochondria) was re-suspended in wash buffer and loaded on a 10-23-40% Percoll gradient (in **wash buffer**) and run for 45 min at 40,000 g. Mitochondria are collected at the 23/40% interphase, and the Percoll is washed by adding at least 10 times the volume of wash buffer and centrifuging at 18,000 g 15 min at 4°C to pellet mitochondria. This process is repeated twice.

Finally, mitochondria are pelleted and resuspended in 1-2mL of **wash buffer** and the quantity is estimated by a Bradford assay.

Extraction buffer:

- **Flower:** 300 mM sucrose, 15 mM tetrasodium-pyrophosphate decahydrate, 2 mM EDTA, 10 mM KH_2PO_4, 1% (w/v) PVP-40, 1% (w/v) BSA, 20 mM ascorbate, 5 mM cysteine, pH 7.5

- **Cells:** 450 mM mannitol, 50 mM tetrasodium-pyrophosphate decahydrate, 0.5% (w/v) PVP-40, 0.5% (w/v) BSA, 20 mM ascorbate, 20 mM cysteine, pH 8

- **Cauliflower:** 300 mM mannitol, 30 mM tetrasodium-pyrophosphate decahydrate, 3 mM EDTA, 0.8% (w/v) PVP-25, 0.5% (w/v) BSA, 20 mM ascorbate, 5 mM cysteine, 2 mM ß mercaptoethanol, pH 7.5

Wash buffer:

- **Flower:** 300 mM sucrose, 10 mM MOPS, 1 mM EGTA, pH 7.5

- **Cells:** 300 mM mannitol, 10 mM TES-KOH, pH 7.5

- **Cauliflower:** 300 mM mannitol, 10 mM phosphate buffer, 1 mM EDTA, pH 7.5

Gradient buffer:

- **Flower:** X% Percoll, 300 mM sucrose, 10 mM MOPS, pH 7.2

- **Cells:** X% Percoll, 300 mM mannitol, 10 mM TES-KOH, pH 7.5
- **Cauliflower:** X% Percoll, 300 mM mannitol, 10 mM phosphate buffer, 1 mM EDTA, pH 7.5

Ribosomes purification from mitochondria

For ribosome purification mitochondria were re-suspended in **lysis buffer** to a concentration of 1 mg/ml (usually starting from at least 20 mg of mitochondria) and incubated for 15 min in 4°C. Lysate was clarified by centrifugation at 30,000 g, 20 min at 4°C. The supernatant was loaded on a 50% sucrose cushion in **monosome buffer** (3 mL of cushion and 7-10 mL of supernatant) and centrifuged at 235,000 g, 3h, 4°C in an ultracentrifuge. The crude ribosomes pellet was re-suspended in **monosome buffer** and loaded on a 10-30% sucrose gradient in the same buffer, formed using Gradient Master device from BioComp, and run for 16 h at 65,000 g in a swing ultracentrifuge. Fractions corresponding to mitoribosomes were collected using BioComp gradient collector and each fraction was pelleted and re-suspended in Monosome buffer, to a protein concentration of 2 μg/μL.

- **Lysis buffer:** 20 mM HEPES-KOH pH 7.6, 100 mM KCl, 30 mM $MgCl_2$, 1 mM DTT, 0.5 mg/ml heparin, 1.6% Triton X-100, 100 μg/ml chloramphenicol, supplemented with proteases inhibitors (Complete EDTA-free)
- **Monosome buffer:** 20 mM HEPES-KOH pH 7.6, 100 mM KCl, 30 mM $MgCl_2$, 1 mM DTT, 50 μg/ml chloramphenicol, supplemented with proteases inhibitors (Complete EDTA-free)

Microscopy

Confocal microscopy imaging

Subcellular localization analysis of recombinant fluorescent proteins was performed in *N.benthamiana* leaves, as described previously. The fluorescent proteins were visualized with the LSM780 confocal microscope from Carl Zeiss. The fluorochromes are excited using argon laser 488 (E-YFP, chlorophylls), after selection of the excitation wavelength. The specific fluorescence emission of the different fluorochromes is filtered in order to isolate the emitted light.

Assessment of ribosome samples by transmission electron microscopy

During sample screening, before cryo-EM and LC-MS/MS analyses, ribosome samples were analyzed by electron microscopy to assess the integrity and the purity of the samples. The samples were visualized with a CM120 100Kv (FEI) transmission electron microscope equipped with a CCD ORIUS 1000 Gatan Camera at the IGBMC (Illkirch). The samples were prepared by negative staining to improve the contrast. Briefly, 4 μL of ribosome solution was applied to a continuous carbon EM

grid and the sample was allowed to adsorb on the grid for 1 min. Excess liquid was discarded and 50 μL of staining solution (2% uranyl acetate solution) was applied to the grid for 2 min. Excess staining solution was discarded and the grid was air dried.

Data collection using cryo-electron microscopy

The cryo-electron microscopy acquisitions and data analysis was performed by Dr. Yaser Hashem (CNRS Strasbourg, INSERM Bordeaux). 4 μL of 70 nM Arabidopsis ribosome fraction prepared from purified mitochondria was applied to the cryo-EM grid, blotted with filter paper from both sides for 1.5 seconds in the temperature- and humidity-controlled Vitrobot apparatus Mark IV (FEI, T = 4°C, humidity 100%, Blot Force 5, Blot waiting time 30 sec) and vitrified in liquid ethane pre-cooled by liquid nitrogen. Data were collected on the Titan Krios S-FEG instrument (FEI) operating at 300 kV acceleration voltage the Falcon II 4096 x 4096 camera and automated data collection with EPU software (FEI). The Falcon II camera was calibrated at nominal magnification of 59,000 X.

Bioinformatic analyses

Subcellular localization predictions were determined with SUBA4 (Hooper et al., 2017). TargetP was also used for target sequence prediction (Emanuelsson et al., 2007).

Protein similarities were analyzed through MUSCLE Alignments (Edgar, 2004).

PPR domain predictions were performed with TPRpred (Karpenahalli et al., 2007).

Tridimensional structure predictions were built with Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2) (Kelley et al., 2015) and molecular representations were prepared with the PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC or UCSF Chimera (Pettersen et al., 2004).

Gene co-expression data was acquired with ATTED-II (Obayashi et al., 2018). Expression data was obtained from the Genevestigator® platform (Hruz et al., 2008).

To identify regions of similarity between biological sequences NCBI's BLAST was used. More specifically to find homologous proteins between organisms tblastn was used (Altschul et al., 1997).

Bibliography

Bibliography

Abrahams, J.P., Leslie, A.G.W., Lutter, R., and Walker, J.E. (1994). Structure at 2.8 Å resolution of F1-ATPase from bovine heart mitochondria. Nature *370*, 621–628.

Adl, S.M., Simpson, A.G.B., Lane, C.E., Lukeš, J., Bass, D., Bowser, S.S., Brown, M.W., Burki, F., Dunthorn, M., Hampl, V., et al. (2012). The Revised Classification of Eukaryotes. J. Eukaryot. Microbiol. *59*, 429–514.

Ahting, U., Thun, C., Hegerl, R., Typke, D., Nargang, F.E., Neupert, W., and Nussberger, S. (1999). The TOM core complex: the general protein import pore of the outer membrane of mitochondria. J. Cell Biol. *147*, 959–968.

Akabane, S., Ueda, T., Nierhaus, K.H., and Takeuchi, N. (2014). Ribosome Rescue and Translation Termination at Non-Standard Stop Codons by ICT1 in Mammalian Mitochondria. PLoS Genet. *10*, e1004616.

Van Aken, O., and Van Breusegem, F. (2015). Licensed to Kill: Mitochondria, Chloroplasts, and Cell Death. Trends Plant Sci. *20*, 754–766.

Allen, J.F. (2003). The function of genomes in bioenergetic organelles. Philos. Trans. R. Soc. B Biol. Sci. *358*, 19–38.

Altman, S. (2007). A view of RNase P. Mol. Biosyst. *3*, 604–607.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*, 3389–3402.

Alverson, A.J., Rice, D.W., Dickinson, S., Barry, K., and Palmer, J.D. (2011). Origins and Recombination of the Bacterial-Sized Multichromosomal Mitochondrial Genome of Cucumber. Plant Cell *23*, 2499–2513.

Amunts, A., Brown, A., Bai, X., Llácer, J.L., Hussain, T., Emsley, P., Long, F., Murshudov, G., Scheres, S.H.W., and Ramakrishnan, V. (2014). Structure of the yeast mitochondrial large ribosomal subunit. Science *343*, 1485–1489.

Amunts, A., Brown, A., Toots, J., Scheres, S.H.W., and Ramakrishnan, V. (2015). The structure of the human mitochondrial ribosome. Science *348*, 95–98.

Andersson, S.G.E., Zomorodipour, A., Andersson, J.O., Sicheritz-Pontén, T., Alsmark, U.C.M., Podowski, R.M., Näslund, A.K., Eriksson, A.-S., Winkler, H.H., and Kurland, C.G. (1998). The genome sequence of Rickettsia prowazekii and the origin ofmitochondria. Nature *396*, 133–140.

Aphasizhev, R., and Aphasizheva, I. (2014). Mitochondrial RNA editing in trypanosomes: Small RNAs in control. Biochimie *100*, 125–131.

Atkinson, G.C., Kuzmenko, A., Kamenski, P., Vysokikh, M.Y., Lakunina, V., Tankov, S., Smirnova, E., Soosaar, A., Tenson, T., and Hauryliuk, V. (2012). Evolutionary and genetic analyses of mitochondrial translation initiation factors identify the missing mitochondrial IF3 in S. cerevisiae. Nucleic Acids Res. *40*, 6122–6134.

Aubourg, S., Boudet, N., Kreis, M., and Lecharny, A. (2000). In Arabidopsis thaliana, 1% of the genome codes for a novel protein family unique to plants. Plant Mol. Biol. *42*, 603–613.

Ban, N., Beckmann, R., Cate, J.H.D., Dinman, J.D., Dragon, F., Ellis, S.R., Lafontaine, D.L.J., Lindahl, L.,

Liljas, A., Lipton, J.M., et al. (2014). A new system for naming ribosomal proteins. Curr. Opin. Struct. Biol.

Banks, J.A., Nishiyama, T., Hasebe, M., Bowman, J.L., Gribskov, M., dePamphilis, C., Albert, V.A., Aono, N., Aoyama, T., Ambrose, B.A., et al. (2011). The Selaginella Genome Identifies Genetic Changes Associated with the Evolution of Vascular Plants. Science *332*, 960–963.

