



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
Facultad de Ciencias Biológicas
Programa de Doctorado en Ciencias Biológicas
Mención Genética Molecular y Microbiología

**IDENTIFICACIÓN Y CARACTERIZACIÓN DE GENES DE PROTEÍNAS
CON REPETICIONES DE PENTATRICOPÉPTIDOS (PPR) NECESARIAS
PARA LA EDICIÓN DE TRANSCRITOS MITOCONDRIALES EN
*Arabidopsis thaliana***

ANA MARIBEL ARENAS MIRANDA

AGOSTO 2015



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
Facultad de Ciencias Biológicas
Programa de Doctorado en Ciencias Biológicas
Mención Genética Molecular y Microbiología

**IDENTIFICACIÓN Y CARACTERIZACIÓN DE GENES DE PROTEÍNAS CON
REPETICIONES DE PENTATRICOPÉPTIDOS (PPR) NECESARIAS PARA LA
EDICIÓN DE TRANSCRITOS MITOCONDRIALES EN *Arabidopsis
thaliana***

Tesis presentada a la Pontificia Universidad Católica de Chile como parte de
los requisitos para optar al grado de Doctor en Ciencias Biológicas mención
Genética Molecular y Microbiología

Por

ANA MARIBEL ARENAS MIRANDA

Director de Tesis: Dr. Xavier Jordana de Buen

Comisión de Tesis: Dra. Loreto Holuigue
Dra. Lee Meisel
Dr. Patricio Arce

AGOSTO 2015

Dedicado a

M. V. B., M. P. y M. A.

*“Aquí teneis mis flores frías
mi simulada sumisión
a los dictámenes del viento
porque yo sobrevivo al agua,
a la sal, a los pescadores,”*

-Alga, Pablo Neruda

AGRADECIMIENTOS

Primero que todo, quiero agradecer al profe Xavier Jordana, mi tutor, por su apoyo y por la autonomía que me dio para elegir mi proyecto. Valoro mucho lo aprendido junto a él, su pulcritud, rigurosidad y ética científica me acompañaran siempre. Gracias por permitirme formar parte de su laboratorio y por enseñarme que se puede realizar investigación en la “frontera de la ciencia” y es posible mover ese límite cuando se trabaja con mucho esfuerzo y dedicación. Sin duda el mejor tutor que pude haber elegido.

A los Dres. Mizuki Takenaka y Axel Brennicke en Alemania, y a todo el laboratorio del Intitut für Molekulare Botanik de la Universität Ulm (Däniil, Bárbara, Anja, Bianca, Mathias, Sachi) por recibirme en su laboratorio con toda la disposición y energía positiva para trabajar y por enseñarme que en la ciencia no existen barreras de idiomas o de nacionalidades, si se trabaja con pasión y amor por lo que se hace.

Chabelita!! el “hágalo simple, hágalo fácil” me acompañará siempre. Gracias por escucharme y por tu apoyo en los momentos mas oscuros. “Quien mucho da...mucho recibe” y así eres tú Chabe todo el cariño que entregas a las personas y pones en tu trabajo se devolverá a ti con creces. Te quiero mucho Chabelita y gracias por todo!

Al Seba “ mi pequeño saltamonte” que al fin creció y ahora brilla con luz propia. Gracias por apoyarme y por creer en este proyecto. No tengo dudas que tendrás un brillante futuro como científico.

A los chicos del lab: Rasta, FranKo, Pauli, Peter, Esteban, Kiko gracias por la compañía y los momentos vividos y por hacer del lab un segundo hogar para mí. A mi querido Cris, con quien compartí tantas clases y horas de laboratorio, gracias por tu disposición a ayudar y a encontrar una solución a los problemas del lab y por ser mi soporte MAC oficial ;). Eres un científico brillante, sólo que un tanto disperso jejeje ...pero lejos el que tiene mas estilo del piso con tus inolvidables zapatillas pintadas con plumón. Te quiero Cris, éxito en todo!

Quiero agradecer a todo “el lab de plantas” por la buena onda y cariño de tantos años (Gigi, Grace, Fran, Lore, Sari, Vero, Dario, Nati, Pauli, Jóse, Trini). Porque para mi las barreras y paredes no existían, siempre me sentí parte de un sólo “gran lab de plantas”. A mis queridos Evars, Aldo y Ariel!!! Gracias por mostrarme muchas veces la luz al final del Túnel (literalmente!!) por la complicidad, carretes y alegrías vividas, los quiero mucho!

A mis queridos amigos Luis y Tati gracias por los consejos, las copuchas, el apoyo y los cafés compartidos a las 16:30 de cada tarde! que ahora extraño tanto!! Aunque la vida nos lleve por distintos rumbos, se que nuestra amistad y cariño se mantendrá.

A mis queridos amigos del ya legendario “equipo ANIP”: Katia, Lucy, Caro, Ro, Jon, Ricardo, Tomás, Carlos, Renzo, Fernando compañeros de tantas reuniones y luchas. Con quienes aprendí que los sueños pueden volverse realidad y que es posible cambiar el

mundo si se aúnan voluntades y organización. Ha sido un orgullo trabajar con ustedes. El límite es el cielo.. por ello no tengo dudas que de aquí saldrá... (en un futuro ya no tan lejano...) algún día el futuro ministro(a) del Min. de C y T que tanto anhelamos... ¿Por qué no??

A mis “amigos de toda la vida” Jek, Kathy y Rod, gracias por la amistad y el cariño mutuo. Los años pasan, pero seguimos siendo los mismos que corríamos por la facultad, Estoy segura que no habrá libreta (o servilleta) que aguante todas las historias que aún nos quedan por vivir ...los quiero mucho.

A mis queridas amigas Nonchis, Vivi, Marivix y Jaco gracias por el apoyo constante, por escucharme cuando mas lo necesité, por su amistad y cariño sincero.

A mi familia Rosita, Maggy, tía Nana, tía Gloria por su preocupación y apoyo constante.

Gracias Don Ángel, Sra. Mary y Miguel Ángel por los años de amor vividos, por permitirme ser parte de su familia y porque gran parte la persona que ahora soy es gracias al afecto y amor que siempre me entregaron, los llevo siempre en mi corazón.

A mi Mamá gracias por estar siempre junto a mí, por apoyarme y por enseñarme a ser una mujer fuerte y a salir adelante frente a la adversidad. A Vicente y la Ale por su cariño y apoyo. A “mis Soles” Maite y Mayra gracias por existir, por iluminar mi vida día a día y por darme una razón para seguir adelante.

A Javier gracias por tu amor y compañía y por ayudarme a sacar adelante la parte mas difícil de este proyecto: “terminar la tesis”!! y por darme la oportunidad de emprender nuevas aventuras juntos... en un lugar maravilloso al Sur del Mundo. ☺

FINANCIAMIENTO

La realización de esta tesis contó con el financiamiento de la Beca CONICYT para Estudios de Doctorado en Chile, de la Beca de Apoyo a la Tesis Doctoral (AT24100161), de los proyectos FONDECYT-CONICYT (1100601 y 1141197), Núcleo Milenio en Genómica Funcional de Plantas (P06-009-F y P10-062-F) y del apoyo de la Vicerrectoría de Investigación y Doctorado de la Pontificia Universidad Católica de Chile (VRID).

ABREVIATURAS

ATP: Adenosina 5' trifosfato

ABRC: Arabidopsis Biological Resource Center

BSA: Albúmina de suero bovino

β -ME: β -mercaptoetanol

CaMV: Virus del mosaico de la coliflor

CDS: Secuencia codificante

Col 0: Columbia 0

dATP: desoxiadenosina 5' trifosfato

dCTP: desoxicitidina 5' trifosfato

dGTP: desoxiguanosina 5' trifosfato

dTTP: desoxitimidina 5' trifosfato

ddNTP: 2',3'-didesoxinucleótidos 5' trifosfato

$[\alpha\text{-}^{32}\text{P}]$ -dCTP: $[\alpha\text{-}^{32}\text{P}]$ -desoxicitidina 5' trifosfato

DNA: Ácido desoxirribonucleico

DEPC: dietilpirocarbonato

DNasa: desoxirribonucleasa

EDTA: Ácido etilendiaminotetraacético

eGFP: “Green fluorescent protein”, proteína fluorescente verde

GABI-kat: Genomanalyse im biologischen-Kölner Arabidopsis T-DNA lines

GUS: β -glucuronidasa.

kb: kilobases

MEF: Mitochondrial editing factor, factor de edición mitocondrial

MES: Ácido 2-(N-morfolino) etanosulfónico

M-MLV: Virus de la leucemia murina de Moloney

nt: nucleótido

pb: pares de bases

PCR: Reacción en cadena de la polimerasa

PPR: Pentatricopeptide repeat, repetición de pentatricopéptido

RNA: Ácido ribonucleico

RNasa: ribonucleasa.

rpm: revoluciones por minuto.

RT-PCR: reverse transcription-polymerase chain reaction

SAIL: Syngenta Arabidopsis Insertional Library.

SDS: Dodecilsulfato de sodio

SIGnAL: Salk Institute Genomic Analysis Laboratory

SNAPshot: Single-nucleotide primer extension

Tris-HCl: clorhidrato de tris(hidroximetil)aminometano.

T-DNA: DNA de transferencia

Temed: Tetrametiletilendiamina

TBE: Tris/Borato/EDTA

RESUMEN

La edición de RNA es un tipo de procesamiento post-transcripcional que altera la identidad de nucleótidos en las moléculas de RNA. En organelos de plantas con flores, la edición de RNA cambia citosinas a uracilos. Más de 400 sitios en mitocondrias de angiospermas son individualmente identificados por proteínas altamente específicas, pero ninguno de estos factores están codificados en el genoma mitocondrial. Todos estos factores de reconocimiento pertenecen a la familia de proteínas con repeticiones de pentatricopéptido (PPR). El principal objetivo de este trabajo es identificar nuevos Factores de Edición Mitocondrial (MEFs). Para lograr este objetivo, se utilizaron los siguientes criterios: (i) Disponibilidad de secuencias de ESTs o cDNAs de genes PPR en bases de datos, (ii) destinación mitocondrial predicha por distintos algoritmos, (iii) proteínas PPR pertenecientes a la subfamilia PLS con motivos E, E+ o DYW, y (iv) reducida identidad de secuencia con otras proteínas PPR de *Arabidopsis*. Utilizando estos criterios, cinco genes PPR candidatos fueron seleccionados. Tres de nuestros candidatos, en experimentos de localización subcelular muestran localización en mitocondrias. Plantas mutantes homocigotas en cada uno de los genes candidatos fueron analizadas con la aproximación de SNaPshot. Esta metodología permite evaluar masivamente los niveles de edición en organelos de *Arabidopsis*, encontrando alteraciones en la edición de diversos transcritos mitocondriales en tres líneas mutantes analizadas. Desde ese momento las proteínas codificados por los genes candidatos fueron nombradas MEF25, MEF26 y MEF31. MEF25 participa en la edición de un sitio en el transcrito *nad1*. MEF26 es necesaria para la edición completa del sitio *cox3-311* y de la edición parcial del sitio *nad4-166*. MEF31 facilita la edición de un sitio en el transcrito *orfX*. Los defectos en la edición en las mutantes *mef25*, *mef26* and *mef31* fueron confirmados por secuenciación de productos de RT-PCR, análisis de una segunda línea mutante o ensayos de complementación. La edición de los sitios diana de estos factores MEF genera un cambio en la naturaleza del aminoácido que finalmente estará presente en la proteína funcional. Sin embargo, el fenotipo de las plantas mutantes *mef25*, *mef26* and *mef31* es indistinguible de las plantas silvestres, sugiriendo que la edición de sus sitios diana no es esencial al menos bajo condiciones estándar de crecimiento. Por otra parte, recientemente ha sido propuesto un código para el reconocimiento del RNA mediado por las proteínas PPR. Usando este código, los dianas predichos para los factores MEF25, MEF26 y MEF31 son coincidentes con los dianas experimentales identificados en este estudio.

ABSTRACT

RNA editing is a post-transcriptional process that alters the identity of nucleotides in RNA molecules. In organelles of flowering plants, RNA editing changes C nucleotides to U nucleotides. More than 400 editing sites in mitochondria of angiosperms are individually addressed by specific proteins, but none of these factors are encoded in the mitochondrial genome. These recognition factors belong to the pentatricopeptide repeat (PPR) family. The main goal of this work is to identify new Mitochondrial Editing Factor (MEFs). To achieve this goal, we used the following criteria: (i) PPR genes with ESTs or cDNA sequences in public databases, (ii) mitochondrial localization predicted by different subcellular algorithms, (iii) PPR members of the PLS subfamily with E, E+ or DYW motifs and (iv) low sequence similarity with other Arabidopsis PPR proteins. Using these criteria, five PPR genes were selected. Three of our candidates showed mitochondrial localization in subcellular localization assays. Homozygous mutant plants for each candidate were analyzed by the SNaPshot approach. This methodology enabled us to massively evaluate the editing levels in Arabidopsis organelles and we found alterations in several mitochondrial transcripts in three mutant lines analyzed. We named the proteins encoded by these candidate genes as MEF25, MEF26 and MEF31. MEF25 is necessary for the edition of one site in the *nad1* transcript. MEF26 is required for the full edition at the *cox3-311* site and the partial editing in the *nad4-166* site. MEF31 facilitates the edition at one site in *orfX* transcript. The editing defects in the *mef25*, *mef26* and *mef31* mutants were confirmed by sequencing of RT-PCR products, analysis of a second mutant line or complementation assay. The target editing sites of these MEF factors generates a change in the nature of the amino acid present in the functional protein. However, the phenotype of *mef25*, *mef26* and *mef31* mutants are indistinguishable from wild type plants, suggesting that the affected editing sites are not essential under standard growth conditions. On the other hand, it has been recently proposed a code for the RNA recognition sequence by the PPR proteins. Using this code, the best predicted targets for MEF25, MEF26 and MEF31 factors are coincident with the experimental targets identified in this study.

ÍNDICE

AGRADECIMIENTOS 5

FINANCIAMIENTO 7

ABREVIATURAS 8

RESUMEN 10

ABSTRACT 11

ÍNDICE 12

CAPÍTULO I 17

INTRODUCCIÓN 17

HIPÓTESIS 34

OBJETIVOS 34

CAPÍTULO II 35

RESULTADOS 35

II.1. Identificación de proteínas PPR candidatas que podrían participar en la edición de transcritos mitocondriales en *Arabidopsis thaliana* 36

II. 2. Caracterización Funcional del Factor de Edición Mitocondrial 25 (MEF25) 39

II. 2.1. Introduction 42

II. 2.2. Materials and Methods 43

II. 2.2.1. Molecular and phenotypic analysis of *mef25* mutant plants. 43

II. 2.2.2. Analysis of RNA editing sites 44

II. 2.2.3. Constructs for MEF25 localization experiments and promoter analysis 44

II. 2.3. Results. 45

II. 2.3.1. At3g25060 encodes an E+-PPR protein targeted to mitochondria	45
II. 2.3.2. MEF25 is required for one mitochondrial RNA editing site	48
II. 2.3.3. Phenotype and pattern of nad1 transcripts are not affected in mef25 mutants	48
II. 2.4. Discussion	55
II. 2.5. Acknowledgments	61
II. 2.6. References	63

II. 3. Caracterización Funcional del Factor de Edición Mitocondrial 26 (MEF26)

68

II. 3.1. Introduction	71
II. 3.2. Materials and methods	72
II. 3.2.1. Genotyping of mef26 mutant plants	72
II. 3.2.2. Screening for affected RNA editing sites and evaluation of editing efficiency	73
II. 3.2.3. MEF26 subcellular localization analysis	74
II. 3.2.4. Complementation of the editing defects in <i>mef26</i> mutant plants	74
II. 3.2.5. Northern blot analysis	75
II. 3.2.6. Prediction of potential MEF26 target editing sites	75
II. 3.3. Results	75
II. 3.3.1. At3g03580 encodes a DYW-PPR protein targeted to mitochondria	75
II. 3.3.2. AT3G03580 (MEF26) is involved in RNA editing of two mitochondrial transcripts	76
II. 3.3.3. Cis-elements of the two MEF26 target sites have similarity	82

II. 3.3.4. Putative MEF26 orthologs	86
II. 3.4. Discussion	93
II. 3.4.1. MEF26 is involved in editing two sites in different mitochondrial transcripts	93
II. 3.4.2. Edited sites targeted by MEF26 share similar cis-elements	98
II. 3.4.3. Conclusions	102
II. 3.5. Acknowledgments	103
II. 3.6. References	104

II. 4. Caracterización Funcional del Factor de Edición Mitocondrial 31 (MEF31)

110

II. 4. Metodología	112
II. 4.1. Crecimiento de <i>Arabidopsis thaliana</i> .	112
II. 4.2. Extracción de ácidos nucleicos de <i>Arabidopsis thaliana</i> .	112
II. 4.2.1. Extracción de DNA genómico.	112
II. 4.2.2. Extracción de RNA total	113
II. 4.3. Técnicas de DNA recombinante	113
II. 4.3.1. Reacción en cadena de la Polimerasa (PCR)	113
II. 4.3.2. Síntesis de cDNA	114
II. 4.3.3. Electroforesis de DNA en geles de agarosa.	114
II. 4.3.4. Clonamiento de fragmentos de DNA en plásmidos.	115
II. 4.3.5. Transformación bacteriana.	115
II. 4.3.6. Extracción de DNA plasmidial.	117
II. 4.3.7. Secuenciación y alineamiento de secuencias.	118

II. 4.4. Generación de plantas transgénicas de <i>Arabidopsis thaliana</i> .	118
II. 4.4.1. Generación de construcciones en vectores binarios.	118
II. 4.4.2. Transformación de <i>Arabidopsis</i> por el método de la inmersión floral.	119
II. 4.4.3. Selección de líneas transgénicas.	120
II. 4.5. Análisis de la localización sub-celular de MEF31	120
II. 4.5.1. Obtención de protoplastos.	120
II. 4.5.2. Análisis de microscopía	121
II. 4.6. Genotipificación de plantas mutantes <i>mef31</i> .	122
II. 4.7. Análisis de la edición del RNA.	122
II. 4.7.1. Secuenciación directa de productos de RT-PCR.	122
II. 4.7.2. Secuenciación de clones.	122
II. 4. Resultados	124
II. 4.1. El gen At2g46050 codifica para una proteína PPR mitocondrial.	124
II. 4.2. Caracterización de una línea mutante en el gen At2g46050	127
II. 4.3. Fenotipo molecular de las plantas mutantes <i>mef31</i> : defectos en la edición de un transcrito mitocondrial.	130
II. 4.4. Los aminoácidos en las posiciones 6 y 1' de los motivos PPR de MEF31 predicen de acuerdo al código de las PPRs que el transcrito <i>orfX</i> es su RNA blanco.	135
II. 4.5. Análisis de la edición de sitios cercanos a <i>orfX-581</i> en plantas mutantes <i>mef31</i>	139
II. 4. Discusión	143
II. 4. Conclusiones	149
<u>CAPÍTULO III</u>	150

DISCUSIÓN Y CONCLUSIONES GENERALES 150

1.- Criterios de selección de proteínas PPR candidatas. 152

2.- Identificación de los factores de edición MEF25, MEF26 y MEF31 154

3.- La edición de los sitios diana de MEF25, MEF26 y MEF31 es necesaria para la incorporación de aminoácidos conservados en las proteínas pero no para el desarrollo de las plantas, al menos en condiciones estándar de crecimiento. 155

Conclusiones Generales 159

CAPÍTULO IV 161

REFERENCIAS BIBLIOGRÁFICAS 161

CAPÍTULO I

INTRODUCCIÓN

I.- Introducción general

Las mitocondrias son organelos semiautónomos de las células eucariontes que tienen un papel metabólico fundamental ya que producen gran parte de la energía que una célula requiere. Llevan a cabo este importante rol gracias a dos procesos metabólicos que tienen lugar en este organelo: el ciclo de Krebs (ciclo del ácido cítrico) y la fosforilación oxidativa. Esta última vía funciona a través de cuatro complejos respiratorios que generan un gradiente de protones, que es finalmente utilizado por el complejo ATP sintasa (complejo V) para la generación de ATP. Sin embargo se han descrito importantes diferencias entre las mitocondrias de plantas y las de otros organismos eucariontes, entre ellas la existencia de vías metabólicas especiales que les permiten participar en los procesos de fotorespiración y síntesis del folato, la presencia de varias NADH deshidrogenasas distintas del complejo I de la cadena respiratoria y de una oxidasa alternativa a la citocromo c oxidasa (AOX) (Welchen E. y cols., 2010).

Se ha estimado que el proteoma mitocondrial de plantas consta de entre 2000 y 3000 productos génicos (Millar y cols., 2003). Sin embargo el genoma mitocondrial de plantas no contiene más de 60 genes que codifican para un set de aproximadamente 30 proteínas, repartidas entre subunidades de los complejos respiratorios y proteínas ribosomales, además codifica para 15-20 RNAs de transferencia (tRNAs) y 3 RNAs ribosomales (rRNAs). Los genomas de mitocondrias de plantas también contienen algunos marcos de lectura abiertos denominados “ORF” (**O**pen **R**ead **F**rame) cuyas funciones génicas aún no han sido descritas del todo (Unsold y cols., 1997). Es muy probable que algunos de estos ORFs fueron generados por rearrreglos genómicos mitocondriales y se considera el hecho de que estos ORFs sean transcritos y editados como una evidencia de que son funcionales.

Es importante señalar que a pesar del reducido número de proteínas codificadas en este organelo, estas proteínas corresponden mayoritariamente a subunidades esenciales para la función de los

complejos de la cadena respiratoria. Los genes mitocondriales están usualmente dispersos a través de todo el genoma organelar, dando origen a transcritos primarios tanto mono como bicistrónicos. Estos transcritos son sometidos a diversas etapas de procesamiento que incluyen la edición del RNA (editing) y la remoción de intrones del grupo II. Se suma a ello la necesidad de la estabilización de algunos transcritos para garantizar un correcto procesamiento y traducción (Figura 1). Dado que el genoma de mitocondrial no codifica para la maquinaria necesaria para la expresión de sus transcritos, incluyendo los procesos de modificación post-transcripcional de éstos, la maquinaria para estas etapas está constituida por productos génicos codificados en el núcleo (Binder y Brennicke, 2003). Estas proteínas deben ser importadas a la mitocondria en donde algunas de ellas son ensambladas en los complejos protéicos, mientras que otras actúan como factores facilitadores de dicho ensamblaje. Este proceso conocido como “biogénesis mitocondrial” requiere de una estrecha y coordinada comunicación “núcleo – mitocondria”, en donde el procesamiento post-transcripcional de los transcritos organelares ha sido sugerido como una de las principales etapas de regulación.

Edición del RNA

Uno de los más importantes y sobrerrepresentado procesamiento post-transcripcional en mitocondrias de plantas es el enigmático proceso de la edición de los transcritos organelares. Dicho proceso implica necesariamente un cambio en la información nucleotídica a nivel de RNA respecto de la información contenida en el DNA genómico. La edición de RNA ha sido identificada en todas las embriofitas (plantas terrestres) excepto en *Marchantia*, y no ocurre en el clado basal de las algas verdes. En el caso de *Marchantia* la explicación mas parsimoniosa para la ausencia de edición es la pérdida secundaria de este tipo de procesamiento, ya que en su clado hermano sí se han encontrado sitios de edición en el transcriptoma mitocondrial de las especies que lo componen (Malek y cols. 1996).

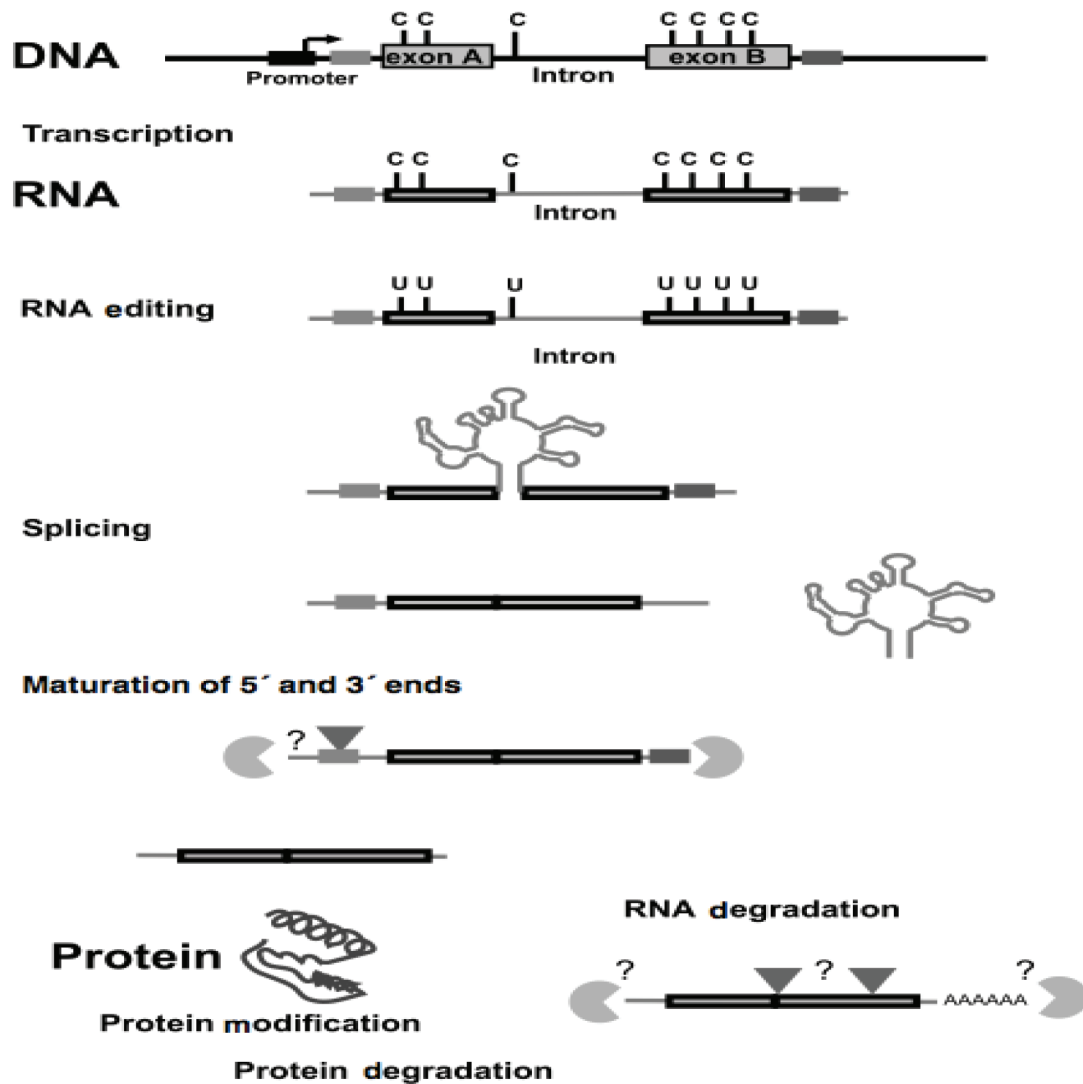


Figura 1. Expresión génica en mitocondrias de plantas. Una compleja serie de etapas de procesamiento post-transcripcional es necesaria para la maduración y generación de transcritos que estén en condiciones adecuadas para ser traducidos. Luego de la transcripción, los transcritos mitocondriales experimentan la edición de citosinas específicas (RNA editing), la remoción de intrones (splicing) y la generación de estructuras secundarias en sus extremos 5' y 3' como parte de su proceso de maduración. Posteriormente, la degradación del RNA mediada por endo y/o exoribonucleasas que es probablemente iniciada por la poliadenilación en el extremo 3' de los mRNA (triángulos grises: sitios de corte endonucleolítico). Modificado desde Gagliardi y Binder (2007).