Barkan, A., and Small, I. (2014). Pentatricopeptide repeat proteins in plants. Annu. Rev. Plant Biol. *65*, 415–442.

Barkan, A., Rojas, M., Fujii, S., Yap, A., Chong, Y.S., Bond, C.S., and Small, I. (2012). A Combinatorial Amino Acid Code for RNA Recognition by Pentatricopeptide Repeat Proteins. PLoS Genet. *8*, 4–11.

Ben-Shem, A., Garreau de Loubresse, N., Melnikov, S., Jenner, L., Yusupova, G., and Yusupov, M. (2011). The Structure of the Eukaryotic Ribosome at 3.0 A Resolution. Science *334*, 1524–1529.

Bentolila, S., Heller, W.P., Sun, T., Babina, A.M., Friso, G., van Wijk, K.J., and Hanson, M.R. (2012). RIP1, a member of an Arabidopsis protein family, interacts with the protein RARE1 and broadly affects RNA editing. Proc. Natl. Acad. Sci. U. S. A. *109*, E1453-61.

Bernhardt, H.S. (2012). The RNA world hypothesis: the worst theory of the early evolution of life (except for all the others). Biol. Direct *7*, 23.

Bernt, M., Braband, A., Schierwater, B., and Stadler, P.F. (2013). Genetic aspects of mitochondrial genome evolution. Mol. Phylogenet. Evol. *69*, 328–338.

Bieri, P., Greber, B.J., and Ban, N. (2018). High-resolution structures of mitochondrial ribosomes and their functional implications. Curr. Opin. Struct. Biol. *49*, 44–53.

Boerema, A.P., Aibara, S., Paul, B., Tobiasson, V., Kimanius, D., Forsberg, B.O., Wallden, K., Lindahl, E., and Amunts, A. (2018). Structure of the chloroplast ribosome with chl-RRF and hibernation-promoting factor. Nat. Plants *4*, 212–217.

Bogenhagen, D.F., Ostermeyer-Fay, A.G., Haley, J.D., and Garcia-Diaz, M. (2018). Kinetics and Mechanism of Mammalian Mitochondrial Ribosome Assembly. Cell Rep. *22*, 1935–1944.

Bonen, L., and Calixte, S. (2006). Comparative Analysis of Bacterial-Origin Genes for Plant Mitochondrial Ribosomal Proteins. Mol. Biol. Evol. *23*, 701–712.

Bonnard, G., Gobert, A., Arrivé, M., Pinker, F., Salinas-Giegé, T., and Giegé, P. (2016). Transfer RNA maturation in Chlamydomonas mitochondria, chloroplast and the nucleus by a single RNase P protein. Plant J. *87*, 270–280.

Bonnefoy, N., and Fox, T.D. (2007). Directed Alteration of Saccharomyces cerevisiae Mitochondrial DNA by Biolistic Transformation and Homologous Recombination. In Methods in Molecular Biology (Clifton, N.J.), pp. 153–166.

Boussardon, C., Salone, V., Avon, A., Berthome, R., Hammani, K., Okuda, K., Shikanai, T., Small, I., and Lurin, C. (2012). Two Interacting Proteins Are Necessary for the Editing of the NdhD-1 Site in Arabidopsis Plastids. Plant Cell *24*, 3684–3694.

Brito Querido, J., Mancera-Martínez, E., Vicens, Q., Bochler, A., Chicher, J., Simonetti, A., and Hashem, Y. (2017). The cryo-EM Structure of a Novel 40S Kinetoplastid-Specific Ribosomal Protein. Structure *25*, 1785–1794.e3.

Brown, J.R. (2003). Ancient horizontal gene transfer. Nat. Rev. Genet. *4*, 121–132.

Brown, A., Amunts, A., Bai, X.-C., Sugimoto, Y., Edwards, P.C., Murshudov, G., Scheres, S.H.W., and Ramakrishnan, V. (2014a). Structure of the large ribosomal subunit from human mitochondria. Science

346, 718–722.

Brown, A., Rathore, S., Kimanius, D., Aibara, S., Bai, X., Rorbach, J., Amunts, A., and Ramakrishnan, V. (2017). Structures of the human mitochondrial ribosome in native states of assembly. Nat. Struct. Mol. Biol. *24*, 866–869.

Brown, G.G., Colas des Francs-Small, C., and Ostersetzer-Biran, O. (2014b). Group II intron splicing factors in plant mitochondria. Front. Plant Sci. *5*, 35.

Bultema, J.B., Braun, H.-P., Boekema, E.J., and Kouřil, R. (2009). Megacomplex organization of the oxidative phosphorylation system by structural analysis of respiratory supercomplexes from potato. Biochim. Biophys. Acta - Bioenerg. *1787*, 60–67.

Burger, G., Gray, M.W., and Franz Lang, B. (2003). Mitochondrial genomes: anything goes. Trends Genet. *19*, 709–716.

Burger, G., Gray, M.W., Forget, L., and Lang, B.F. (2013). Strikingly bacteria-like and gene-rich mitochondrial genomes throughout jakobid protists. Genome Biol. Evol. *5*, 418–438.

Cahoon, A.B., and Qureshi, A.A. (2018). Leaderless mRNAs are circularized in Chlamydomonas reinhardtii mitochondria. Curr. Genet. *0*, 1–13.

Canino, G., Bocian, E., Barbezier, N., Echeverría, M., Forner, J., Binder, S., and Marchfelder, A. (2009). Arabidopsis encodes four tRNase Z enzymes. Plant Physiol. *150*, 1494–1502.

Cavalier-Smith, T. (1989). Archaebacteria and Archezoa. Nature *339*, 100–101.

Cazalet, C., Gomez-Valero, L., Rusniok, C., Lomma, M., Dervins-Ravault, D., Newton, H.J., Sansom, F.M., Jarraud, S., Zidane, N., Ma, L., et al. (2010). Analysis of the Legionella longbeachae Genome and Transcriptome Uncovers Unique Strategies to Cause Legionnaires' Disease. PLoS Genet. *6*, e1000851.

Cech, T.R. (1986). The generality of self-splicing RNA: Relationship to nuclear mRNA splicing. Cell *44*, 207–210.

Chen, L., and Liu, Y.-G. (2014). Male Sterility and Fertility Restoration in Crops. Annu. Rev. Plant Biol. *65*, 579–606.

Cheng, S., Gutmann, B., Zhong, X., Ye, Y., Fisher, M.F., Bai, F., Castleden, I., Song, Y., Song, B., Huang, J., et al. (2016). Redefining the structural motifs that determine RNA binding and RNA editing by pentatricopeptide repeat proteins in land plants. Plant J. *85*, 532–547.

Cho, Y., Qiu, Y.L., Kuhlman, P., and Palmer, J.D. (1998). Explosive invasion of plant mitochondria by a group I intron. Proc. Natl. Acad. Sci. U. S. A. *95*, 14244–14249.

Christianson, T., and Rabinowitz, M. (1983). Identification of multiple transcriptional initiation sites on the yeast mitochondrial genome by in vitro capping with guanylyltransferase. J. Biol. Chem. *258*, 14025–14033.

Chrzanowska-Lightowlers, Z., Rorbach, J., and Minczuk, M. (2017). Human mitochondrial ribosomes can switch structural tRNAs – but when and why? RNA Biol. *6286*, 00–00.

Chrzanowska-Lightowlers, Z.M.A., Pajak, A., and Lightowlers, R.N. (2011). Termination of protein synthesis in mammalian mitochondria. J. Biol. Chem. *286*, 34479–34485.

Clapham, D.E. (2007). Calcium Signaling. Cell *131*, 1047–1058.

Claros, M.G., Perea, J., Shu, Y., Samatey, F.A., Popot, J.-L., and Jacq, C. (1995). Limitations to in vivo Import of Hydrophobic Proteins into Yeast Mitochondria. The Case of a Cytoplasmically Synthesized Apocytochrome b. Eur. J. Biochem. *228*, 762–771.

Clough, S.J., and Bent, A.F. (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J. *16*, 735–743.

Colcombet, J., Lopez-Obando, M., Heurtevin, L., Bernard, C., Martin, K., Berthomé, R., and Lurin, C. (2013). Systematic study of subcellular localization of Arabidopsis PPR proteins confirms a massive targeting to organelles. RNA Biol. *10*, 1557–1575.

Coquille, S., Filipovska, A., Chia, T., Rajappa, L., Lingford, J.P., Razif, M.F.M., Thore, S., and Rackham, O. (2014). An artificial PPR scaffold for programmable RNA recognition. Nat. Commun. *5*, 5729.

Cotton, J.A., and McInerney, J.O. (2010). Eukaryotic genes of archaebacterial origin are more important than the more numerous eubacterial genes, irrespective of function. Proc. Natl. Acad. Sci. U. S. A. *107*, 17252–17255.

D.N.J. de Grey, A. (2005). Forces maintaining organellar genomes: is any as strong as genetic code disparity or hydrophobicity? BioEssays *27*, 436–446.

Daems, W.T., and Wisse, E. (1966). Shape and attachment of the cristae mitochondriales in mouse hepatic cell mitochondria. J. Ultrastruct. Res. *16*, 123–140.

Darwin, C. (1859). On the origin of species by means of natural selection, or, the preservation of favoured races in the struggle for life (London).

Davies, K.M., Strauss, M., Daum, B., Kief, J.H., Osiewacz, H.D., Rycovska, A., Zickermann, V., and Kuhlbrandt, W. (2011). Macromolecular organization of ATP synthase and complex I in whole mitochondria. Proc. Natl. Acad. Sci. *108*, 14121–14126.

Dekker, P.J., Ryan, M.T., Brix, J., Müller, H., Hönlinger, A., and Pfanner, N. (1998). Preprotein translocase of the outer mitochondrial membrane: molecular dissection and assembly of the general import pore complex. Mol. Cell. Biol. *18*, 6515–6524.

Denslow, N.D., Anders, J.C., and O'Brien, T.W. (1991). Bovine mitochondrial ribosomes possess a high affinity binding site for guanine nucleotides. J. Biol. Chem. *266*, 9586–9590.

Derbikova, K.S., Levitsky, S.A., Chicherin, I. V., Vinogradova, E.N., and Kamenski, P.A. (2018). Activation of Yeast Mitochondrial Translation: Who Is in Charge? Biochem. *83*, 87–97.

Desai, N., Brown, A., Amunts, A., and Ramakrishnan, V. (2017). The structure of the yeast mitochondrial ribosome. Science *355*, 528–531.