En angiospermas, la edición de RNAs organelares involucra exclusivamente cambios de citosinas por uracilos (C→U) (Covello y Gray, 1989; Hiesel y cols., 1989). En cloroplastos han sido descritos 34 sitios de edición (Chateigner-Boutin y Small, 2007), mientras que en mitocondrias de *Arabidopsis* más de 450 sitios son editados (Bentolli y cols., 2008; Giege y Brennicke, 1999; Zehrmann y cols., 2008). La mayoría de estos cambios ocurre en regiones codificantes de los RNAs mensajeros y con menor frecuencia se encuentran en intrones o tRNAs. La expresión génica en mitocondrias de plantas terrestres necesita de la edición de sus transcritos ya que este procesamiento es esencial para la generación de codones de inicio de la traducción, la corrección de codones que codifican para aminoácidos conservados o la generación de codones de término de la traducción (Takenaka y cols., 2008) (Figura 2).

Las razones para la existencia y establecimiento de la edición de transcritos organelares en plantas continúan siendo enigmáticas, no obstante en los últimos años se ha realizado un importante progreso en el entendimiento de cómo la maquinaria de edición es capaz de reconocer citosinas (C) específicas en los transcritos, a pesar de que no se han descrito secuencias consenso cuando se comparan diferentes elementos en *cis* en los transcritos mitocondriales. Sin embargo gracias a experimentos *in vitro* e *in organello* se ha logrado acotar una región *cis* de entre -20 y +6 nucleótidos respecto al sitio de edición como necesaria y suficiente para la edición de un sitio (Choury y cols., 2004; Farré y Araya., 2001; Takenaka y cols., 2004)

Una pregunta que aún queda pendiente es la identidad de la enzima que catalizaría la edición de los sitios. Hasta ahora es aceptada la hipótesis de que se trataría de una reacción de desaminación y no de una transaminación la que mediaría la transformación de citosinas a uracilos, ya que en experimentos *in vitro* no se han identificado putativas moléculas receptoras de los grupos aminos que pudiesen participar en transaminaciones (Takenaka y cols., 2008). Los restantes componentes del “editosoma de plantas” han sido parcialmente develados durante los últimos años a través de aproximaciones genéticas, así como de análisis bioquímicos en extractos plastidiales usando entrecruzamiento UV. Se han logrado detectar factores protéicos que tienen

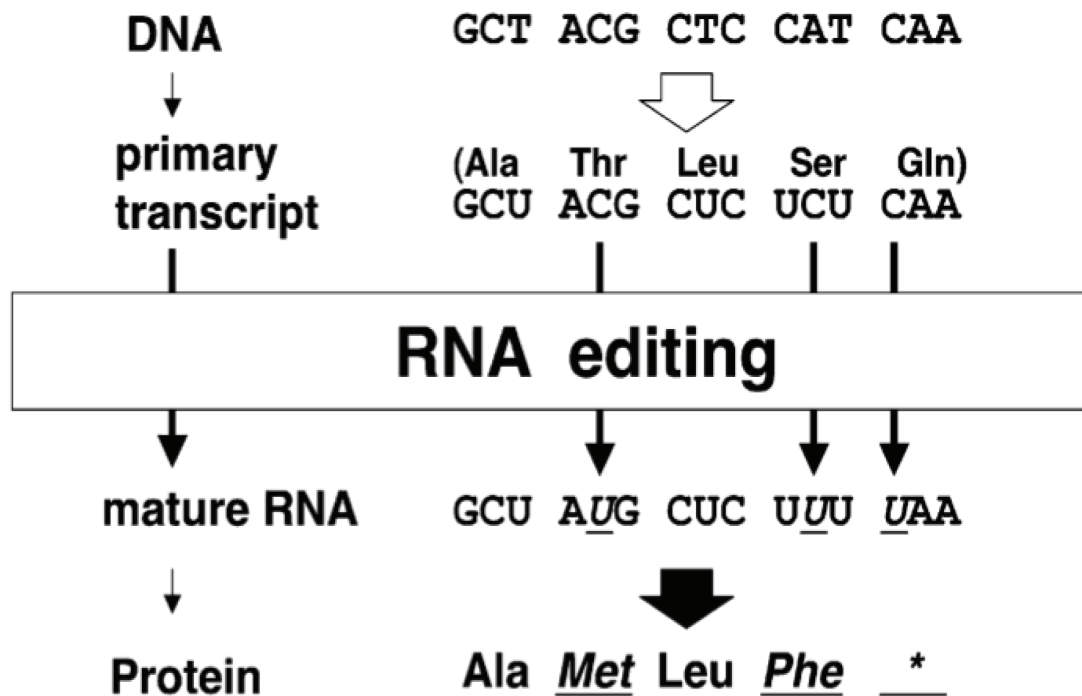


Figura 2. La edición de RNA en organelos. La edición de RNA en mitocondrias de angiospermas es un procesamiento post-transcripcional en donde citosinas específicas (C) del transcrito primario son modificadas a uracilos (U) en el RNA mensajero maduro. La secuencia aminoacídica codificada por el transcrito maduro completamente editado es distinta de la secuencia proteica codificada por el DNA genómico y el transcrito primario. En una secuencia hipotética se muestran diversos ejemplos de las consecuencias de la alteración de codones a raíz de la edición del RNA: Creación de codones de inicio de la traducción (ACG → AUG), de término de la traducción (CAA → UAA, CGA → UGA, CAG → UAG) y “corrección” de codones que codifican para aminoácidos conservados (UCU que codifica para serina (*Ser*) en UUU que codifica para *una* fenilalanina (*Phe*) conservada, por ejemplo). Modificado desde Takenaka y cols., 2008.

la capacidad de interactuar con RNA y que mayoritariamente pertenecen a una muy particular familia génica que codifica para las proteínas con repeticiones de pentatricopéptido “PPR” (Lurin y cols., 2004). Más recientemente, nuevos factores generales de editing han sido descritos. Éstos son capaces de interactuar con proteínas de tipo PPR y pertenecen a una segunda familia de proteínas, denominada MORF por “Multiple Organellar RNA editing factors” (Takenaka y cols., 2012) o RIP por “RNA-editing factor interacting protein ” (Bentolila y cols., 2012). Sin embargo, mientras las proteínas PPRs son factores de especificidad de la edición, que reconocen a las citosinas que deben ser editadas (como se verá mas adelante), la función de las proteínas MORF continúa siendo desconocida.

Familia de las proteínas PPR

Con la secuenciación del genoma de *Arabidopsis* se hizo evidente la existencia de una familia génica muy numerosa que codifica para aproximadamente 450 proteínas PPR (Lurin y cols., 2004; Small y Peters, 2002). Los genomas de otras plantas terrestres codifican para un número aún mayor de ellas, como es el caso de *Populus* y *Vitis* con 600 genes de proteínas PPR cada uno (Jaillon y cols., 2007; Tuskan y cols., 2006). No obstante estas impresionantes cifras, los genes de proteínas PPR están ampliamente distribuidos en los eucariontes, pero su número es mucho menor fuera de las plantas (Lurin y cols., 2004; O’Toole y cols., 2008). Hasta ahora, no se han encontrado genes de proteínas con motivos PPR en procariontes (Lurin y cols., 2004; Schmitz-Linneweber y Small, 2008). Esta familia de proteínas ha sido definida por la presencia de motivos canónicos de 35 aminoácidos, repetidos en tándem desde 2 hasta 30 veces, a través de los cuales podrían unir secuencias específicas de RNA. En el trabajo realizado por Lurin y cols., (2004), estas proteínas fueron clasificadas de acuerdo a las características de los motivos PPR y a la presencia o ausencia de dominios adicionales en el carboxilo terminal. Así se dividió a la familia en dos subfamilias: la subfamilia P y la subfamilia PLS. La subfamilia P está formada por proteínas PPR que presentan sólo motivos PPR canónicos, es decir de 35 aminoácidos. La

subfamilia PLS está formada por proteínas PPR que presentan repeticiones de tipo P (canónicos) y además repeticiones de tipo S de 31-32 aminoácidos y repeticiones de tipo L1 y L2 de 35 a 40 aminoácidos, con variaciones en la secuencia. Dentro de esta subfamilia es posible definir diferentes clases, dependiendo de si las proteínas presentan motivos adicionales en el extremo C-terminal: proteínas PPR tipo PLS sin dominios adicionales en el carboxilo terminal, proteínas PPR tipo PLS con un dominio E adicional en el C-terminal, proteínas PPR tipo PLS con dos dominios adicionales E y E+, y proteínas PPR tipo PLS con tres dominios adicionales E, E+ y DYW (Lurin y cols., 2004) (Figura 3).

Las proteínas PPR han sido propuestas como adaptadores moleculares que se unirían al RNA y facilitarían actividades del metabolismo del RNA organelar. En el trabajo de Lurin y cols., (2004) se evaluaron *in silico* los péptidos de destinación predichos de todas las proteínas PPR descritas en *Arabidopsis*, y se encontró que la mayoría de las proteínas de la subfamilia P (canónicas) tendrían destinación mitocondrial, en cambio las de la subfamilia PLS serían destinadas a plastidios o a mitocondrias (Lurin y cols., 2004). Hasta el momento, todos los roles fisiológicos confirmados de las proteínas PPR en plantas son mitocondriales o cloroplásticos, existiendo hasta la fecha sólo dos excepciones: la proteína PPR de *Arabidopsis* GPR23, que interactuaría con la RNA polimerasa II en el núcleo (Ding y cols., 2006) y la proteína PPR PNM1 que presenta localización dual en mitocondria y núcleo en *Arabidopsis* (Hammani y cols., 2011).

La primera PPR descrita en la edición de un transcrito organelar fue la proteína CRR4 (chlororespiratory reduction 4). Esta proteína PPR es esencial para la edición del codón de inicio del transcrito cloroplástico *ndhD* (Kotera y cols., 2005). De allí en adelante, entre los años 2005 y 2009, otras 14 proteínas PPR fueron descritas como necesarias para la edición de distintos transcritos plastidiales, todas ellas miembros de las subclases E o DYW (Cai y cols., 2009; Chateigner-Boutin y cols., 2008; Hammani y cols., 2009; Okuda y cols., 2009; Robbins y cols., 2009; Yu y cols., 2009; Zhou y cols., 2008). Este exponencial aumento en la caracterización de

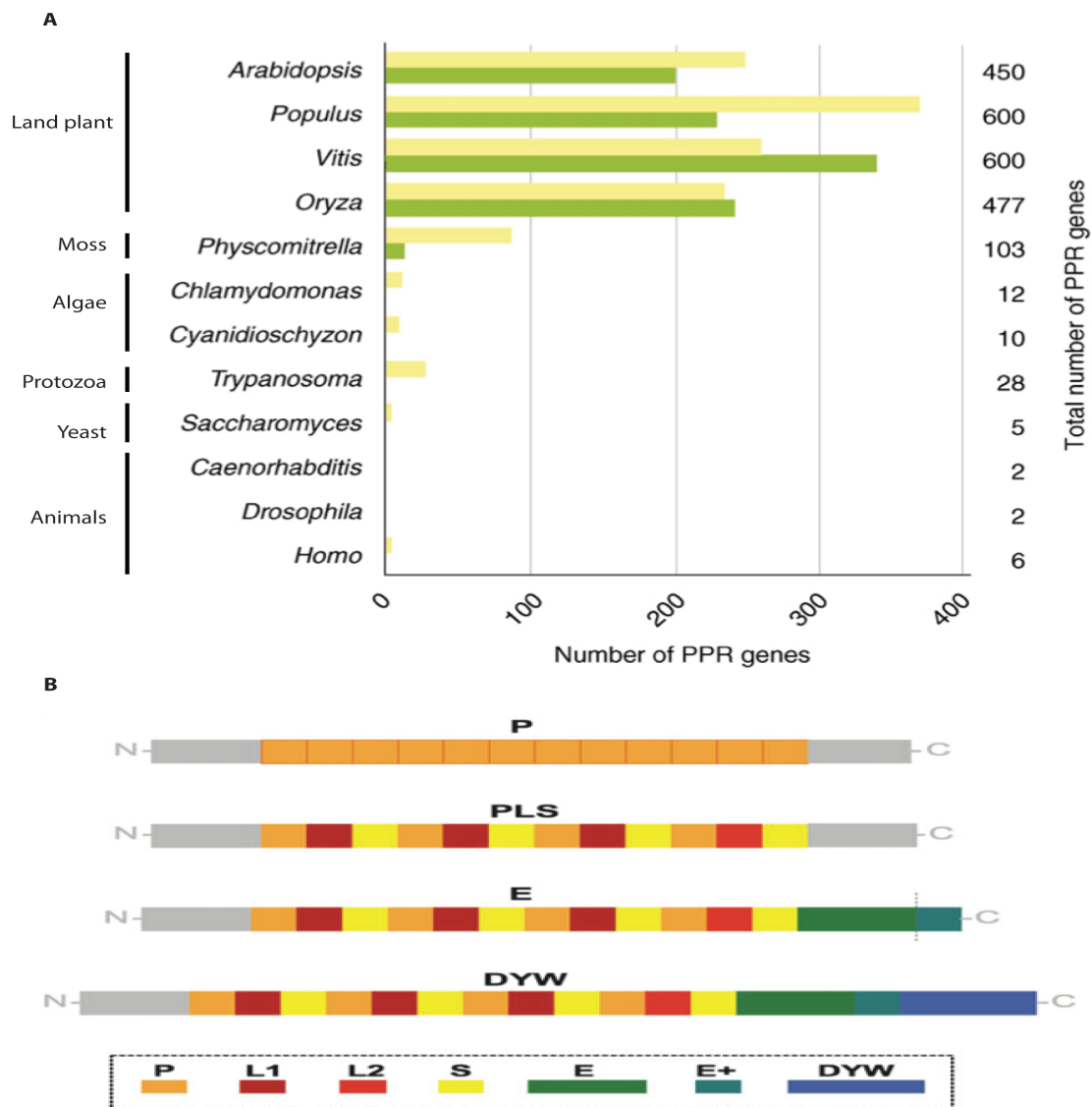


Figura 3. Distribución de los genes PPR en eucariontes y estructura de las proteínas PPR. (A) Se muestra la distribución de los genes de PPR en eucariontes. Las barras amarillas representan el número de genes de proteínas PPR tipo P y las barras verdes el número de genes de proteínas PPR de la subfamilia PLS. A la derecha en el eje vertical se muestra el número total de genes PPR en cada especie. Modificado desde Schmitz-Linneweber y Small, 2008. (B) Representación esquemática de los motivos PPR definidos por Lurin y cols., (2004) para cada subclase. El número y orden de las repeticiones pueden variar dependiendo de cada proteína. No obstante, se ha observado que numerosas proteínas poseen arreglos que mantienen la organización P-L-S. La línea punteada entre los dominios E y E+ señala que existen algunos ejemplos en los cuales E+ no está presente. Modificado desde O'Toole y cols., 2008.

proteínas PPR implicadas en la edición plastidial en desmedro de las PPRs involucradas en la edición de transcritos mitocondriales, es explicado básicamente por el reducido número de sitios de edición a analizar en el cloroplasto (sólo 34 sitios en *Arabidopsis*) y a que muchas de las mutantes en estos genes de proteínas PPR cloroplásticas presentan fenotipos característicos de albinismo clororespiratorio. Sin embargo, en el año 2009 Zehrmann y cols. describen el primer “Factor de Edición Mitocondrial” (MEF1), una proteína PPR de la subfamilia PLS que participa en la edición de tres mRNAs mitocondriales en *Arabidopsis*: *rps4*, *nad7* y *nad2* (Zehrmann y cols., 2009). Este grupo desarrolló una novedosa estrategia denominada “SNAPshot” (Takenaka y Brennicke, 2009) para el análisis masivo de los sitios de edición mitocondriales, contando con varios paneles de amplificación y secuenciación para examinar más de 350 sitios de edición en pocos ensayos (Figura 4). Por otra parte, ese mismo año se demostró que OGR1 (opaque and growth retardation 1), una proteína PPR mitocondrial de arroz perteneciente a la subclase DYW era esencial para la edición de siete sitios en cinco transcritos mitocondriales distintos (Kim y cols., 2009). La mayoría de las proteínas PPRs implicadas en la edición mitocondrial son necesarios para la edición de transcritos que codifican para subunidades de los complejos de la cadena transportadora de electrones “ETC” (complejos I, III y IV) o para proteínas implicadas en la biogenesis de estos complejos. No obstante, algunos factores PPR tales como REME1, MEF11, MEF14 y SLO2 (Bentolia y cols., 2010; Verbitskiy y cols., 2010; Verbitskiy y cols., 2011; Zhou y cols., 2012) han sido descritos como necesarios para la edición de transcritos de genes con funciones “putativas” como es el caso de los genes *orf*. Como ejemplos de estos genes en el genoma mitocondrial de *Arabidopsis* están el gen *matR*, que codificaría para una maturasa de intrones, y *orfX* (*mttB*) por “Membrane Targeting and Translocation”, que codificaría para un análogo mitocondrial de la proteína bacteriana *tatC* (van der Merve y Dubery, 2007).

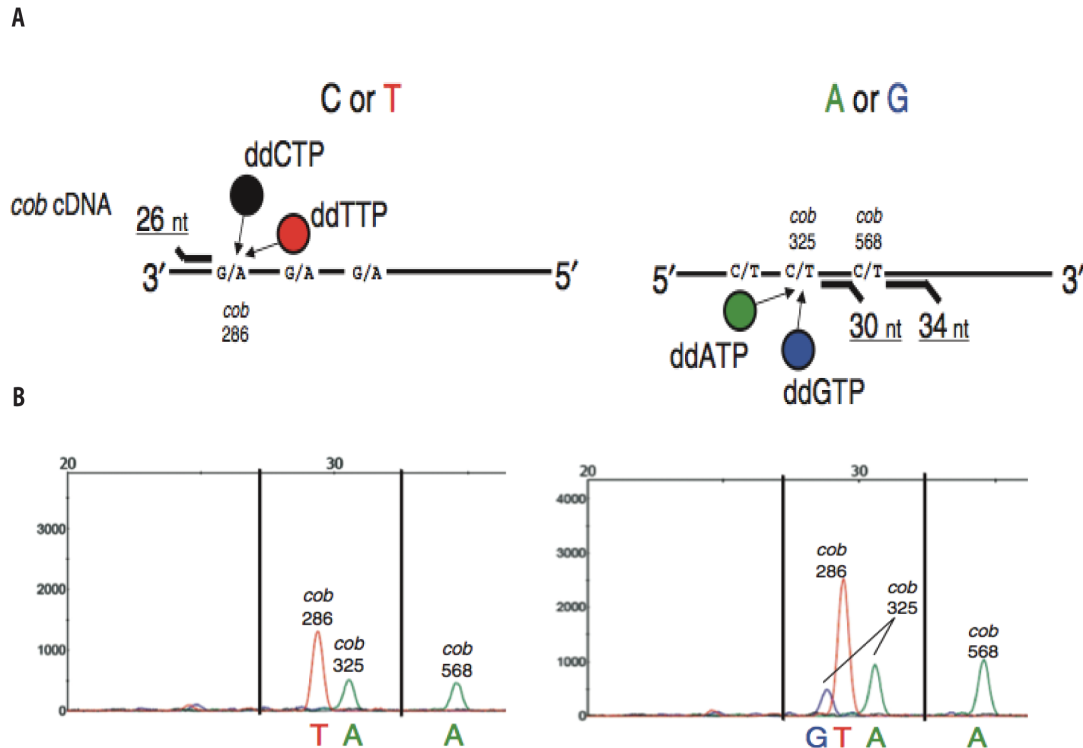


Figura 4. Análisis masivo de sitios de edición organelares utilizando la metodología de SNAPshot. (A) Se diseñan partidores específicos para cada sitio de edición que se quiere analizar (fragmentos de 26 nt, 30 nt y 34 nt), con la característica que hibridan justo un nucleótido antes de la citosina a editar, para luego realizar una reacción de extensión en la que se incorporará un didesoxinucleótido marcado con un fluoróforo para cada uno de los siguientes pares ddCTP/ddTTP (para la hebra anti-sentido) o ddATP/ddGTP (para la hebra sentido) dependiendo de si este sitio fue editado o no. Los sets de partidores en cada panel de sitios a analizar presentan temperaturas de hibridación similares y diferencias de tamaño de 4 nucleótidos entre ellos. Se muestran en la representación algunos sitios de edición del transcrito mitocondrial que codifica para citocromo b (*cob-286*, *cob-325* y *cob-568*) con sus respectivos partidores y el tamaño en nucleótidos (nt) del producto de extensión esperado. Los sitios están numerados con respecto a la adenina (A) +1 del codón de inicio de la traducción ATG. (B) Se observan las señales de fluorescencia obtenidas para cada uno de los nucleótidos marcados que fueron incorporados en los diferentes sitios de edición *cob-286*, *cob-325* y *cob-568*, utilizando como molde cDNA de

plantas silvestres (panel izquierdo) y cDNA de plantas mutantes que presentan alteraciones en la edición del sitio *cob-325* (panel derecho). Modificado desde Takenaka y Brennicke, 2009. La deficiencia parcial en la edición del sitio *cob-325* es detectada porque el oligonucleótido correspondiente es extendido con una ddGTP (C no editada en la hebra complementaria), además con una ddATP (T editada en la hebra complementaria). Estos resultados permiten concluir que en las plantas mutantes existe una reducción en la edición del sitio *cob-325* ya que este sitio esta siendo editado parcialmente. Cabe hacer notar que a pesar de que los oligonucleótidos extendidos con ddGTP o ddATP tienen el mismo tamaño migran de forma diferente, debido a su composición nucleotídica diferente.

En resumen hasta el inicio de este trabajo de tesis, todos los factores de tipo PPR implicados en la edición de distintos transcritos tienen en común pertenecer a la subfamilia PLS y la presencia de dominios adicionales de tipo E, E+ o DYW en el extremo C-terminal. Además, la expansión de la subfamilia PLS en plantas esta correlacionada con la presencia de la edición como un tipo de procesamiento de RNA en organelos (Figura 3) (Fuji y Small , 2011).

Por otra parte, se ha sugerido que el dominio DYW, presente en algunas PPR de la subfamilia PLS, podría tener la actividad citidina desaminasa en la edición de los transcritos. Esta proposición está basada en análisis filogenéticos que correlacionan la edición del RNA organelar en plantas con la presencia de aminoácidos conservados en este dominio y en el sitio catalítico de enzimas citidina deaminasa de otros organismos (O'Toole y cols., 2008; Rudinger y cols., 2008; Salone y cols., 2007). Sin embargo, hasta ahora no ha sido posible demostrar que alguna proteína PPR con dominio DYW tenga actividad desaminasa. Hasta este punto, parece indiscutible el involucramiento de proteínas PPR en el proceso de edición de transcritos organelares. No obstante, quedan preguntas pendientes sobre los mecanismos moleculares a través de los cuales estos factores reconocen los sitios de edición. Es por ello que ejemplos de nuevos factores sitio específicos podrían ayudar a entender este proceso. Además, al inicio de este trabajo de tesis muy pocos factores mitocondriales implicados en la edición habían sido descritos, por lo que se decidió identificar nuevas proteínas PPR que fueran esenciales para la edición de transcritos mitocondriales, y eventualmente caracterizar sus mecanismos de acción. Para ello se seleccionaron proteínas PPR de acuerdo a ciertos criterios (descritos en resultados) con el objetivo de aumentar las posibilidades de que fueran proteínas PPR mitocondriales que estuvieran implicadas en la edición de transcritos organelares, y de que se observaran defectos en la edición en mutantes nulas de los genes que codifican para ellas.

Recientemente se han realizado importantes avances para entender las bases que gobiernan el reconocimiento del RNA a través de los motivos PPRs. Utilizando aproximaciones computacionales, experimentales y la información obtenida a partir de la caracterización de más

de una decena de proteínas PPR descritas durante los últimos años (Bentolila y cols., 2010; Ohtani y cols., 2010; Sung y cols., 2010; Takenaka y cols., 2010; Tasaki y cols., 2010; Hammani y cols., 2011; Uchida y cols., 2011; Rüdinger y cols., 2011; Verbitskiy y cols., 2012; Murayama y cols., 2012; incluidos algunos de los resultados publicados de esta tesis), es que se ha sugerido como las repeticiones de PPR estarían mediando la unión secuencia específica a su RNA blanco en una relación 1 : 1 entre un motivo PPR y un nucleótido. Se ha propuesto así un código de reconocimiento en que ciertos pares de aminoácidos, específicamente el sexto residuo de un motivo PPR y el primer aminoácido del motivo siguiente, muestran preferencias por ciertos nucleótidos (Tabla 1). El código propuesto en primera instancia por Barkan y colaboradores en el 2012, señala a los aminoácidos en la posiciones 6 y 1' (definido como el primer aminoácido de la siguiente repetición) de los motivos PPR de tipo P y S como los participantes en la interacción directa con el RNA (Figura 5) y a los motivos L como “espaciadores” entre los elementos de tipo P y S generando así una organización de tipo PLS en las proteínas (Rivals y cols., 2006; Barkan y cols., 2012).

En el caso de las PPR implicadas en la edición de transcritos estos nucleótidos están situados río arriba de la citosina a editar en el elemento *cis* necesario y suficiente para la edición (Barkan y cols., 2012; Takenaka y cols., 2013). El nucleótido -4 con respecto al sitio a editar sería reconocido por el primer motivo PPR ubicado hacia el carboxilo terminal de la proteína, y luego los motivos PPR hacia el extremo amino terminal irían reconociendo consecutivamente los nucleótidos desde la posición -5 en dirección al extremo 5' del transcrito. En general, el número de nucleótidos potencialmente reconocidos por la proteína PPR correspondería al número de motivos PPR presentes en ella (Barkan y cols., 2012). Este mecanismo de reconocimiento modular propuesto para las proteínas PPR, que difiere de los mecanismos previamente descritos para la interacción “proteína – RNA”, ha sido denominado “código de las PPRs” y abre de paso insospechadas posibilidades no sólo para la predicción de posibles RNAs blanco de una proteína PPR en particular (Takenaka y cols., 2013) sino que también plantea la posibilidad real de diseñar

proteínas PPR a medida que pudiesen interactuar con RNAs de interés en aplicaciones biotecnológicas (Barkan y cols., 2012; Barkan y Small, 2014).

Tabla 1. Combinaciones de aminoácidos (AA) en motivos PPR que determinan especificidad de unión con nucleótidos específicos.

Motif type	AA 6	AA 1'	Base	
			Correlation ^a	Experimental ^b
P, S	T	N	A	A >>> G,C,U
P, S	S	N	A	
P	S	S	A	
P	N	N	C	C = U >>> G,A
P	N	S	C	C > U >>> G,A
P, S	T	D	G	G >>> A,C,U
S	S	D	G	
P, S	N	D	U	U > C >>> G,A
L	P	D	U	

(a) Se utilizaron mas de 40 proteínas PPR descritas en literatura con sus respectivos RNAs diana (Barkan y cols., 2012; Takenaka y cols., 2013; Yagi y cols., 2013), desde las cuales se identificaron y alinearon aproximadamente 1000 motivos PPR presentes en ellas, con lo que se generó una matriz de frecuencias aminoacídicas v/s identidad nucleotídicas de sus RNA dianas estableciendo las correlaciones mostradas. (b) La correspondencia experimental se basa en afinidades relativas a ensayos *in vitro* de unión de variantes de la proteína PPR10 con RNAs modificados (Barkan y cols., 2012). AA 6: sexto aminoácido de un motivo PPR. AA 1': el primer aminoácido del motivo siguiente. Modificado desde Barkan y Small, 2014.