Dorrell, R.G., and Howe, C.J. (2012). What makes a chloroplast? Reconstructing the establishment of photosynthetic symbioses. J. Cell Sci. *125*, 1865–1875.

Duarte, I., Nabuurs, S.B., Magno, R., and Huynen, M. (2012). Evolution and Diversification of the Organellar Release Factor Family. Mol. Biol. Evol. *29*, 3497–3512.

Dube, P., Wieske, M., Stark, H., Schatz, M., Stahl, J., Zemlin, F., Lutsch, G., and van Heel, M. (1998). The 80S rat liver ribosome at 25 å resolution by electron cryomicroscopy and angular reconstitution. Structure *6*, 389–399.

Dyson, M.R., Lee, C.P., Mandal, N., Seong, B.L., Varshney, U., and RajBhandary, U.L. (1993). Identity of a Prokaryotic Initiator tRNA. In The Translational Apparatus, (Boston, MA: Springer US), pp. 23–33.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797.

Emanuelsson, O., Brunak, S., von Heijne, G., and Nielsen, H. (2007). Locating proteins in the cell using TargetP, SignalP and related tools. Nat. Protoc. *2*, 953–971.

Eme, L., Spang, A., Lombard, J., Stairs, C.W., and Ettema, T.J.G. (2017). Archaea and the origin of eukaryotes. Nat. Rev. Microbiol. *15*, 711–723.

Eubel, H., Jänsch, L., and Braun, H.-P. (2003). New insights into the respiratory chain of plant mitochondria. Supercomplexes and a unique composition of complex II. Plant Physiol. *133*, 274–286.

Feagin, J.E., Harrell, M.I., Lee, J.C., Coe, K.J., Sands, B.H., Cannone, J.J., Tami, G., Schnare, M.N., and Gutell, R.R. (2012). The Fragmented Mitochondrial Ribosomal RNAs of Plasmodium falciparum. PLoS One *7*, e38320.

Fernie, A.R., Carrari, F., and Sweetlove, L.J. (2004). Respiratory metabolism: glycolysis, the TCA cycle and mitochondrial electron transport. Curr. Opin. Plant Biol. *7*, 254–261.

Fey, J., Dietrich, A., Cosset, A., Desprez, T., and Maréchal-Drouard, L. (1997). Evolutionary aspects of "chloroplast-like" trnN and trnH expression in higher-plant mitochondria. Curr. Genet. *32*, 358–360.

Filipovska, A., and Rackham, O. (2012). Modular recognition of nucleic acids by PUF, TALE and PPR proteins. Mol. Biosyst. *8*, 699.

Forner, J., Weber, B., Thuss, S., Wildum, S., and Binder, S. (2007). Mapping of mitochondrial mRNA termini in Arabidopsis thaliana: t-elements contribute to 5' and 3' end formation. Nucleic Acids Res. *35*, 3676–3692.

Fournier, D., Palidwor, G.A., Shcherbinin, S., Szengel, A., Schaefer, M.H., Perez-Iratxeta, C., and Andrade-Navarro, M.A. (2013). Functional and genomic analyses of alpha-solenoid proteins. PLoS One *8*, e79894.

Frugier, M., Bour, T., Ayach, M., Santos, M. a S., Rudinger-Thirion, J., Théobald-Dietrich, A., and Pizzi, E. (2010). Low Complexity Regions behave as tRNA sponges to help co-translational folding of plasmodial proteins. FEBS Lett. *584*, 448–454.

Gagliardi, D., and Leaver, C.J. (1999). Polyadenylation accelerates the degradation of the mitochondrial mRNA associated with cytoplasmic male sterility in sunflower. EMBO J. *18*, 3757–3766.

Gaur, R., Grasso, D., Datta, P.P., Krishna, P.D. V, Das, G., Spencer, A., Agrawal, R.K., Spremulli, L., and Varshney, U. (2008). A single mammalian mitochondrial translation initiation factor functionally replaces two bacterial factors. Mol. Cell *29*, 180–190.

Giegé, P., Konthur, Z., Walter, G., and Brennicke, A. (1998). An ordered Arabidopsis thaliana mitochondrial cDNA library on high-density filters allows rapid systematic analysis of plant gene expression: a pilot study. Plant J. *15*, 721–726.

van der Giezen, M. (2009). Hydrogenosomes and mitosomes: conservation and evolution of functions. J. Eukaryot. Microbiol. *56*, 221–231.

Gipson, A.B., Morton, K.J., Rhee, R.J., Simo, S., Clayton, J.A., Perrett, M.E., Binkley, C.G., Jensen, E.L., Oakes, D.L., Rouhier, M.F., et al. (2017). Disruptions in valine degradation affect seed development and germination in Arabidopsis. Plant J. *90*, 1029–1039.

Gleave, A.P. (1992). A versatile binary vector system with a T-DNA organisational structure conducive to efficient integration of cloned DNA into the plant genome. Plant Mol. Biol. *20*, 1203–1207.

Gobert, A., Gutmann, B., Taschner, A., Gössringer, M., Holzmann, J., Hartmann, R.K., Rossmanith, W., and Giegé, P. (2010). A single Arabidopsis organellar protein has RNase P activity. Nat. Struct. Mol. Biol. *17*, 740–744.

Gold, V.A., Chroscicki, P., Bragoszewski, P., and Chacinska, A. (2017). Visualization of cytosolic ribosomes on the surface of mitochondria by electron cryo-tomography. EMBO Rep.

Goldschmidt-Reisin, S., Kitakawa, M., Herfurth, E., Wittmann-Liebold, B., Grohmann, L., and Graack, H.R. (1998). Mammalian mitochondrial ribosomal proteins. N-terminal amino acid sequencing, characterization, and identification of corresponding gene sequences. J. Biol. Chem. *273*, 34828–34836.

Gray, M.W. (2012). Mitochondrial evolution. Cold Spring Harb. Perspect. Biol. *4*, a011403.

Gray, M.W. (2015). Mosaic nature of the mitochondrial proteome: Implications for the origin and evolution of mitochondria. Proc. Natl. Acad. Sci. U. S. A. *112*, 10133–10138.

Gray, M.W., Burger, G., and Lang, B.F. (1999). Mitochondrial evolution. Science *283*, 1476–1481.

Greber, B.J., and Ban, N. (2016). Structure and Function of the Mitochondrial Ribosome. Annu. Rev. Biochem. *85*, annurev-biochem-060815-014343.

Greber, B.J., Boehringer, D., Leitner, A., Bieri, P., Voigts-Hoffmann, F., Erzberger, J.P., Leibundgut, M., Aebersold, R., and Ban, N. (2013). Architecture of the large subunit of the mammalian mitochondrial ribosome. Nature *505*, 515–519.

Greber, B.J., Bieri, P., Leibundgut, M., Leitner, A., Aebersold, R., Boehringer, D., and Ban, N. (2015). The complete structure of the 55S mammalian mitochondrial ribosome. Science *348*, 303–308.

Gregori J, Sanchez A, V.J. (2013). msmsTests: LC-MS/MS Differential Expression Tests. R package version 1.18.0.

Griparic, L., and van der bliek, A.M. (2001). The Many Shapes of Mitochondrial Membranes. Traffic *2*, 235–244.

Gu, J., Wu, M., Guo, R., Yan, K., Lei, J., Gao, N., and Yang, M. (2016). The architecture of the mammalian respirasome. Nature *537*, 639–643.

Gualberto, J.M., and Newton, K.J. (2017). Plant Mitochondrial Genomes: Dynamics and Mechanisms of Mutation. Annu. Rev. Plant Biol. *68*, 225–252.

Gualberto, J.M., Lamattina, L., Bonnard, G., Weil, J.-H., and Grienenberger, J.-M. (1989). RNA editing in wheat mitochondria results in the conservation of protein sequences. Nature *341*, 660–662.

Gualerzi, C.O., and Pon, C.L. (1990). Initiation of mRNA translation in prokaryotes. Biochemistry *29*, 5881–5889.

Gutmann, B., Gobert, A., and Giegé, P. (2012a). Mitochondrial Genome Evolution and the Emergence of PPR Proteins. In Advances in Botanical Research, pp. 253–313.

Gutmann, B., Gobert, A., and Giegé, P. (2012b). PRORP proteins support RNase P activity in both organelles and the nucleus in Arabidopsis. Genes Dev. *26*, 1022–1027.

Haffter, P., McMullin, T.W., and Fox, T.D. (1990). A genetic link between an mRNA-specific translational activator and the translation system in yeast mitochondria. Genetics *125*, 495–503.

Hahn, A., Parey, K., Bublitz, M., Mills, D.J., Zickermann, V., Vonck, J., Kühlbrandt, W., and Meier, T. (2016). Structure of a Complete ATP Synthase Dimer Reveals the Molecular Basis of Inner Mitochondrial Membrane Morphology. Mol. Cell *63*, 445–456.

Haïli, N., Arnal, N., Quadrado, M., Amiar, S., Tcherkez, G., Dahan, J., Briozzo, P., Colas des Francs-Small, C., Vrielynck, N., and Mireau, H. (2013). The pentatricopeptide repeat MTSF1 protein stabilizes the nad4 mRNA in Arabidopsis mitochondria. Nucleic Acids Res. *41*, 6650–6663.

Haïli, N., Planchard, N., Arnal, N., Quadrado, M., Vrielynck, N., Dahan, J., des Francs-Small, C.C., and Mireau, H. (2016). The MTL1 Pentatricopeptide Repeat Protein Is Required for Both Translation and

Splicing of the Mitochondrial NADH DEHYDROGENASE SUBUNIT7 mRNA in Arabidopsis. Plant Physiol. *170*, 354–366.

Hammani, K., and Giegé, P. (2014). RNA metabolism in plant mitochondria. Trends Plant Sci. *19*, 380–389.

Hammani, K., Gobert, A., Hleibieh, K., Choulier, L., Small, I., and Giegé, P. (2011). An Arabidopsis dual-localized pentatricopeptide repeat protein interacts with nuclear proteins involved in gene expression regulation. Plant Cell *23*, 730–740.

Hancock, K., and Hajduk, S.L. (1990). The mitochondrial tRNAs of Trypanosoma brucei are nuclear encoded. J. Biol. Chem. *265*, 19208–19215.

Hao, W., Richardson, A.O., Zheng, Y., and Palmer, J.D. (2010). Gorgeous mosaic of mitochondrial genes created by horizontal transfer and gene conversion. Proc. Natl. Acad. Sci. *107*, 21576–21581.

Haque, M.E., Elmore, K.B., Tripathy, A., Koc, H., Koc, E.C., and Spremulli, L.L. (2010). Properties of the C-terminal Tail of Human Mitochondrial Inner Membrane Protein Oxa1L and Its Interactions with Mammalian Mitochondrial Ribosomes. J. Biol. Chem. *285*, 28353–28362.

Hashem, Y., des Georges, A., Fu, J., Buss, S.N., Jossinet, F., Jobe, A., Zhang, Q., Liao, H.Y., Grassucci, R.A., Bajaj, C., et al. (2013). High-resolution cryo-electron microscopy structure of the Trypanosoma brucei ribosome. Nature *494*, 385–389.

Hauler, A., Jonietz, C., Stoll, B., Stoll, K., Braun, H.-P., and Binder, S. (2013). RNA PROCESSING FACTOR 5 is required for efficient 5′ cleavage at a processing site conserved in RNAs of three different mitochondrial genes in Arabidopsis thaliana. Plant J. *74*, 593–604.

Hayashi, T., and Stuchebrukhov, A.A. (2010). Electron tunneling in respiratory complex I. Proc. Natl. Acad. Sci. U. S. A. *107*, 19157–19162.

von Heijne, G. (1986). Why mitochondria need a genome. FEBS Lett. *198*, 1–4.

Heinemeyer, J., Braun, H.-P., Boekema, E.J., and Kouřil, R. (2007). A Structural Model of the Cytochrome c Reductase/Oxidase Supercomplex from Yeast Mitochondria. J. Biol. Chem. *282*, 12240–12248.

Herbert, C.J., Golik, P., and Bonnefoy, N. (2013a). Yeast PPR proteins, watchdogs of mitochondrial gene expression. RNA Biol. *10*, 1477–1494.

Herrmann, J.M., Woellhaf, M.W., and Bonnefoy, N. (2013). Control of protein synthesis in yeast mitochondria: The concept of translational activators. Biochim. Biophys. Acta - Mol. Cell Res. *1833*, 286–294.

Hoch, F.L. (1992). Cardiolipins and biomembrane function. Biochim. Biophys. Acta *1113*, 71–133.

Hodge, T., and Colombini, M. (1997). Regulation of metabolite flux through voltage-gating of VDAC channels. J. Membr. Biol. *157*, 271–279.

Holec, S., Lange, H., Kuhn, K., Alioua, M., Borner, T., and Gagliardi, D. (2006). Relaxed Transcription in Arabidopsis Mitochondria Is Counterbalanced by RNA Stability Control Mediated by Polyadenylation and Polynucleotide Phosphorylase. Mol. Cell. Biol. *26*, 2869–2876.

Holsters, M., Silva, B., Van Vliet, F., Genetello, C., De Block, M., Dhaese, P., Depicker, A., Inzé, D., Engler, G., Villarroel, R., et al. (1980). The functional organization of the nopaline A. tumefaciens plasmid pTiC58. Plasmid *3*, 212–230.

Holzmann, J., Frank, P., Löffler, E., Bennett, K.L., Gerner, C., and Rossmanith, W. (2008). RNase P without RNA: Identification and Functional Reconstitution of the Human Mitochondrial tRNA

Processing Enzyme. Cell *135*, 462–474.

Hooper, C.M., Castleden, I.R., Tanz, S.K., Aryamanesh, N., and Millar, A.H. (2017). SUBA4: the interactive data analysis centre for Arabidopsis subcellular protein locations. Nucleic Acids Res. *45*, D1064–D1074.

Horn, R., Gupta, K.J., and Colombo, N. (2014). Mitochondrion role in molecular basis of cytoplasmic male sterility. Mitochondrion.

Howard, M.J., Lim, W.H., Fierke, C.A., and Koutmos, M. (2012). Mitochondrial ribonuclease P structure provides insight into the evolution of catalytic strategies for precursor-tRNA 5' processing. Proc. Natl. Acad. Sci. *109*, 16149–16154.

Hruz, T., Laule, O., Szabo, G., Wessendorp, F., Bleuler, S., Oertle, L., Widmayer, P., Gruissem, W., and Zimmermann, P. (2008). Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. Adv. Bioinformatics *2008*, 420747.

Huchon, D., Szitenberg, A., Shefer, S., Ilan, M., and Feldstein, T. (2015). Mitochondrial group I and group II introns in the sponge orders Agelasida and Axinellida. BMC Evol. Biol. *15*, 278.

Jalal, A., Schwarz, C., Schmitz-Linneweber, C., Vallon, O., Nickelsen, J., and Bohne, A.-V. (2015). A Small Multifunctional Pentatricopeptide Repeat Protein in the Chloroplast of Chlamydomonas reinhardtii. Mol. Plant *8*, 412–426.

Jühling, F., Pütz, J., Florentz, C., and Stadler, P.F. (2012). Armless mitochondrial tRNAs in enoplea (nematoda). RNA Biol. *9*, 1161–1166.

Jühling, T., Duchardt-Ferner, E., Bonin, S., Wöhnert, J., Pütz, J., Florentz, C., Betat, H., Sauter, C., and Mörl, M. (2018). Small but large enough: structural properties of armless mitochondrial tRNAs from the nematode Romanomermis culicivorax. Nucleic Acids Res. *46*, 9170–9180.

Kalanon, M., and McFadden, G.I. (2010). Malaria, Plasmodium falciparum and its apicoplast. Biochem. Soc. Trans. *38*, 775–782.

Karnkowska, A., Vacek, V., Zubáčová, Z., Treitli, S.C., Petrželková, R., Eme, L., Novák, L., Žárský, V., Barlow, L.D., Herman, E.K., et al. (2016). A Eukaryote without a Mitochondrial Organelle. Curr. Biol. *26*, 1274–1284.

Karpenahalli, M.R., Lupas, A.N., and Söding, J. (2007). TPRpred: a tool for prediction of TPR-, PPR- and SEL1-like repeats from protein sequences. BMC Bioinformatics *8*, 2.

Ke, J., Chen, R.-Z., Ban, T., Zhou, X.E., Gu, X., Tan, M.H.E., Chen, C., Kang, Y., Brunzelle, J.S., Zhu, J.-K., et al. (2013). Structural basis for RNA recognition by a dimeric PPR-protein complex. Nat. Struct. Mol. Biol. *20*, 1377–1382.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J.E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc. *10*, 845–858.

Khatter, H., Myasnikov, A.G., Natchiar, S.K., and Klaholz, B.P. (2015). Structure of the human 80S ribosome. Nature *520*, 640–645.

Kirby, L.E., and Koslowsky, D. (2017). Mitochondrial dual-coding genes in Trypanosoma brucei. PLoS Negl. Trop. Dis. *11*, e0005989.

Kitakawa, M., Graack, H.R., Grohmann, L., Goldschmidt-Reisin, S., Herfurth, E., Wittmann-Liebold, B., Nishimura, T., and Isono, K. (1997). Identification and characterization of the genes for mitochondrial ribosomal proteins of Saccharomyces cerevisiae. Eur. J. Biochem. *245*, 449–456.

Koc, E.C., Burkhart, W., Blackburn, K., Moseley, A., Koc, H., and Spremulli, L.L. (2000). A proteomics

approach to the identification of mammalian mitochondrial small subunit ribosomal proteins. J. Biol. Chem. *275*, 32585–32591.

Koc, E.C., Burkhart, W., Blackburn, K., Moyer, M.B., Schlatzer, D.M., Moseley, A., and Spremulli, L.L. (2001). The Large Subunit of the Mammalian Mitochondrial Ribosome. J. Biol. Chem. *276*, 43958–43969.

Koehler, C.M., Jarosch, E., Tokatlidis, K., Schmid, K., Schweyen, R.J., and Schatz, G. (1998). Import of mitochondrial carriers mediated by essential proteins of the intermembrane space. Science *279*, 369–373.

Kolitz, S.E., and Lorsch, J.R. (2010). Eukaryotic initiator tRNA: Finely tuned and ready for action. FEBS Lett. *584*, 396–404.

Körte, A., Forsbach, V., Gottenöf, T., and Rödel, G. (1989). In vitro and in vivo studies on the mitochondrial import of CBS1, a translational activator of cytochrome b in yeast. Mol. Gen. Genet. *217*, 162–167.

Kruszewski, J., and Golik, P. (2016). Pentatricopeptide motifs in the N-terminal extension domain of yeast mitochondrial RNA polymerase Rpo41p are not essential for its function. Biochem. *81*, 1101–1110.

Kühlbrandt, W. (2014). The Resolution Revolution. Science *343*, 1443–1444.

Kühn, K., Richter, U., Meyer, E.H., Delannoy, E., de Longevialle, A.F., O'Toole, N., Börner, T., Millar, A.H., Small, I.D., and Whelan, J. (2009). Phage-type RNA polymerase RPOTmp performs gene-specific transcription in mitochondria of Arabidopsis thaliana. Plant Cell *21*, 2762–2779.

Kummer, E., Leibundgut, M., Rackham, O., Lee, R.G., Boehringer, D., Filipovska, A., and Ban, N. (2018). Unique features of mammalian mitochondrial translation initiation revealed by cryo-EM. Nature *560*, 263–267.

Kuzmenko, A., Atkinson, G.C., Levitskii, S., Zenkin, N., Tenson, T., Hauryliuk, V., and Kamenski, P. (2014). Mitochondrial translation initiation machinery: Conservation and diversification. Biochimie *100*, 132–140.

Lafontaine, D.L.J., and Tollervey, D. (2001). The function and synthesis of ribosomes. Nat. Rev. Mol. Cell Biol. *2*, 514–520.

Lagouge, M., Mourier, A., Ju Lee, H., Spåhr, H., Wai, T., Kukat, C., Silva Ramos, E., Motori, E., Busch, J.D., Siira, S., et al. (2015). SLIRP Regulates the Rate of Mitochondrial Protein Synthesis and Protects LRPPRC from Degradation. PLOS Genet. | DOI.

Lane, N. (2011). Energetics and genetics across the prokaryote-eukaryote divide. Biol. Direct *6*, 35.

Lane, N. (2017). Serial endosymbiosis or singular event at the origin of eukaryotes? J. Theor. Biol. *434*, 58–67.

Lane, N., and Martin, W. (2010). The energetics of genome complexity. Nature *467*, 929–934.

Lang, B.F., Burger, G., O'Kelly, C.J., Cedergren, R., Golding, G.B., Lemieux, C., Sankoff, D., Turmel, M., and Gray, M.W. (1997). An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. Nature *387*, 493–497.

Lau, W.C.Y., and Rubinstein, J.L. (2012). Subnanometre-resolution structure of the intact Thermus thermophilus H+-driven ATP synthase. Nature *481*, 214–218.