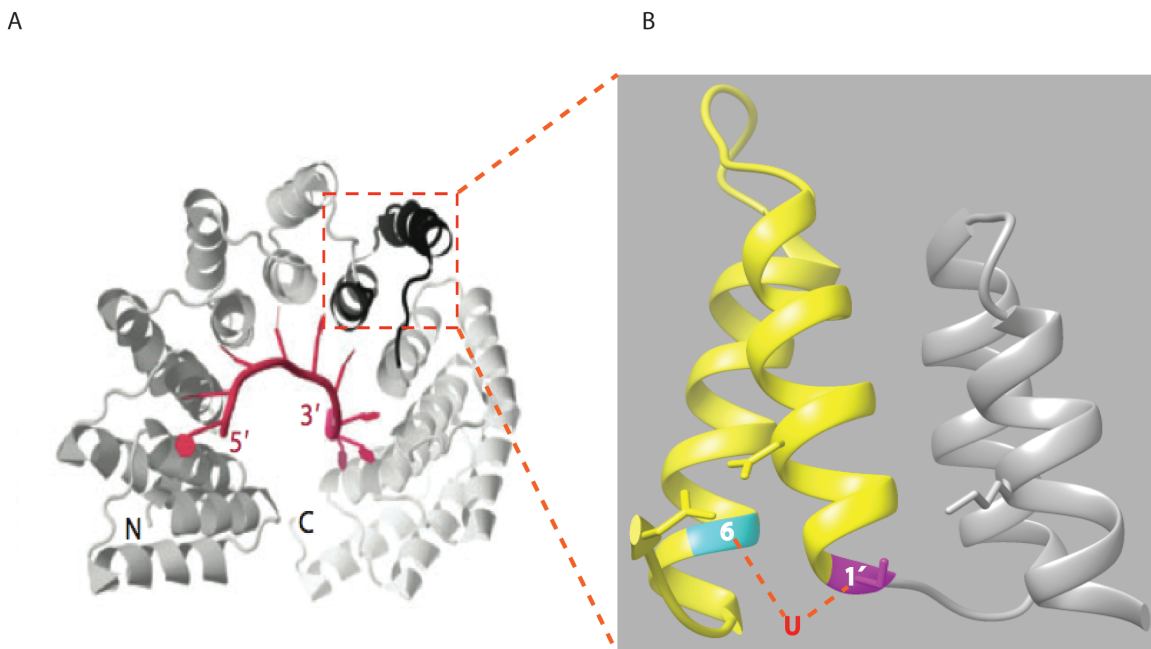


Figura 5. Estructura de los motivos PPR y modelo de unión de dominios PPR a su RNA blanco. Cada repetición (motivo PPR) está formada por un par de hélices α -antiparalelas conectadas por un pequeño “loop”. (A) Modelo de unión propuesto para una proteína “PPR con su RNA blanco”, ejemplificado con 10 motivos PPR (cada repetición en distintos tonos de gris) interactuando con un segmento de RNA de poliuridinas (poliU) de 9 nucleótidos en color rojo. (B) Estructura 3D de dos motivos PPR de la proteína RNasaP (PRORP1) de *Arabidopsis* se muestra específicamente la región que comprende los residuos 96 al 175 (PDB ID 4G23). Los residuos 6 y 1' del sexto motivo PPR (en amarillo) han sido propuestos como determinantes en la especificidad de unión a un nucleótido específico. En el ejemplo se muestra un uracilo representado por “U”. Modificado desde Barkan y Small (2014). Visualizado con el programa Chimera <http://www.cgl.ucsf.edu/chimera/>.

HIPÓTESIS

“Proteínas PPR candidatas de la subfamilia PLS - E o DYW son mitocondriales y necesarias para la edición de uno o más sitios en transcritos de este organelo en *Arabidopsis thaliana*. Estas proteínas PPR interactúan con sus RNA dianas en forma consistente con el código propuesto para el reconocimiento del RNA por los motivos PPR”

OBJETIVOS

Objetivo General

Identificar y caracterizar factores PPR implicados en la edición de transcritos mitocondriales en *Arabidopsis thaliana*

Objetivos Específicos

- 1.- Seleccionar genes de proteínas PPR potencialmente implicadas en la edición de transcritos mitocondriales en *Arabidopsis thaliana*.
- 2.- Determinar la localización subcelular de las proteínas PPR seleccionadas.
- 3.- Caracterizar la edición de los transcritos organelares en plantas mutantes insercionales en los genes de proteínas PPR candidatas.
- 4.- Caracterizar el fenotipo de las plantas mutantes insercionales en los genes de proteínas PPR candidatas.
- 5.- Determinar si el reconocimiento de los sitios editados por las PPR seleccionadas es consistente con el código propuesto para la interacción RNA - proteína PPR.

CAPÍTULO II

RESULTADOS

**II.1 Identificación de proteínas PPR candidatas que podrían participar en la edición
de transcritos mitocondriales en *Arabidopsis thaliana***

Para la identificación de proteínas PPR que podrían participar en la edición de transcritos mitocondriales en *Arabidopsis thaliana* se realizó una selección de genes candidatos, utilizando los siguientes criterios diseñados para aumentar las posibilidades de identificar factores esenciales en la edición mitocondrial:

- i. Disponibilidad de secuencias de ESTs o cDNAs de genes de proteínas PPR de *Arabidopsis* en bases de datos.
- ii. Destinación mitocondrial predicha por al menos dos algoritmos predictivos.
- iii. Proteínas PPR pertenecientes a la subfamilia PLS, de las subclases E, E+ o DYW.
- iv. Proteínas PPR con identidad de secuencia reducida con otras proteínas PPR de *Arabidopsis*, para disminuir la posibilidad de redundancia funcional y aumentar así la posibilidad de observar una alteración en la edición en mutantes homocigotas.

Es así como de un universo de 441 genes de proteínas PPR descritos en *Arabidopsis* se seleccionaron 5 genes candidatos de acuerdo a los criterios anteriormente descritos (Tabla 1). Se obtuvieron plantas homocigotas mutantes para cada gen, y se analizó el estatus de la edición de más de 300 sitios en mitocondrias y de todos los sitios cloroplásticos a través de la metodología de SNaPshot (Takenaka y Brennicke, 2009). Este trabajo fue realizado en colaboración con el grupo del Dr. Mizuki Takenaka, lo que permitió determinar que en líneas mutantes insercionales de los genes At1g26900 y At1g77170 no habían alteraciones en la edición de los transcritos evaluados. Por el contrario, en las líneas mutantes en los genes At3g25060, At3g03580 y At2g46050 se detectan defectos en la edición de distintos transcritos mitocondriales que van desde una reducción parcial de la edición hasta la anulación de ésta, lo que sugiere la participación de estas proteínas en el procesamiento específico de estos transcritos organelares. En base a este fenotipo molecular, desde ahora estas proteínas PPR son referidas como MEF25, MEF26 y MEF31 respectivamente, por “**Mitochondrial Editing Factor**” y serán descritas en detalle a continuación.

Tabla 2 Las proteínas PPR candidatas analizadas en este estudio

AGI	Subclase	Ranking destinación *	Identidad (%) **	Nº mutantes / Genotipo ***	SNaPshot	Sitios editing afectados
AT1G26900	E	4	33	1 / Homocigota	Si	No
AT1G77170	E	2	41	1 / Homocigota	Si	No
AT3G25060	E	4	33	2 / Homocigota	Si	1 sitio / 1 transcrito
AT3G03580	DYW	2	39	2 / Homocigota	Si	2 sitios / 2 transcritos
AT2G46050	E	4	33	1 / Homocigota	Si	1 sitio / 1 transcrito

Tabla 2. Proteínas PPR candidatas. AGI: Arabidopsis Gene Identifier; (*): Predictores de destinación usados: MitoProtII, TargetP, Predotar e iPSORT, agrupados en la base de datos SUBA (<http://www.plantenergy.uwa.edu.au/applications/suba2/index.php>). Las categorías asignan puntuaciones desde 4 a 0 para cada proteína PPR analizada, siendo asignados cuatro puntos cuando todos los predictores señalan que la probable destinación de la proteína es mitocondrial, y un puntaje de cero cuando ninguno de los programas predice destinación mitocondrial para dicha proteína; (**): Identidad de secuencia de las proteínas PPR candidatas con otras proteínas PPR de Arabidopsis. Se seleccionaron proteínas PPR que presentasen una identidad de secuencia menor al 45% con otras proteínas PPR de Arabidopsis. Con ello se jerarquizó las proteínas candidatas, seleccionando aquellas con la menor identidad posible para reducir las posibilidades de redundancia funcional y aumentar las posibilidades de observar alteraciones en la edición en mutantes insercionales en dichos genes. Para esto se realizó el alineamiento de las secuencias aminoacídicas de las proteínas PPR candidatas utilizando el programa BLAST, disponible en la base de datos “UniProt” (<http://www.ebi.ac.uk/uniprot/search/SearchTools.html>), utilizando los parámetros de búsqueda optimizados por la base de datos para el cálculo de las identidades de las secuencias; (***): Se contó con uno o dos mutantes insercionales independientes para cada uno de los genes At1g26900 (SAIL_205_B05), At1g77170 (SAIL_1291_C04), At3g25060 (SAIL_672_D08 y SALK_091381C), At3g03580 (GT_5_22445 y GK_895H11), At2g46050 (SAIL_451_B04).

II. 2 Caracterización Funcional del Factor de Edición Mitocondrial 25 (MEF25)¹

¹ Arenas-M Anita, Takenaka M, Moreno S, Gómez I, Jordana X (2013) Contiguous RNA editing sites in the mitochondrial *nad1* transcript of *Arabidopsis thaliana* are recognized by different proteins. FEBS Lett 587: 887–891.

En esta sección se describe la identificación y caracterización de un miembro de la familia de proteínas PPR a quien se denominó MEF25, por “Mitochondrial Editing Factor 25”, ya que está implicado específicamente en la edición de un sitio en el transcrito *nad1* que codifica para la subunidad NAD1 del complejo NADH-deshidrogenasa mitocondrial (complejo I).

Los principales resultados concernientes a esta parte del trabajo de tesis son presentados y discutidos en la forma de artículo científico, el cual fue publicado en la revista FEBS letters.

TITLE:

Contiguous RNA editing sites in the mitochondrial *nad1* transcript of *Arabidopsis thaliana* are recognized by different proteins.

Authors:

Anita Arenas-M¹, Mizuki Takenaka², Sebastián Moreno¹, Isabel Gómez¹, and Xavier Jordana^{1*}

¹Departamento de Genética Molecular y Microbiología, Facultad de Ciencias Biológicas, Pontificia Universidad Católica de Chile, Casilla 114-D, Santiago, Chile; ²Molekulare Botanik, Universität Ulm, 89069 Ulm, Germany

*Corresponding author: Tel: 56-2-6862668; Fax: 56-2-2225515; E-mail: xjordana@bio.puc.cl

(Accepted 5 February 2013)

ABSTRACT:

Pentatricopeptide repeat (PPR) proteins have been identified as site specific factors for RNA editing in plant organelles. These proteins recognize *cis*-elements near the editing site. It is unclear how contiguous sites are addressed, whether one or two factors are required. We here characterize the PPR MEF25 to be essential for RNA editing at the *nad1*-308 site in *Arabidopsis* mitochondria. Another editing site just one nucleotide upstream, *nad1*-307, is edited normally in *mef25* mutant lines. This finding shows that two independent factors recognizing similar *cis*-elements are involved at these contiguous sites without competing each other in vivo.

Keywords: RNA editing factor, plant mitochondria, Pentatricopeptide Repeat Protein, PPR protein, *nad1*, MEF25

1. Introduction

RNA editing changes the nucleotide sequence of mature transcripts away from that of its DNA templates. In flowering plants, this change involves the deamination of more than 400 specific cytidines to uridines in organellar transcripts [1]. Since RNA editing was discovered in plant mitochondria, the two basic questions of the specificity determinants in the RNA substrate and the components of the "editosome" machinery have been answered only partially. Experiments with *in organello* [2] or *in vitro* [3] editing systems showed that a *cis*-region in the RNA between -20 and +6 nucleotides relative to the target site is generally necessary and sufficient for editing. However, until now the precise bases for specificity have not yet been established for any site. Recently, the MORF-protein family (Multiple Organellar Editing Factor) was found to be important for editing of many sites in mitochondria and all sites in plastids [4], but their precise role is still open. These MORFs can interact with selected trans-acting proteins necessary for editing single or few specific cytidines, the Pentatricopeptide Repeat (PPR) proteins [5-7]. Distinct functions have been assigned to several PPRs, all of them related to organellar RNA metabolism including splicing, endo- and exonucleolytic processing, translation initiation and editing [8]. PPR proteins are characterized by degenerate motifs of 35 amino acids arranged as tandem repeats. Some PPR proteins (PLS subfamily) contain repeats shorter (S) or longer (L) than the canonical (P) 35 amino acid repeat, and one, two or three additional domains in the C-terminal region: The E-domain, the E+-domain and the DYW-domain [5]. All of the organelle editing factors required for editing at specific sites contain the E, the E and E+ or the E, E+ and DYW domains. PPR proteins most likely bind specifically to RNA with the PPR motifs, and in the case of RNA editing recognize the *cis*-element at the target site [e.g. 9-11].

Several RNA editing sites in plant organelles are located near or even contiguous to other sites. Two alternative scenarios are plausible for the recognition of contiguous nucleotides. In the first scenario one PPR protein binds to its cognate *cis*-element and facilitates RNA editing at both sites. In this case, the PPR protein must allow at least one nucleotide flexibility in the distance between

the binding site and editing sites. In the second scenario, two PPR proteins independently mediate editing at the two sites. In this case, the *cis*-elements of the contiguous sites are expected to overlap for the two trans-factors and the mutually exclusive binding of the two PPR proteins must accommodate the potential competition. We here report the identification and analysis of the novel Mitochondrial RNA Editing Factor 25, MEF25, which very specifically addresses only one of two contiguous RNA editing sites in the *nad1* transcript.

2. Materials and Methods

2.1. Molecular and phenotypic analysis of *mef25* mutant plants.

The wild type *Arabidopsis thaliana* ecotype used was Columbia-0 (Col-0). The T-DNA insertion mutant lines SAIL 672 (*mef25-1*) and SALK_091381C (*mef25-2*) in the Col-0 background were obtained from the ABRC Stock Center. Seeds were sown on half-concentrated Murashige and Skoog medium supplemented with 1% (w/v) sucrose and solidified with 0.8% (w/v) agar. After 2 weeks in a 16-h/8-h day/night cycle at 22°C, seedlings were transferred to soil and grown for 4-6 weeks under the same conditions. To isolate homozygous mutant plants, genotyping was performed by PCR as described [12], using the specific primers indicated in Supplementary Table 1 (see also Supplementary Fig. 2). Growth and phenotype of wild type and *mef25* mutant plants were analyzed at the seedling stage and in adult plants. Alexander staining of pollen was performed on anthers of recently opened flowers as described [12].

To characterize *nad1* transcripts in wild type and *mef25* mutant plants, total RNA was isolated from leaves of six-week-old plants or 15-days-old seedlings with the TRIzol reagent following the manufacturer's protocol (Invitrogen). Northern-blot analyses were performed as described, using a *nad1* ³²P-labeled probe [13].