Lechner, M., Rossmanith, W., Hartmann, R.K., Thölken, C., Gutmann, B., Giegé, P., and Gobert, A. (2015). Distribution of Ribonucleoprotein and Protein-Only RNase P in Eukarya. Mol. Biol. Evol. *32*,

msv187.

Liere, K., and Börner, T. (2011). Plant Mitochondria (New York, NY: Springer New York).

Liere, K., Weihe, A., and Börner, T. (2011). The transcription machineries of plant mitochondria and chloroplasts: Composition, function, and regulation. J. Plant Physiol. *168*, 1345–1360.

Lightowlers, R.N., and Chrzanowska-Lightowlers, Z.M.A. (2013). Human pentatricopeptide proteins: only a few and what do they do? RNA Biol. *10*, 1433–1438.

Lill, R., Diekert, K., Kaut, A., Lange, H., Pelzer, W., Prohl, C., and Kispal, G. (1999). The Essential Role of Mitochondria in the Biogenesis of Cellular Iron-Sulfur Proteins. Biol. Chem. *380*, 1157–1166.

Litonin, D., Sologub, M., Shi, Y., Savkina, M., Anikin, M., Falkenberg, M., Gustafsson, C.M., and Temiakov, D. (2010). Human Mitochondrial Transcription Revisited. J. Biol. Chem. *285*, 18129–18133.

Liu, M., and Spremulli, L. (2000). Interaction of Mammalian Mitochondrial Ribosomes with the Inner Membrane. J. Biol. Chem. *275*, 29400–29406.

Lurin, C., Andrés, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyère, C., Caboche, M., Debast, C., Gualberto, J., Hoffmann, B., et al. (2004). Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. Plant Cell *16*, 2089–2103.

Malek, O., and Knoop, V. (1998). Trans-splicing group II introns in plant mitochondria: the complete set of cis-arranged homologs in ferns, fern allies, and a hornwort. RNA *4*, 1599–1609.

Manna, S. (2015). An overview of pentatricopeptide repeat proteins and their applications. Biochimie *113*, 93–99.

Manthey, G.M., and McEwen, J.E. (1995). The product of the nuclear gene PET309 is required for translation of mature mRNA and stability or production of intron-containing RNAs derived from the mitochondrial COX1 locus of Saccharomyces cerevisiae. EMBO J. *14*, 4031–4043.

Marienfeld, J., Unseld, M., and Brennicke, A. (1999). The mitochondrial genome of Arabidopsis is composed of both native and immigrant information. Trends Plant Sci. *4*, 495–502.

Martijn, J., Vosseberg, J., Guy, L., Offre, P., and Ettema, T.J.G. (2018). Deep mitochondrial origin outside the sampled alphaproteobacteria. Nature *557*, 101–105.

Martin, W., and Russell, M.J. (2003). On the origins of cells: a hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells. Philos. Trans. R. Soc. B Biol. Sci. *358*, 59–85.

Martin, W., Rujan, T., Richly, E., Hansen, A., Cornelsen, S., Lins, T., Leister, D., Stoebe, B., Hasegawa, M., and Penny, D. (2002). Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. Proc. Natl. Acad. Sci. *99*, 12246–12251.

Maxwell, D.P., Wang, Y., and McIntosh, L. (1999). The alternative oxidase lowers mitochondrial reactive oxygen production in plant cells. Proc. Natl. Acad. Sci. *96*, 8271–8276.

McDonald, A.E., and Vanlerberghe, G.C. (2006). Origins, evolutionary history, and taxonomic distribution of alternative oxidase and plastoquinol terminal oxidase. Comp. Biochem. Physiol. Part D Genomics Proteomics *1*, 357–364.

Melnikov, S., Ben-Shem, A., Garreau de Loubresse, N., Jenner, L., Yusupova, G., and Yusupov, M. (2012). One core, two shells: bacterial and eukaryotic ribosomes. Nat. Struct. Mol. Biol. *19*, 560–567.

Merrick, W.C. (2004). Cap-dependent and cap-independent translation in eukaryotic systems. Gene

332, 1–11.

Mihara, K., and Sato, R. (1985). Molecular cloning and sequencing of cDNA for yeast porin, an outer mitochondrial membrane protein: a search for targeting signal in the primary structure. EMBO J. *4*, 769–774.

Minczuk, M., He, J., Duch, A.M., Ettema, T.J., Chlebowski, A., Dzionek, K., Nijtmans, L.G.J., Huynen, M.A., and Holt, I.J. (2011). TEFM (c17orf42) is necessary for transcription of human mtDNA. Nucleic Acids Res. *39*, 4284–4299.

Miranda, R.G., McDermott, J.J., and Barkan, A. (2018). RNA-binding specificity landscapes of designer pentatricopeptide repeat proteins elucidate principles of PPR-RNA interactions. Nucleic Acids Res. *46*, 2613–2623.

Mitchell, P. (1961). Coupling of Phosphorylation to Electron and Hydrogen Transfer by a Chemi-Osmotic type of Mechanism. Nature *191*, 144–148.

Mueller, S.J., and Reski, R. (2015). Mitochondrial Dynamics and the ER: The Plant Perspective. Front. Cell Dev. Biol. *3*, 78.

Mühleip, A.W., Joos, F., Wigge, C., Frangakis, A.S., Kühlbrandt, W., and Davies, K.M. (2016). Helical arrays of U-shaped ATP synthase dimers form tubular cristae in ciliate mitochondria. Proc. Natl. Acad. Sci. U. S. A. *113*, 8442–8447.

Myasnikov, A.G., Simonetti, A., Marzi, S., and Klaholz, B.P. (2009). Structure–function insights into prokaryotic and eukaryotic translation initiation. Curr. Opin. Struct. Biol. *19*, 300–309.

Noeske, J., Wasserman, M.R., Terry, D.S., Altman, R.B., Blanchard, S.C., and Cate, J.H.D. (2015). High-resolution structure of the Escherichia coli ribosome. Nat. Struct. Mol. Biol. *22*, 336–341.

Nosek, J., and Tomáška, Ľ. (2003). Mitochondrial genome diversity: evolution of the molecular architecture and replication strategy. Curr. Genet. *44*, 73–84.

Nunnari, J., and Suomalainen, A. (2012). Mitochondria: In sickness and in health. Cell.

O'Malley, R.C., Barragan, C.C., and Ecker, J.R. (2015). A user's guide to the Arabidopsis T-DNA insertion mutant collections. Methods Mol. Biol. *1284*, 323–342.

O'Toole, N., Hattori, M., Andres, C., Iida, K., Lurin, C., Schmitz-Linneweber, C., Sugita, M., and Small, I. (2008). On the Expansion of the Pentatricopeptide Repeat Gene Family in Plants. Mol. Biol. Evol. *25*, 1120–1128.

Obayashi, T., Aoki, Y., Tadaka, S., Kagaya, Y., and Kinoshita, K. (2018). ATTED-II in 2018: A Plant Coexpression Database Based on Investigation of the Statistical Property of the Mutual Rank Index. Plant Cell Physiol. *59*, e3.

Ojala, D., Merkel, C., Gelfand, R., and Attardi, G. (1980). The tRNA genes punctuate the reading of genetic information in human mitochondrial DNA. Cell *22*, 393–403.

Ojala, D., Montoya, J., and Attardi, G. (1981). tRNA punctuation model of RNA processing in human mitochondria. Nature *290*, 470–474.

Okuda, K., Myouga, F., Motohashi, R., Shinozaki, K., and Shikanai, T. (2007). Conserved domain structure of pentatricopeptide repeat proteins involved in chloroplast RNA editing. Proc. Natl. Acad. Sci. *104*, 8178–8183.

Okuda, K., Chateigner-Boutin, A.-L., Nakamura, T., Delannoy, E., Sugita, M., Myouga, F., Motohashi, R., Shinozaki, K., Small, I., and Shikanai, T. (2009). Pentatricopeptide Repeat Proteins with the DYW Motif Have Distinct Molecular Functions in RNA Editing and RNA Cleavage in Arabidopsis Chloroplasts.

PLANT CELL ONLINE *21*, 146–156.

Omura, T. (1998). Mitochondria-targeting sequence, a multi-role sorting sequence recognized at all steps of protein import into mitochondria. J. Biochem. *123*, 1010–1016.

Ott, M., and Herrmann, J.M. (2010). Co-translational membrane insertion of mitochondrially encoded proteins. Biochim. Biophys. Acta - Mol. Cell Res. *1803*, 767–775.

Ott, M., Prestele, M., Bauerschmitt, H., Funes, S., Bonnefoy, N., and Herrmann, J.M. (2006). Mba1, a membrane-associated ribosome receptor in mitochondria. EMBO J. *25*, 1603–1610.

Ott, M., Amunts, A., and Brown, A. (2016). Organization and Regulation of Mitochondrial Protein Synthesis. Annu. Rev. Biochem. *85*, annurev-biochem-060815-014334.

Palade, G.E. (1955). A small particulate component of the cytoplasm. J. Cell Biol. *1*, 59–68.

Pearce, S.F., Rebelo-Guiomar, P., D'Souza, A.R., Powell, C.A., Van Haute, L., and Minczuk, M. (2017). Regulation of Mammalian Mitochondrial Gene Expression: Recent Advances. Trends Biochem. Sci. *42*, 625–639.

Perrin, R., Lange, H., Grienenberger, J.-M., and Gagliardi, D. (2004). AtmtPNPase is required for multiple aspects of the 18S rRNA metabolism in Arabidopsis thaliana mitochondria. Nucleic Acids Res. *32*, 5174–5182.

Peters, K., Belt, K., and Braun, H.-P. (2013). 3D Gel Map of Arabidopsis Complex I. Front. Plant Sci. *4*, 153.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera?A visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612.

Pfeffer, S., Woellhaf, M.W., Herrmann, J.M., and Förster, F. (2015). Organization of the mitochondrial translation machinery studied in situ by cryoelectron tomography. Nat. Commun. *6*, 6019.

Pinker, F., Bonnard, G., Gobert, A., Gutmann, B., Hammani, K., Sauter, C., Gegenheimer, P. a, and Giegé, P. (2013). PPR proteins shed a new light on RNase P biology. RNA Biol. *10*, 1457–1468.

Pinker, F., Schelcher, C., Fernandez-Millan, P., Gobert, A., Birck, C., Thureau, A., Roblin, P., Giegé, P., and Sauter, C. (2017). Biophysical analysis of Arabidopsis protein-only RNase P alone and in complex with tRNA provides a refined model of tRNA binding. J. Biol. Chem. *292*, 13904–13913.

Poole, A., and Penny, D. (2007). Engulfed by speculation. Nature *447*, 913–913.

Portereiko, M.F. (2006). NUCLEAR FUSION DEFECTIVE1 Encodes the Arabidopsis RPL21M Protein and Is Required for Karyogamy during Female Gametophyte Development and Fertilization. Plant Physiol. *141*, 957–965.

Poutre, C.G., and Fox, T.D. (1987). PET111, a Saccharomyces cerevisiae nuclear gene required for translation of the mitochondrial mRNA encoding cytochrome c oxidase subunit II. Genetics *115*, 637–647.

Prikryl, J., Rojas, M., Schuster, G., and Barkan, A. (2011). Mechanism of RNA stabilization and translational activation by a pentatricopeptide repeat protein. Proc. Natl. Acad. Sci. U. S. A. *108*, 415–420.

Pusnik, M., Small, I., Read, L.K., Fabbro, T., and Schneider, A. (2007). Pentatricopeptide repeat proteins in Trypanosoma brucei function in mitochondrial ribosomes. Mol. Cell. Biol. *27*, 6876–6888.

Ramrath, D.J.F., Niemann, M., Leibundgut, M., Bieri, P., Prange, C., Horn, E.K., Leitner, A.,

Boehringer, D., Schneider, A., and Ban, N. (2018). Evolutionary shift toward protein-based architecture in trypanosomal mitochondrial ribosomes. Science

Richter, R., Rorbach, J., Pajak, A., Smith, P.M., Wessels, H.J., Huynen, M.A., Smeitink, J.A., Lightowlers, R.N., and Chrzanowska-Lightowlers, Z.M. (2010). A functional peptidyl-tRNA hydrolase, ICT1, has been recruited into the human mitochondrial ribosome. EMBO J. *29*, 1116–1125.

Ringel, R., Sologub, M., Morozov, Y.I., Litonin, D., Cramer, P., and Temiakov, D. (2011). Structure of human mitochondrial RNA polymerase. Nature *478*, 269–273.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Roll-Mecak, A., Shin, B.S., Dever, T.E., and Burley, S.K. (2001). Engaging the ribosome: universal IFs of translation. Trends Biochem. Sci. *26*, 705–709.

Ruzzenente, B., Metodiev, M.D., Wredenberg, A., Bratic, A., Park, C.B., Cámara, Y., Milenkovic, D., Zickermann, V., Wibom, R., Hultenby, K., et al. (2012). LRPPRC is necessary for polyadenylation and coordination of translation of mitochondrial mRNAs. EMBO J. *31*, 443–456.

Sagan, L. (1967). On the origin of mitosing cells. J. Theor. Biol. *14*, 225-IN6.

Salinas-Giegé, T., Giegé, R., and Giegé, P. (2015). tRNA Biology in Mitochondria. Int. J. Mol. Sci. *16*, 4518–4559.

Salinas-Giegé, T., Cavaiuolo, M., Cognat, V., Ubrig, E., Remacle, C., Duchêne, A.-M., Vallon, O., and Maréchal-Drouard, L. (2017). Polycytidylation of mitochondrial mRNAs in Chlamydomonas reinhardtii. Nucleic Acids Res. 1–11.

Salinas, T., Duchêne, A.M., and Maréchal-Drouard, L. (2008). Recent advances in tRNA mitochondrial import. Trends Biochem. Sci. *33*, 320–329.

Salone, V., Rüdinger, M., Polsakiewicz, M., Hoffmann, B., Groth-Malonek, M., Szurek, B., Small, I., Knoop, V., and Lurin, C. (2007). A hypothesis on the identification of the editing enzyme in plant organelles. FEBS Lett. *581*, 4132–4138.

Santo-Domingo, J., and Demaurex, N. (2012). Perspectives on: SGP symposium on mitochondrial physiology and medicine: the renaissance of mitochondrial pH. J. Gen. Physiol. *139*, 415–423.

Sapp, J. (2005). The prokaryote-eukaryote dichotomy: meanings and mythology. Microbiol. Mol. Biol. Rev. *69*, 292–305.

Schaefke, B., Sun, W., Li, Y.-S., Fang, L., and Chen, W. (2018). The evolution of posttranscriptional regulation. Wiley Interdiscip. Rev. RNA *9*, e1485.

Schelcher, C., Sauter, C., and Giegé, P. (2016). Mechanistic and Structural Studies of Protein-Only RNase P Compared to Ribonucleoproteins Reveal the Two Faces of the Same Enzymatic Activity. Biomol. 2016, Vol. 6, Page 30 *6*, 30.

Schimper, A.F. (1883). Über die Entwicklung der Chlorophyllkörner und Farbkörper.

Schmitz-Linneweber, C., and Small, I. (2008). Pentatricopeptide repeat proteins: a socket set for organelle gene expression. Trends Plant Sci. *13*, 663–670.

Scolnick, E., Tompkins, R., Caskey, T., and Nirenberg, M. (1968). Release factors differing in specificity for terminator codons. Proc. Natl. Acad. Sci. U. S. A. *61*, 768–774.

Shadel, G.S. (2004). Coupling the mitochondrial transcription machinery to human disease. Trends Genet. *20*, 513–519.

Sharma, M.R., Koc, E.C., Datta, P.P., Booth, T.M., Spremulli, L.L., and Agrawal, R.K. (2003). Structure of the mammalian mitochondrial ribosome reveals an expanded functional role for its component proteins. Cell *115*, 97–108.

Sharma, M.R., Booth, T.M., Simpson, L., Maslov, D.A., and Agrawal, R.K. (2009). Structure of a mitochondrial ribosome with minimal RNA. Proc. Natl. Acad. Sci. U. S. A. *106*, 9637–9642.

Shen, C., Zhang, D., Guan, Z., Liu, Y., Yang, Z., Yang, Y., Wang, X., Wang, Q., Zhang, Q., Fan, S., et al. (2016). Structural basis for specific single-stranded RNA recognition by designer pentatricopeptide repeat proteins. Nat. Commun. *7*, 11285.

Shikanai, T., and Okuda, K. (2011). In Vitro RNA-Binding Assay for Studying Trans-Factors for RNA Editing in Chloroplasts. In Methods in Molecular Biology (Clifton, N.J.), pp. 199–208.

Shine, J., and Dalgarno, L. (1974). The 3'-terminal sequence of Escherichia coli 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. U. S. A. *71*, 1342–1346.

Shoji, S., Dambacher, C.M., Shajani, Z., Williamson, J.R., and Schultz, P.G. (2011). Systematic Chromosomal Deletion of Bacterial Ribosomal Protein Genes. J. Mol. Biol. *413*, 751–761.

Siira, S.J., Spåhr, H., Shearwood, A.-M.J., Ruzzenente, B., Larsson, N.-G., Rackham, O., and Filipovska, A. (2017). LRPPRC-mediated folding of the mitochondrial transcriptome. Nat. Commun. *8*, 1532.

De Silva, D., Tu, Y.-T., Amunts, A., Fontanesi, F., and Barrientos, A. (2015). Mitochondrial ribosome assembly in health and disease. Cell Cycle *14*, 2226–2250.

Sirrenberg, C., Bauer, M.F., Guiard, B., Neupert, W., and Brunner, M. (1996). Import of carrier proteins into the mitochondrial inner membrane mediated by Tim22. Nature *384*, 582–585.

Sloan, D.B., Alverson, A.J., Chuckalovcak, J.P., Wu, M., McCauley, D.E., Palmer, J.D., and Taylor, D.R. (2012). Rapid Evolution of Enormous, Multichromosomal Genomes in Flowering Plant Mitochondria with Exceptionally High Mutation Rates. PLoS Biol. *10*, e1001241.

Sloan, D.B., Wu, Z., and Sharbrough, J. (2018a). Correction of Persistent Errors in Arabidopsis Reference Mitochondrial Genomes. Plant Cell *30*, 525–527.

Sloan, D.B., Warren, J.M., Williams, A.M., Wu, Z., Abdel-Ghany, S.E., Chicco, A.J., and Havird, J.C. (2018b). Cytonuclear integration and co-evolution. Nat. Rev. Genet. *19*, 635–648.

van der Sluis, E.O., Bauerschmitt, H., Becker, T., Mielke, T., Frauenfeld, J., Berninghausen, O., Neupert, W., Herrmann, J.M., and Beckmann, R. (2015). Parallel Structural Evolution of Mitochondrial Ribosomes and OXPHOS Complexes. Genome Biol. Evol. *7*, 1235–1251.

Small, I.D., and Peeters, N. (2000). The PPR motif - a TPR-related motif prevalent in plant organellar proteins. Trends Biochem. Sci. *25*, 46–47.

Smith, D.R., Hua, J., and Lee, R.W. (2010). Evolution of linear mitochondrial DNA in three known lineages of Polytomella. Curr. Genet. *56*, 427–438.

Smits, P., Smeitink, J.A.M., van den Heuvel, L.P., Huynen, M.A., and Ettema, T.J.G. (2007). Reconstructing the evolution of the mitochondrial ribosomal proteome. Nucleic Acids Res. *35*, 4686–4703.

Soto, I.C., Fontanesi, F., Myers, R.S., Hamel, P., and Barrientos, A. (2012). A Heme-Sensing Mechanism in the Translational Regulation of Mitochondrial Cytochrome c Oxidase Biogenesis. Cell Metab. *16*, 801–813.

Sousa, J.S., Mills, D.J., Vonck, J., and Kühlbrandt, W. (2016). Functional asymmetry and electron flow

in the bovine respirasome. Elife *5*.

Sousa, J.S., D'Imprima, E., and Vonck, J. (2018). Mitochondrial Respiratory Chain Complexes. pp. 167–227.

Sousa, P.M.F., Videira, M.A.M., Santos, F.A.S., Hood, B.L., Conrads, T.P., and Melo, A.M.P. (2013). The bc:caa3 supercomplexes from the Gram positive bacterium Bacillus subtilis respiratory chain: A megacomplex organization? Arch. Biochem. Biophys. *537*, 153–160.

Staples, R.R., and Dieckmann, C.L. (1994). Suppressor analyses of temperature-sensitive cbp1 strains of Saccharomyces cerevisiae: the product of the nuclear gene SOC1 affects mitochondrial cytochrome b mRNA post-transcriptionally. Genetics *138*, 565–575.

Stehling, O., and Lill, R. (2013). The role of mitochondria in cellular iron-sulfur protein biogenesis: mechanisms, connected processes, and diseases. Cold Spring Harb. Perspect. Biol. *5*, a011312.