2.2. Analysis of RNA editing sites

mef25-1 and *mef25-2* homozygous mutant plants were screened by multiplexed single base extension “SNapShot analysis” for altered RNA editing at specific sites [14]. To confirm the editing defect observed, specific cDNA fragments were generated by RT-PCR amplification [15] using primer *nad1ssampAC580R* for cDNA synthesis and primers *nad1ssamp-37F* and *nad1ssampCA376R* for PCR amplification (Supplementary Table 1). The status of the respective editing site was determined by sequence analysis (Macrogen, Seoul, Korea or 4base lab, Reutlingen, Germany). The cDNA sequences were compared for C to T differences resulting from RNA editing.

2.3. Constructs for MEF25 localization experiments and promoter analysis

MEF25 localization was investigated by GFP fusion to the MEF25 amino terminal 144 amino acids in vector pK7FWG2 [16]. The DNA encoding this amino terminal sequence was obtained by PCR with primers *amino-mef25F* and *amino-mef25R* (Supplementary Table 1). To construct GUS promoter fusions, two PCR fragments of 1,000 and 500 bp from upstream of the MEF25 initiation codon were cloned into the pGEM-T easy plasmid (Promega), digested with *SalI* and *NcoI* (restriction sites introduced by the primers, see Supplementary Fig. 1) and ligated to the GUS frame in pCAMBIA1381 (www.cambia.org). Primers were *Pro1mef25-SalIF* and *Promef25-NcoIR* for the 1.0 kb DNA fragment and *Pro2mef25-SalIF* and *Promef25-NcoIR* for the 0.5 kb fragment. Constructs were verified by DNA sequencing, introduced into *A. tumefaciens* GV3101 by electroporation and *Arabidopsis* plants were transformed by floral dip [17]. Several independent transgenic lines were obtained for each construct. GFP and Mitotracker Orange (Invitrogen) fluorescence were analyzed with a confocal microscope (Nikon C2+) in seedling roots and protoplasts prepared as described [18]. Histochemical GUS staining was performed as described [19].

3. Results.

3.1. *At3g25060* encodes an E+-PPR protein targeted to mitochondria

To identify specific trans-factors involved in RNA editing of plant mitochondrial transcripts, we selected PPR candidate genes based on four criteria: i) PPR genes with ESTs or cDNA sequences in public databases; ii) Mitochondrial destination predicted by different subcellular sorting algorithms (Predotar, TargetP, iPSORT, Mitoprot2); iii) Members of the PLS subfamily with E, E+ or DYW motifs; iv) Low sequence similarity with other *Arabidopsis* PPR proteins. One of the PPR genes thus selected is At3g25060, which encodes an E+-PPR protein.

To test the prediction that the protein translated from this gene, At3g25060, is targeted to mitochondria, a DNA fragment encoding the N-terminal 144 amino acids (up to the second PPR motif) was fused upstream of the GFP gene and stable transgenic *Arabidopsis* lines were generated. In leaf protoplasts and seedling roots from these transgenic lines GFP fluorescence showed the punctuate pattern characteristic of mitochondria (Fig.1). Furthermore, GFP and Mitotracker-Orange fluorescence overlapped and were clearly distinct from chloroplast fluorescence, confirming that the AT3G25060 protein is targeted to mitochondria and not to chloroplasts.

To characterize *MEF25* gene expression, we fused either 1,000 bp or 500 bp fragments from just upstream of the *MEF25* AUG codon to a GUS reporter gene. Transgenic lines for both constructs showed similar expression patterns, suggesting that the main promoter activity is located within 500 bp upstream of the *MEF25* AUG start codon (Supplementary Fig. 1). GUS activity was detected in cotyledons and hypocotyls of seedlings, in leaves, and in sepals, filaments and the pistil of flowers.

3.2 *MEF25* is required for one mitochondrial RNA editing site

With the mitochondrial localization of AT3G25060 confirmed, we analyzed the RNA editing phenotypes in At3g25060 disrupted mutants. Two independent mutants were retrieved from the

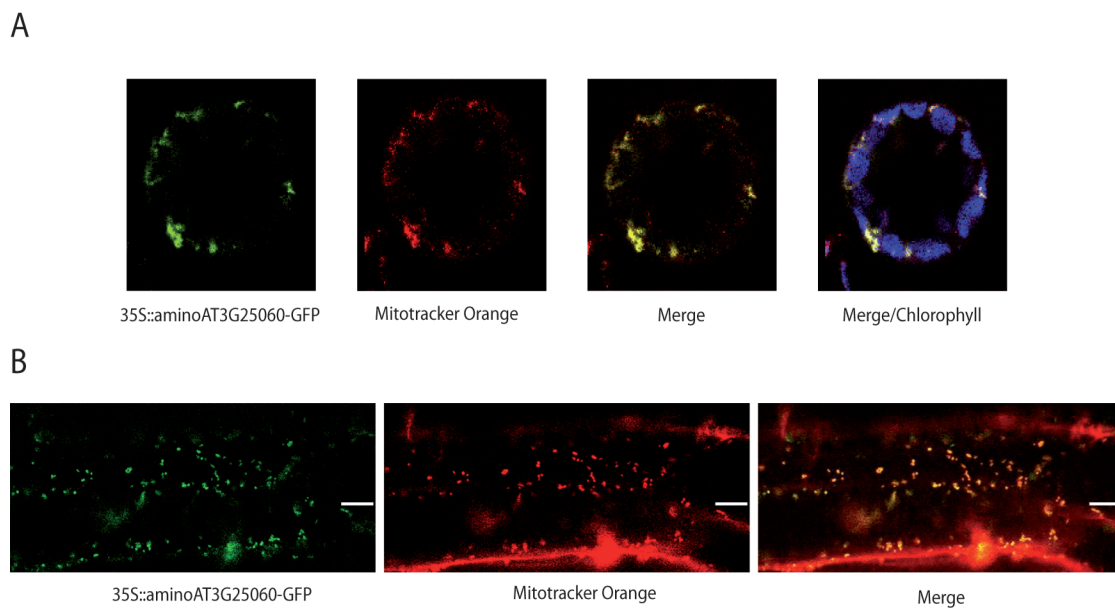
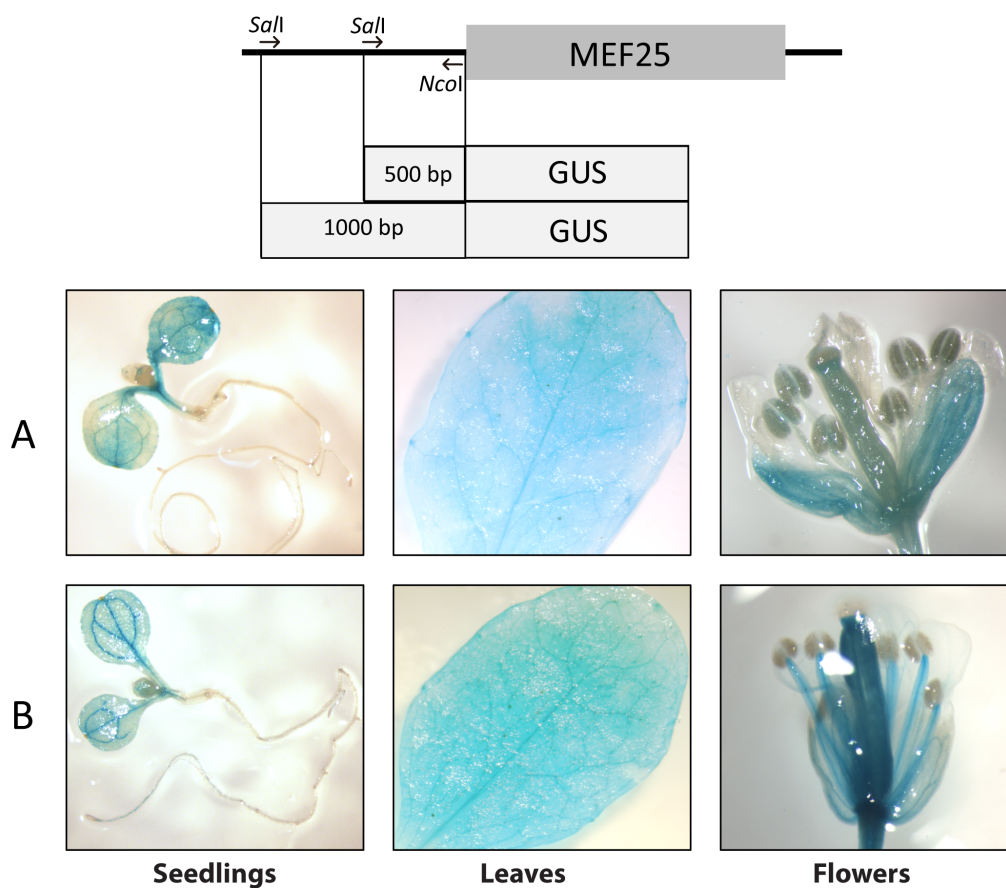


Figure 1. The AT3G25060 (MEF25) amino terminus targets GFP to mitochondria. The aminoAT3G25060-GFP fusion protein was expressed in transgenic *Arabidopsis* plants under control of the CaMV35S promoter. (A) Protoplasts isolated from leaves and incubated with Mitotracker Orange, from left to right: AminoAT3G25060-GFP fluorescence; Mitotracker Orange signal; merged aminoAT3G25060-GFP and Mitotracker Orange signals (yellow indicates signal co-localization); overlay between the merged image of aminoAT3G25060-GFP and Mitotracker Orange signals, and chlorophyll autofluorescence. (B) Roots of a seedling, from left to right: aminoAT3G25060-GFP signal; Mitotracker Orange stain; merged GFP and Mitotracker Orange signals. Scale bar = 10 μ m.



Supplementary Fig. 1. *MEF25* promoter activity. Upper panel: Schematic of the *MEF25* promoter fusions to the GUS reporter gene. One thousand or 500 bp upstream of the *MEF25* ATG start codon were fused in frame to the GUS coding sequence. Histochemical localization of GUS activity is shown for representative transgenic lines carrying (A) the 1,000 bp construct or (B) the 500 bp construct. From left to right: GUS activity staining of 7-days-old seedlings, rosette leaves and inflorescences from adult 8-week-old plants.

SALK and SAIL collections, homozygous plants were selected and the positions of both independent T-DNA insertions were confirmed by PCR and sequencing to be in the coding region (Fig. 2A, Supplementary Fig. 2).

For the identification of mitochondrial RNA-editing defects in the mutants, RNA from leaves was screened for deficiencies in RNA editing with the multiplexed SNaPshot approach [14,15]. With this assay, 367 annotated editing sites were analyzed, in addition, 45 further mitochondrial editing sites were examined by direct RT-PCR amplification and sequencing. In this analysis we found a defect in editing at one of the 25 editing events in the *nad1* transcript, site *nad1*-308 is not edited in both mutants (Fig. 2B and C). Surprisingly, the immediately adjacent site *nad1*-307 is completely edited in both wild type and mutant plants, suggesting that nucleotides 307 and 308 are recognized by different factors. When we analyzed the *nad1* RNA editing status by RT-PCR, we found that among 43 mutant cDNA clones, all were unedited at *nad1*-308 and edited at *nad1*-307 (Supplementary Fig. 3A). In contrast, among 82 wild type cDNA clones, all were edited at *nad1*-307 and 80 were also edited at *nad1*-308. When normally edited at positions 307 and 308, the CCG (pro) codon is changed to UUG (leu) codon. The leucine amino acid is highly conserved in plants. In both mutants the CCG codon is edited to UCG, encoding serine. The gene locus (TAIR ID: AT3G25060) was renamed *MEF25* and the mutant lines *mef25-1* and *mef25-2*, respectively.

3.3. Phenotype and pattern of *nad1* transcripts are not affected in *mef25* mutants

When the mutants were compared to wild type plants, no changes were seen in the macroscopic phenotype or in the development of homozygous *mef25-1* and *mef25-2* plants under standard growth conditions (Supplementary Fig. 4). Neither was pollen viability affected. These results suggest that the lack of editing at site *nad1*-308 has little effect on normal mitochondrial function. RNA processing of the *nad1* precursor to the mature transcript includes *cis*- and *trans*-splicing and 25 RNA editing events (Fig. 2B). To evaluate if the lack of RNA editing at the *nad1*-308 site