Stock, D. (1999). Molecular Architecture of the Rotary Motor in ATP Synthase. Science *286*, 1700–1705.

Stoll, B., and Binder, S. (2016). Two NYN domain containing putative nucleases are involved in transcript maturation in Arabidopsis mitochondria. Plant J. *85*, 278–288.

Stoll, B., Zendler, D., and Binder, S. (2014). RNA Processing Factor 7 and Polynucleotide Phosphorylase Are Necessary for Processing and Stability of nad2 mRNA in Arabidopsis Mitochondria. RNA Biol. *11*, 968–976.

Sun, T., Bentolila, S., and Hanson, M.R. (2016). The Unexpected Diversity of Plant Organelle RNA Editosomes. Trends Plant Sci. *21*, 962–973.

Szyrach, G., Ott, M., Bonnefoy, N., Neupert, W., Herrmann, J.M., and Stuart, R. (2003). Ribosome binding to the Oxa1 complex facilitates co-translational protein insertion in mitochondria. EMBO J. *22*, 6448–6457.

Taanman, J.-W. (1999). The mitochondrial genome: structure, transcription, translation and replication. Biochim. Biophys. Acta - Bioenerg. *1410*, 103–123.

Takenaka, M., Verbitskiy, D., van der Merwe, J.A., Zehrmann, A., and Brennicke, A. (2008). The process of RNA editing in plant mitochondria. Mitochondrion *8*, 35–46.

Temperley, R., Richter, R., Dennerlein, S., Lightowlers, R.N., and Chrzanowska-Lightowlers, Z.M. (2010). Hungry codons promote frameshifting in human mitochondrial ribosomes. Science *327*, 301.

Tomecki, R., Dmochowska, A., Gewartowski, K., Dziembowski, A., and Stepien, P.P. (2004). Identification of a novel human nuclear-encoded mitochondrial poly(A) polymerase. Nucleic Acids Res. *32*, 6001–6014.

Tyagi, S., Pande, V., and Das, A. (2014). Whole mitochondrial genome sequence of an Indian Plasmodium falciparum field isolate. Korean J. Parasitol. *52*, 99–103.

Unseld, M., Marienfeld, J.R., Brandt, P., and Brennicke, A. (1997). The mitochondrial genome of Arabidopsis thaliana contains 57 genes in 366,924 nucleotides. Nat. Genet. *15*, 57–61.

Uyttewaal, M., Mireau, H., Rurek, M., Hammani, K., Arnal, N., Quadrado, M., and Giegé, P. (2008). PPR336 is Associated with Polysomes in Plant Mitochondria. J. Mol. Biol. *375*, 626–636.

Vance, J.E. (2014). MAM (mitochondria-associated membranes) in mammalian cells: Lipids and beyond. Biochim. Biophys. Acta - Mol. Cell Biol. Lipids *1841*, 595–609.

Vanlerberghe, G. (2013). Alternative Oxidase: A Mitochondrial Respiratory Pathway to Maintain

Metabolic and Signaling Homeostasis during Abiotic and Biotic Stress in Plants. Int. J. Mol. Sci. *14*, 6805–6847.

Verschoor, A., Warner, J.R., Srivastava, S., Grassucci, R.A., and Frank, J. (1998). Three-dimensional structure of the yeast ribosome. Nucleic Acids Res. *26*, 655–661.

Wang, C., and Youle, R.J. (2009). The Role of Mitochondria in Apoptosis. Annu. Rev. Genet. *43*, 95.

Wang, S., Bai, G., Wang, S., Yang, L., Yang, F., Wang, Y., Zhu, J.-K., and Hua, J. (2016). Chloroplast RNA-Binding Protein RBD1 Promotes Chilling Tolerance through 23S rRNA Processing in Arabidopsis. PLOS Genet. *12*, e1006027.

Weisser, M., Schäfer, T., Leibundgut, M., Böhringer, D., Aylett, C.H.S., and Ban, N. (2017). Structural and Functional Insights into Human Re-initiation Complexes. Mol. Cell *67*, 447–456.e7.

Wende, S., Platzer, E.G., Jühling, F., Pütz, J., Florentz, C., Stadler, P.F., and Mörl, M. (2014). Biological evidence for the world's smallest tRNAs. Biochimie *100*, 151–158.

Wolters, J.F., Chiu, K., and Fiumera, H.L. (2015). Population structure of mitochondrial genomes in Saccharomyces cerevisiae. BMC Genomics *16*, 451.

Yagi, Y., Hayashi, S., Kobayashi, K., Hirayama, T., and Nakamura, T. (2013). Elucidation of the RNA Recognition Code for Pentatricopeptide Repeat Proteins Involved in Organelle RNA Editing in Plants. PLoS One *8*, e57286.

Yamashita, T., Nakamaru-Ogiso, E., Miyoshi, H., Matsuno-Yagi, A., and Yagi, T. (2007). Roles of bound quinone in the single subunit NADH-quinone oxidoreductase (Ndi1) from Saccharomyces cerevisiae. J. Biol. Chem. *282*, 6012–6020.

Yin, P., Li, Q., Yan, C., Liu, Y., Liu, J., Yu, F., Wang, Z., Long, J., He, J., Wang, H.-W., et al. (2013). Structural basis for the modular recognition of single-stranded RNA by PPR proteins. Nature *504*, 168–171.

Yonath, A. (2009). Large facilities and the evolving ribosome, the cellular machine for genetic-code translation. J. R. Soc. Interface *6*, S575–S585.

Yusupova, G., Jenner, L., Rees, B., Moras, D., and Yusupov, M. (2006). Structural basis for messenger RNA movement on the ribosome. Nature *444*, 391–394.

Zaremba-Niedzwiedzka, K., Caceres, E.F., Saw, J.H., Bäckström, Di., Juzokaite, L., Vancaester, E., Seitz, K.W., Anantharaman, K., Starnawski, P., Kjeldsen, K.U., et al. (2017). Asgard archaea illuminate the origin of eukaryotic cellular complexity. Nature *541*, 353–358.

Zeng, R., Smith, E., and Barrientos, A. (2018). Yeast Mitoribosome Large Subunit Assembly Proceeds by Hierarchical Incorporation of Protein Clusters and Modules on the Inner Membrane. Cell Metab. *27*, 645–656.e7.

Zeng, X., Hourset, A., and Tzagoloff, A. (2006). The Saccharomyces cerevisiae ATP22 Gene Codes for the Mitochondrial ATPase Subunit 6-Specific Translation Factor. Genetics *175*, 55–63.

Ziaja, K., Michaelis, G., and Lisowsky, T. (1993). Nuclear Control of the Messenger RNA Expression for Mitochondrial ATPase Subunit 9 in a New Yeast Mutant. J. Mol. Biol. *229*, 909–916.

Zimmer, C. (2009). On the Origin of Eukaryotes. Science *325*, 666–668.

Résumé en français

Résumé de thèse en français

Introduction

La traduction mitochondriale fait l'objet d'un intérêt considérable car elle combine des caractéristiques bactériennes associées à des traits spécifiques ayant évolué dans les cellules eucaryotes. La mitochondrie est un sujet d'étude crucial pour la santé humaine, car son dysfonctionnement est associé à un nombre croissant de maladies héréditaires et est impliqué dans des maladies courantes telles que les maladies neurodégénératives, les cardiomyopathies, le syndrome métabolique, le cancer et l'obésité (Nunnari and Suomalainen, 2012; De Silva et al., 2015). Chez les végétaux, la mitochondrie a également provoque aussi un intérêt considérable car elle spécifie un trait largement étendu conduisant à une incapacité des plantes à produire du pollen fonctionnel, appelé «stérilité mâle cytoplasmique». Les mitochondries végétales présentent donc un intérêt agronomique considérable (Chen and Liu, 2014).

La synthèse des protéines, ou traduction, est le processus final par lequel l'information génétique intermédiaire (ARNm) est décodée pour produire des protéines. La traduction est réalisée par les ribosomes, machines moléculaires omniprésentes dans le vivant qui lisent les ARNm et catalysent l'assemblage des acides aminés pour former une chaîne polypeptidique. Même si les ribosomes sont omniprésents dans tous les organismes vivants, leurs structures et leurs compositions, ainsi que les facteurs qui contribuent à la traduction, sont extrêmement différents. Les ribosomes sont particulièrement divergents entre les procaryotes et les eucaryotes, mais d'autres ribosomes sont également trouvés dans les mitochondries et les chloroplastes, nécessaires pour effectuer la traduction des ARNm encore codés par les génomes organellaires. Alors que le ribosome chloroplastique ressemble fortement aux ribosomes bactériens (Boerema et al., 2018), les ribosomes mitochondriaux, ou mitoribosomes, qui font l'objet de cette thèse, ont divergé de manière significative par rapport à leurs homologues bactériens au cours de l'évolution des eucaryotes. Chez l'homme et la levure, de récentes études ont montré que les ribosomes mitochondriaux étaient radicalement différents, non seulement de ceux trouvés chez les procaryotes et dans les cytosols des eucaryotes, mais qu'ils étaient également différents entre eux, chacun des mitoribosomes de l'homme et de la levure faisant intervenir des composants spécifiques (Bieri et al., 2018; Desai et al., 2017; Greber et al., 2015).

Chez les plantes, le mécanisme de traduction ainsi que la composition du mitoribosome demeurent particulièrement méconnus. Son ribosome implique des sous-unités spécifiques et son

initiation ne fait pas appel à des séquences de type Shine-Dalgarno. Il est donc probable que des facteurs en *trans* reconnaissant des séquences spécifiques soient impliqués dans l'initiation de la traduction et le recrutement des ribosomes. Ces facteurs pourraient appartenir à de nouvelles familles de protéines modulaires de liaison à l'ARN, telles que les protéines à « pentatricopeptide repeat » (PPR). Ces dernières composent une énorme classe de protéines liant l'ARN, omniprésentes chez les eucaryotes et impliquées dans une grande variété de mécanismes post-transcriptionnels (Gutmann et al., 2012a). Les protéines PPR sont principalement localisées dans les mitochondries et les chloroplastes (Barkan and Small, 2014). Elles ont été trouvé comme agissant dans tous les mécanismes post-transcriptionnels tels que le processing, l'édition de l'ARN, l'épissage, la stabilité ou encore la traduction d'ARN spécifiques au niveau des organelles (Schmitz-Linneweber and Small, 2008). De plus, le mitoribosome animal contient deux protéines PPRs, l'une d'entre-elles étant impliqué dans le recrutement de l'ARNm messager, permettant ainsi l'initiation de la traduction (Kummer et al., 2018).

Mon sujet de thèse principal porte sur la caractérisation de l'appareil traductionnel des mitochondries d'Arabidopsis. Le travail s'est organisé en deux parties. La première partie a consisté à mettre au point un protocole adéquat de purification des ribosomes mitochondriaux. Le but étant d'obtenir des échantillons suffisamment purs pour pouvoir les analyser à la fois en protéomique, pour déterminer de façon précise la composition du mitoribosome d'Arabidopsis, et par cryo-microscopie électronique afin d'obtenir une structure à haute résolution du ribosome mitochondrial d'*A.thaliana*.

La seconde partie a consisté à caractériser les facteurs protéiques spécifiques associés au ribosome mitochondrial d'*A.thaliana.* Il s'agit en particulier des protéines PPR, dont des études précédentes et des résultats préliminaires suggèrent une association aux ribosomes mitochondriaux. D'autres facteurs protéiques, ne faisant pas partis de la famille des protéines PPR ont également été identifié. Pour caractériser ses facteurs j'ai analysé de nombreuses lignées mutantes pour identifier leurs rôles. J'ai également analysé l'influence d'un facteur en particulier, rPPR1, sur la traduction mitochondriale par « ribosome profiling ».

En plus de mon travail sur les mitoribosomes, j'ai également participé à d'autres projets du laboratoire, portant notamment sur la caractérisation fonctionnelle de protéines PRORPs chez le parasite *Plasmodium falciparum* et le nématode *Romanomermis culicivorax*. J'ai aussi été impliqué dans la caractérisation de la protéine MNU2, un partenaire protéique de la PRORP mitochondriale d'*Arabidopsis thaliana.*

Résultats

Au cours de ma thèse, j'ai tout d'abord établi un protocole efficace de purification de mitochondries d'*Arabidopsis thaliana* à partir de cultures cellulaires, ainsi qu'à partir d'inflorescence de plantes. J'ai ensuite pu mettre au point un protocole de purification de ribosomes mitochondriaux, pour permettre à la fois leur analyse protéomique et leur analyse structurale. Travailler à partir de cultures cellulaires permet d'obtenir des mitochondries en grande quantité et de façon rapide, ce qui m'a permis de tester de nombreuses conditions de purification de ribosomes.

Pour vérifier la pureté des échantillons, les fractions finales de purification contenant les ribosomes mitochondriaux ont été analysés par spectrométrie de masse en collaboration avec Philippe Hamman (Plateforme protéomique, Strasbourg-Esplanade), mais également directement sur microscope électronique à transmission en collaboration avec Yaser Hashem (IBMC, Strasbourg puis INSERM Bordeaux). De plus un suivie de la purification a été réalisé à l'aide d'anticorps permettant soit la détection d'une protéine ribosomale cytosolique soit d'une protéine ribosomale mitochondriale.

Pour me permettre d'analyser de façon efficace la composition du ribosome mitochondrial d'Arabidopsis, j'ai créé une lignée de plante complémentée exprimant avec une version étiquetée d'une de mes protéines candidates, PPR336 renommée rPPR1, ayant déjà été caractérisée comme interagissant avec le ribosome mitochondrial (Uyttewaal et al., 2008).

J'ai utilisé cette lignée pour purifier des mitochondries et réaliser des co-immuno-précipitation de rPPR1HA. La co-immuno-précipitation a permis de confirmer que rPPR1 était un composant du mitoribosome d'Arabidopsis, associé à la petite sous-unité.

En croisant les données de protéomiques obtenues par purification biochimique classique avec mes données de co-immuno-précipitation, j'ai pu obtenir une liste complète des composants protéiques du mitoribosome d'Arabidopsis. Cela m'a permis d'établir que le mitoribosome est constitué de 81 protéines différentes, dont 19 sont spécifiques des plantes. Parmi ces protéines, 10 font partie de la famille des protéines PPR. En séparant biochimiquement petite et grande sous-unités du ribosome, j'ai pu attribuer chacune des protéines identifiées à leur sous-unité respective.

L'analyse par cryo-électron microscopie des mitoribosomes d'Arabidopsis a permis de déterminer une structure à 16A de résolution de ce dernier (Figure 1). Même si la structure n'est pas d'assez haute résolution pour y identifier tous les composants, cela a permis de révéler les caractéristiques structurales uniques du mitoribosome de plantes. En particulier, il se distingue

par une très grande petite sous-unité, plus grande que la grande sous-unité. La petite sous-unité présentant une extension caractéristique très importante au niveau de la tête (> 200Å). Cette extension est due à une insertion de 370 nt dans l'ARN ribosomique 18S, constituant un tout nouveau domaine. La comparaison de l'architecture du mitoribosome d'Arabidopsis avec celui d'animaux et de levure a révélé qu'il était significativement plus grand que ces-derniers, notamment pour la petite sous-unité qui présente une architecture unique.

Pour la seconde partie de mon projet j'ai travaillé à la caractérisation des mutants PPR trouvés comme faisant partis du mitoribosome. Pour cela, j'ai travaillé avec dix lignées de plantes simples mutantes knock-out générées par insertion d'ADN-T. J'avais également à ma disposition un double mutant, issu du croissement de deux des simples mutants, qui était déjà disponible au laboratoire. C'est ce double mutant qui a été mon outil de travail principal. Originellement nommé *336/336L,* puis renommé *rPPR1/336L,* ce double mutant a un phénotype de retard de croissance par rapport aux plantes sauvages et aux simples mutants. Ce double mutant a été utilisé pour la création de la lignée rPPR1HA qui a permis la caractérisation de protéome du mitoribosome d'Arabidopsis. L'analyse des mutants d'insertion a montré que l'effet de ces mutations pouvait avoir un effet plus ou moins important sur la plantes. Pour certaines PPR les mutations sont létales, certaines sont stériles, d'autres affectées dans leur croissances et pour certaines il n'y a pas d'effet.

Pour analyser plus finement le rôle de ces protéines j'ai également cloné la majorité d'entre-elles. Cela permettra plus tard de confirmer la localisation subcellulaire de toutes les protéines PPR étudiées nouvellement identifiées mais également de créer d'autres lignées de plantes exprimant des protéines étiquetées. De plus j'ai également réalisé le même travail pour une partie des protéines spécifiques du mitoribosome de plantes ne faisant pas parties de la famille des protéines PPR.

En collaboration avec Hakim Mireau de l'INRA de Versailles j'ai également pu réaliser une analyse de profilage des ribosomes sur le mutant rPPR1, me permettant ainsi de conclure sur son rôle global dans la traduction mitochondrial, bien que le mutant ne présente pas de phénotype macroscopique très marqué.

Conclusions

Ainsi mon travail a permis de montrer que le ribosome mitochondrial d'Arabidopsis est complètement différent de tous les ribosomes connus jusqu'à présent, que ce soit en termes de composition ou de structure. Bien que d'origine bactérienne, ces ribosomes ont significativement et spécifiquement divergé de leur ancêtre procaryote. Par exemple, chez les plantes, les mitoribosomes ont une énorme SSU caractérisée par une grande extension au niveau de la tête, dû à des modifications majeures des ARNr et ont recruté de nombreuses protéines spécifiques des eucaryotes, parmi lesquelles 10 sont des protéines PPR.

Florent WALTZ

Characterization of the mitochondrial translation apparatus of *Arabidopsis thaliana*

Résumé

Dans les cellules eucaryotes, différents types de ribosomes coexistent. Les ribosomes mitochondriaux synthétisent les quelques protéines codées par l'ADN mitochondrial, qui sont essentielles au fonctionnement de l'organisme. Ces ribosomes sont particulièrement divergents des ribosomes procaryotes, mais sont également très différents entre les eucaryotes. Mon travail de thèse s'est concentré sur la caractérisation de la structure et de la composition en protéines du ribosome mitochondrial de la plante modèle *Arabidopsis thaliana*. Des approches biochimiques complémentaires ont permis d'identifier 19 protéines uniquement trouvées dans le mitoribosome de plante, parmi lesquelles 10 sont des protéines PPR, des protéines particulièrement abondantes chez les plantes. Les mutations des gènes codant pour ces PPR ribosomales (rPPR) mènent à l'apparition de phénotypes macroscopiques distincts, notamment une létalité ou des retards de croissance importants. L'analyse moléculaire du mutant *rppr1* par profilage des ribosomes, ainsi que l'analyse du taux de protéines mitochondriales, révèlent que la protéine rPPR1 est un facteur de traduction générique, ce qui constitue une nouvelle fonction des protéines PPR. De plus, la cryo-électron microscopie a été utilisée pour déterminer l'architecture tridimensionnelle de ce mitoribosome. Cette approche a révélé la structure unique du mitoribosome de plante, caractérisée par une très grande petite sous-unité ribosomale ayant un domaine additionnel jamais décrit jusqu'à présent. Globalement, mes résultats ont montré que le mitoribosome d'Arabidopsis est complètement différent des ribosomes bactériens et des autres mitoribosomes eucaryotes, à la fois en terme de structure mais aussi de composition, permettant ainsi de mieux comprendre l'évolution de ce composant central de l'expression génétique.

Mots-clés: protéines PPR, mitochondrie, traduction, mitoribosome, *Arabidopsis thaliana*

Résumé en anglais

Ribosomes are the molecular machines translating the genetic information carried by mRNA into protein. Different translation machineries co-exist in eukaryote cells. While cytosolic translation is comparatively well characterized, it remains the most elusive step of gene expression in mitochondria. In plants, while numerous pentatricopeptide repeat (PPR) proteins are involved in all steps of gene expression, their function in translation remains unclear. My work focused on the biochemical characterisation of Arabidopsis mitochondrial ribosomes and the identification of its protein composition. Complementary biochemical approaches identified 19 plant specific mitoribosome proteins, among which 10 are PPR proteins. The knock out mutations of ribosomal PPR (rPPR) genes result in distinct macroscopic phenotypes including lethality or severe growth delays. The molecular analysis of rPPR1 mutants, using ribosome profiling, as well as the analysis of mitochondrial protein levels, revealed that rPPR1 is a generic translation factor, which is a novel function for PPR proteins. Finally, single particle cryo-electron microscopy was used and revealed the unique structural architecture of Arabidopsis mitoribosomes, characterised by a very large small ribosomal subunit, larger than the large subunit, with a novel head domain. Overall, my results showed that Arabidopsis mitoribosomes are completely distinct from bacterial and other eukaryote mitoribosomes, both in terms of structure and of protein content.

Keywords: PPR proteins, mitochondria, translation, mitoribosome, *Arabidopsis thaliana*