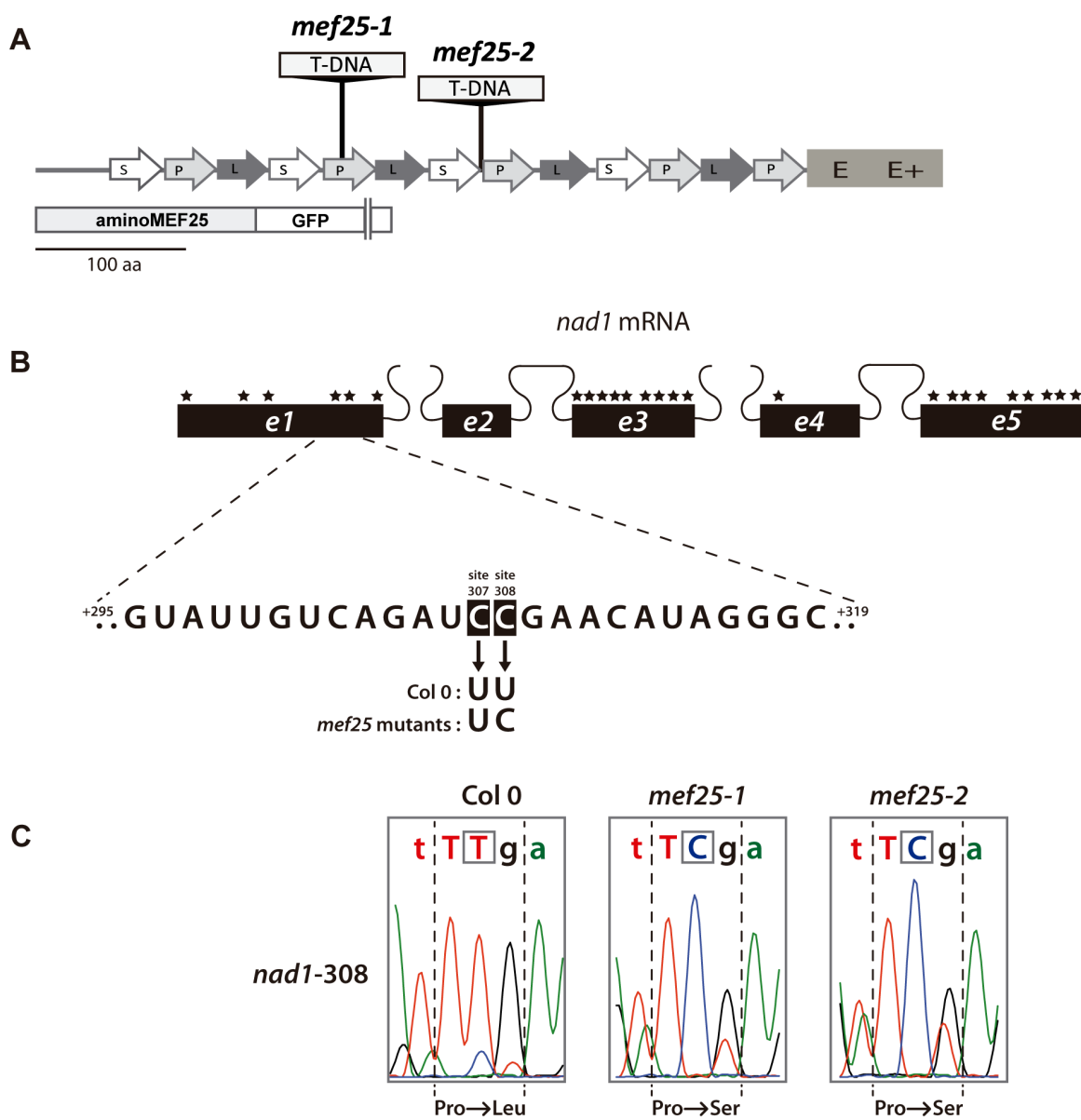


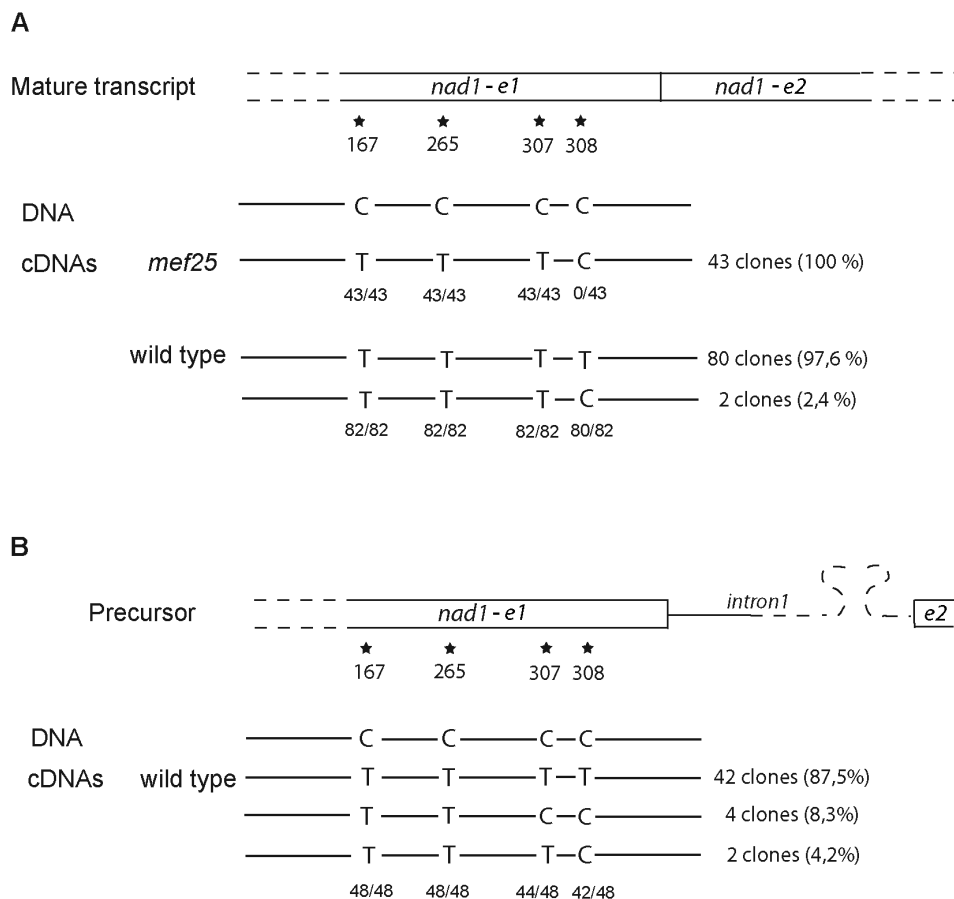
Figure 2.

Figure. 2. AT3G25060 (MEF25) is necessary for one RNA editing event in the mitochondrial *nad1* transcript. (A) AT3G25060 (MEF25) structure and T-DNA insertion sites in the *mef25-1* and *mef25-2* mutants. MEF25 contains thirteen pentatricopeptide (PPR) motifs (arrows). P represents the classic PPR motif, L and S the variant “long” and “short” PPR motifs, and E and E+ the extension motifs at the C-terminus (box). The N-terminal sequence fused to GFP to analyze the subcellular localization is shown below the MEF25 structure. (B) Structure of the *nad1* mRNA. Five exons (black boxes), two *trans*- and two *cis*-splicing events (open and closed lines, respectively), and twenty-five editing sites (stars) are indicated. The zoomed sequence below gives the nucleotides from +295 to +319 nt in exon 1 (e1) with the *nad1*-308 site which is not edited in the *mef25* mutants. Nucleotides are counted from the AUG. (C) Sequence analyses of *nad1* cDNAs obtained by RT-PCR from wild type and *mef25* mutant plants are displayed. The codon containing both 307 and 308 editing sites is indicated by dashed lines, editing sites are in capitals and site *nad1*-308 is framed.

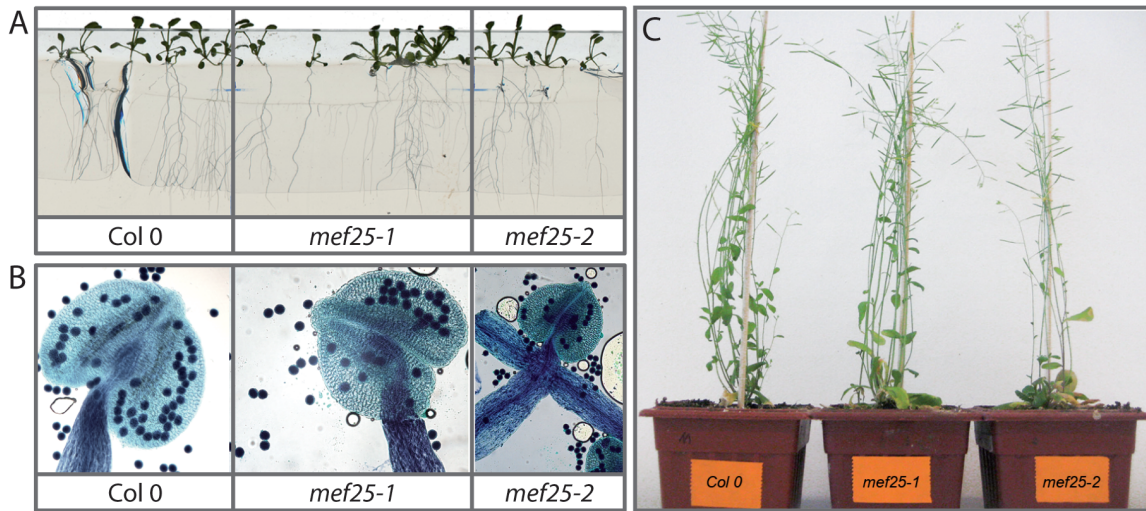
T-DNA TAAGTTGTCTAAGCGTCAATTTGTTTACACCACAATATATCCTGCCACCAGCCAGCCAACAGCTCCCCGACCGGCAGCTC
mef25-2 TAAGTTGTCTAAGCGTCAATTTGTTTACACCACAATATATCAGCAATCGCAAGTCGGGTCATTAAAGACTGGTAGATTAG
Col-0 GTTTC AACCGATTTAGTGAGCGCTTGTGGAGTACTTGTGGCGTGTTCGCAAGTCGGGTCATTAAAGACTGGTAGATTAG

+ 985

Supplementary Fig. 2. T-DNA insertion sites in the *mef25* mutants. PCR amplification of *MEF25* gene/T-DNA junctions was performed with the following primers: for *mef25-1* (A), the left border primer (*LB1-SAIL*) in combination with either an upstream (*mef25-1-2F*) or downstream (*mef25-1R*) gene specific primer; for *mef25-2* (B), the left border primer *LBb1-SALK* in combination with either an upstream (*mef25-1-2F*) or a downstream (*mef25-2R*) gene specific primer. Sequencing of the PCR products demonstrate that the T-DNA left border was present at both ends of the insertion in both mutants and, more importantly, established that no major deletions or chromosomal rearrangements took place during the insertional event. (A) A 135 bp deletion (between +641 and +775) and a 9 bp insertion (in red, unknown origin) occurred in *mef25-1*. (B) In *mef25-2*, a 49 bp deletion (between +936 and +984) and two insertions (8 and 5 bp, in red) were found. In all alignments the first row corresponds to the T-DNA sequence, the second row to the junction between *MEF25* and T-DNA and the third row to the wild type *MEF25* sequence. T-DNA sequences are in blue and *MEF25* sequences in green. Numbers are in relation to the first nucleotide of the ATG start codon from *MEF25*.



Supplementary Fig. 3. (A) Analysis of editing in mature *nad1* transcripts. (B) Analysis of editing in unspliced *nad1* exon1 transcripts. RNA was isolated from 15 days-old-wild type and *mef25* mutant seedlings. (A) cDNAs corresponding to mature transcripts were obtained by RT-PCR amplification using primer *nad1ssampAC580R* for cDNA synthesis and primers *nad1ssamp-37F* and *nad1ssampCA376R* for PCR amplification. (B) cDNAs encompassing exon 1 and 88 nucleotides of intron 1 were generated by RT-PCR amplification using primer *intron1_nad1-R2B* for cDNA synthesis and primers *nad1ssamp-37F* and *intron1_nad1-R1* for PCR amplification. The status of editing sites 167, 265, 307 and 308 (stars) was assessed by cDNA sequence analysis. Lines represent the sequenced cDNAs, edited or unedited nucleotides are indicated by T and C, respectively. The ratio of edited/total clones at each site is indicated.



Supplementary Fig. 4. Phenotypically both *mef25* mutant lines grow and develop indistinguishably from wild type plants. (A) Wild type, *mef25-1* and *mef25-2* seedlings were grown for 13 days on vertical plates with half concentrated Murashige and Skoog medium. (B) Pollen viability is not affected in the mutant plants as probed by Alexander's staining. Pollen grains show the intensive purple staining of the cytoplasm, characteristic of viable pollen. (C) Mature 12 week-old *mef25-1* and *mef25-2* plants show heights and seed sets comparable to those of wild type plants.

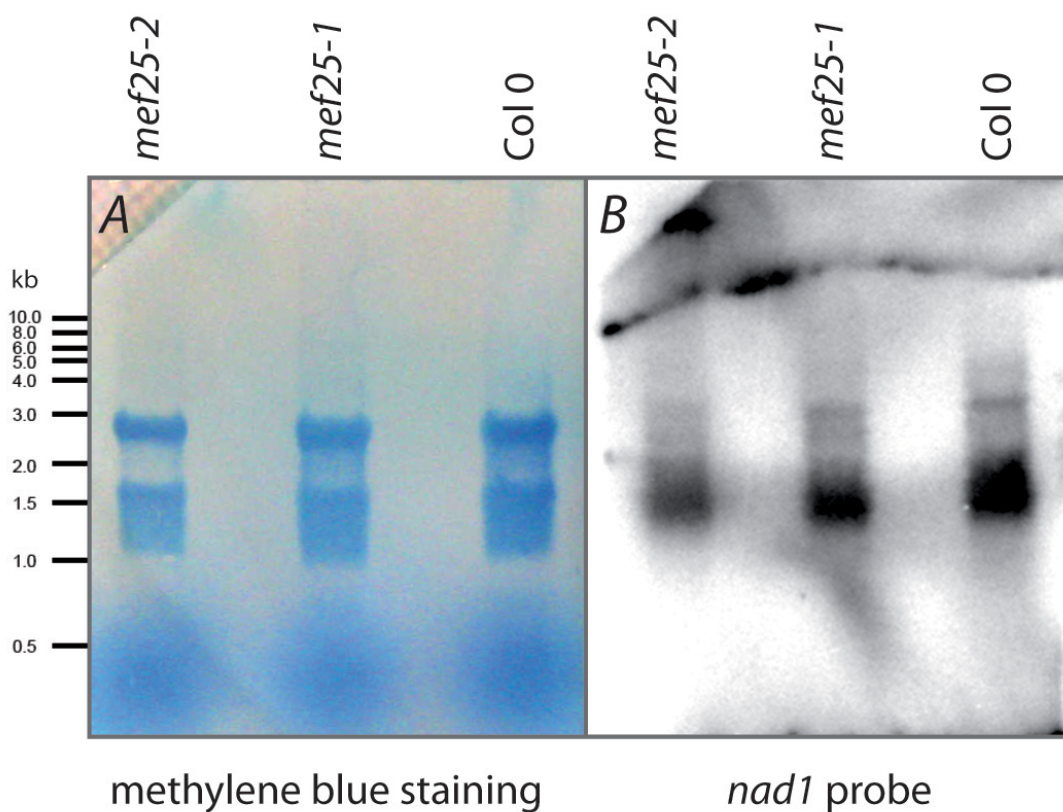
alters other RNA processing steps in the *nad1* transcript, we assayed an RNA gel blot with a probe specific for the *nad1* transcript. No difference in the *nad1* transcript profiles was seen in leaves of wild type and *mef25-1* and *mef25-2* mutant plants, suggesting that stability, splicing and 5' or 3' processing are not affected (Supplementary Fig. 5). This result also precludes that the molecular phenotype of the loss of RNA editing at site *nad1*-308 is a secondary effect generated by defects in transcription or processing.

4. Discussion

Here we identify MEF25 as a mitochondrial PPR protein required for RNA editing specifically at site *nad1*-308. This assignment is supported by the lack of editing at this site in two independent *mef25* mutant lines. The unedited C changes a highly conserved leucine to a serine in the NAD1 protein. NAD1 is one of more than 40 subunits of respiratory Complex I (NADH dehydrogenase) in plants and one of nine coded in the mitochondrial genome [20,21]. Modular subunit arrangements proposed by Klodmann et al. [20] localize this subunit in the membrane arm, to the periphery at the interface of the two complex I arms.

The *mef25* mutant lines are phenotypically indistinguishable from wild type plants, at least under standard growth conditions (Supplementary Fig. 4), suggesting that the amino acid alteration caused by the absence of the *nad1*-308 editing event is tolerated in the NAD1 protein despite the conservation of the amino acid encoded by the edited codon in many plants (Fig. 3). Amino acid changes caused by other editing events in the *nad1* transcript may be more detrimental and it will be interesting to see if there is any cumulative effect detectable from the combination of several such amino acid alterations. Until now the only factors implicated in editing of other sites in the *nad1* mRNA are MORF factors acting at sites 500, 743 and 755 (Supplementary Fig. 6) [4].

There are several contiguous pairs of RNA editing events in *Arabidopsis* mitochondria, for example sites *mttB*-144 and 145 [22]. In the *slo2* mutants, RNA editing at both *mttB*-144 and 145 sites is completely lost, suggesting that binding of the SLO2 protein to its *cis*-element is essential



Supplementary Fig. 5. Northern blot analysis of *nad1* transcripts in wild type and *mef25* mutant plants. Total leaf RNA (7 μ g each lane) from wild type, *mef25-1* and *mef25-2* plants was separated by electrophoresis on a formaldehyde agarose gel and transferred onto a nylon membrane. (A) Methylene blue staining of RNA as a loading and transfer control. (B) Blot hybridization with a *nad1* probe encompassing exons 1 to 5, obtained by RT-PCR using primers *nad1ssamp-37F* and *nad1-rev*. The *nad1* transcript profiles were no different in wild type and *mef25* plants, indicating that neither the quantity, stability nor processing are affected. Positions of DNA size standards are indicated on the left in kilobase pairs.



Figure 3. Conservation of nucleotide and amino acid sequences around the *nad1*-308 editing site.

(A) Alignment of nucleotide sequences around the editing site targeted by MEF25 in land plants.

Nucleotide 308 (counted from the AUG) and other RNA editing sites are shown as unedited C or

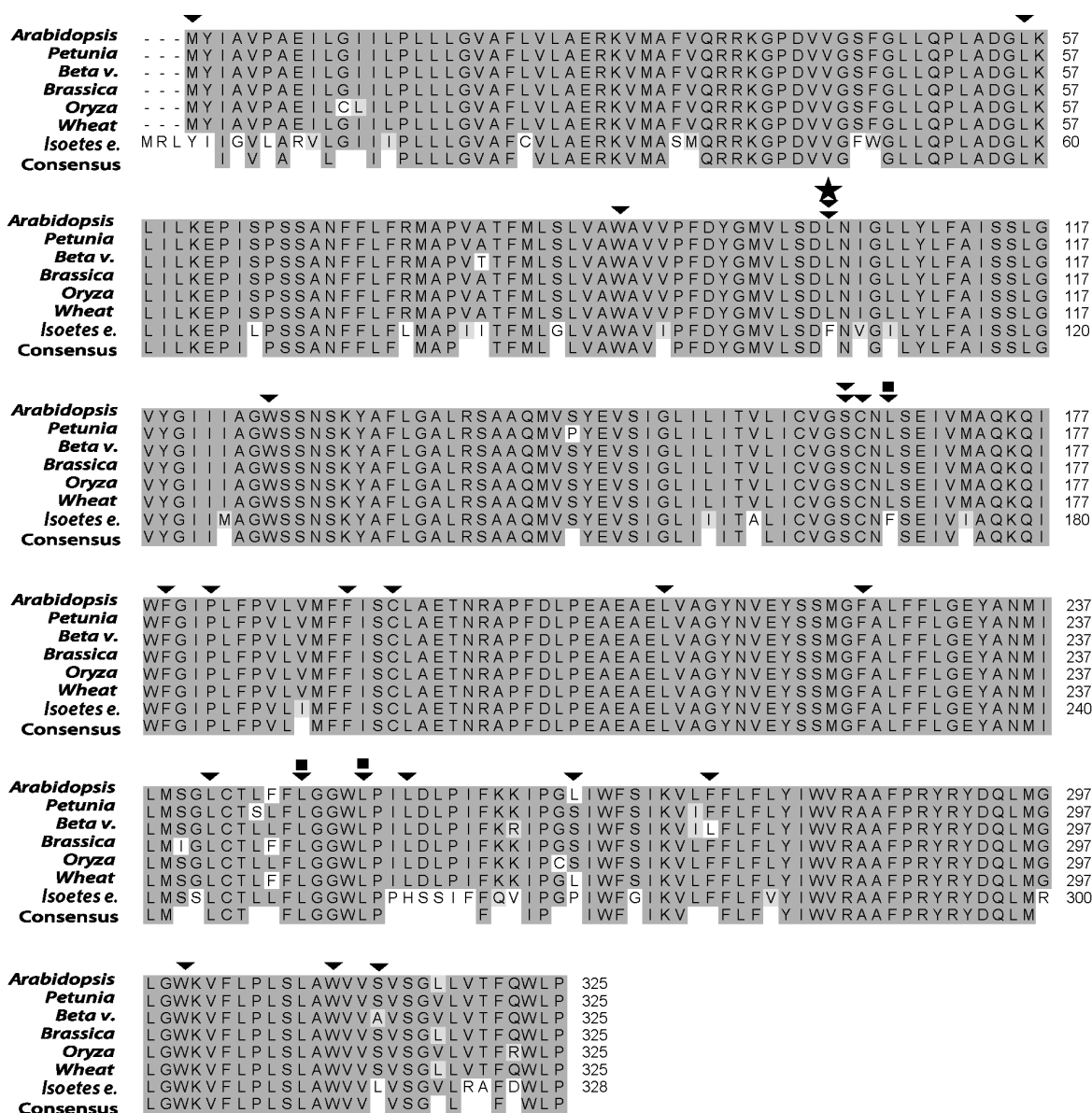
U in red. (B) Alignment of amino acid sequences around leucine 103 of the NAD1 protein.

Amino acid sequences are from the UniProt database (<http://www.uniprot.org>), mRNA sequences

from REDIdb (http://biologia.unical.it/py_script/REDIdb/) and lycophyte sequences from Grewe

et al. [23]. Identical residues are in grey, similar amino acids are in light grey. Amino acids

altered by RNA editing events in angiosperms are box.



for RNA editing at both sites. The action of MEF25 at contiguous sites *nad1*-307 and 308 is likely different from SLO2, since editing is lost only at site *nad1*-308 while the adjacent *nad1*-307 nucleotide is normally edited. More than forty independent clones of the *nad1* cDNA from mutant plants confirm this specificity (Supplementary Fig. 3A). Therefore, these two adjacent editing sites must be recognized by (at least) two different editing factors. These two factors appear to solely act on their cognate sites, the factor specific for site *nad1*-307 cannot edit *nad1*-308 and site *nad1*-307 does not require MEF25. Furthermore, site *nad1*-307 does not require site *nad1*-308 to be edited and is recognized and processed independently from the identity of the downstream nucleotide at *nad1*-308. The contrasting behavior of SLO2 and MEF25 at respectively contiguous RNA editing events may be induced by different cofactors, however, it is still possible that SLO2 acts similar to MEF25. SLO2 may be responsible for editing at only one of the two *mttB* sites which may be a pre-requisite for binding of another trans-factor and editing at the second site.

Among the most closest situations in which mitochondrial editing at only one of two nearby sites is impaired in PPR mutants are sites *ccmF*-1 and *ccmF*-2 in *Physcomitrella patens*, separated by 18 nucleotides, and *cox3*-413 and *cox3*-422 in *Arabidopsis*, separated by 8 nucleotides [24, 25]. In the *Physcomitrella* PpPPR_71 mutant editing at *ccmF*-2 was impaired while *ccmF*-1 was unaffected. In this case editing at *ccmF*-1 may be required for editing at *ccmF*-2 since cDNA sequence analysis suggested that *ccmF*-1 editing occurs before *ccmF*-2 editing, and the PpPPR_71 protein preferentially binds RNA edited at *ccmF*-1. In the *Arabidopsis* *mef11* mutant no detectable editing was observed at site *cox3*-422 while *cox3*-413 editing was unaffected, and editing at *cox3*-413 increases the similarity between the presumed *cox3*-422 recognition sequence and that of another editing site targeted by MEF11 (*nad4*-124). However, in contrast to the situation of *ccmF* sites, cDNA sequence analysis indicated that editing at the upstream *cox3*-413 is not required for editing at the *cox3*-422 site. It is clear from our results that editing at *nad1*-307 does not require previous editing at *nad1*-308. However, whether editing at *nad1*-308 requires

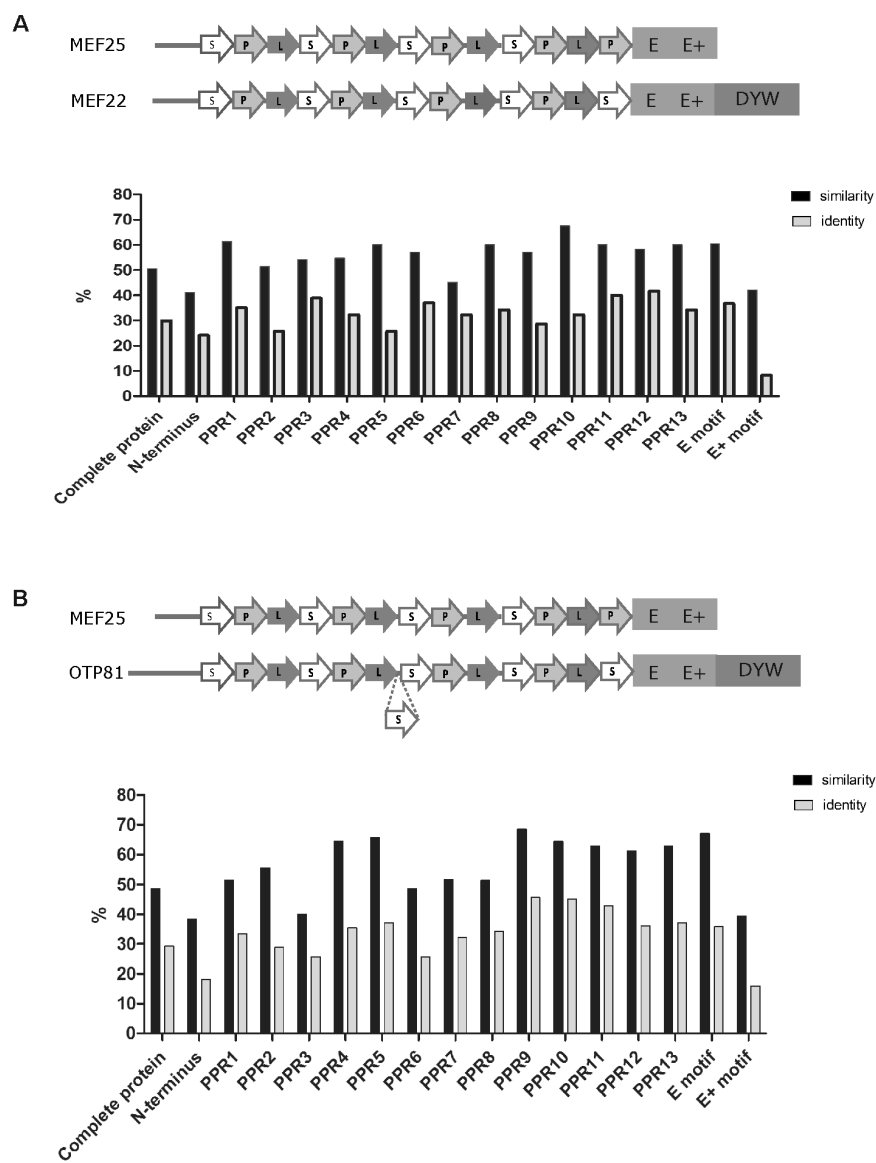
nad1-307 to be edited or not is unresolved. cDNA sequence analysis showed that both sites are nearly fully edited in mature transcripts from wild type seedlings (82/82 clones edited at *nad1-307* and 80/82 clones edited at *nad1-308*, Supplementary Fig. 3A). In an attempt to analyze partially edited transcripts, we sequenced 48 cDNAs containing exon 1 and part of intron 1 (Supplementary Fig. 3B). Forty two cDNA clones (87.5%) were edited at both sites, four (8.3%) were unedited at both sites and two (4.2%) were edited at *nad1-307* only. Although the number of cDNA clones edited at *nad1-307* only is low (two for mature transcripts and two for unspliced transcripts), the absence of cDNAs edited at *nad1-308* only may suggest that *nad1-307* editing occurs before and is required for editing at *nad1-308*. Alternatively, both sites may be edited independently with *nad1-307* editing more efficient and faster. Unfortunately we did not obtain recombinant MEF25 protein to evaluate MEF25 *in vitro* binding to RNA edited or unedited at *nad1-307*.

Recently an amino acid code has been proposed for RNA recognition by PPR proteins [26]. This amino acid code does not assign any role to the -1 position of *nad1-307* in *nad1-308* editing. Based on the work of Lurin et al. [5] we carefully analyzed MEF25 amino acid sequence and manually predicted the structure of PPR, E and E+ motifs shown in Fig. 2. This structure is similar to that annotated in the UniProt database (<http://www.uniprot.org>) and allows us to identify in each motif the 6 and 1' amino acids proposed to constitute the amino acid code [26]. However, application of this code to MEF25 does not show preferential targeting of *nad1-308* over *nad1-307*, and does not allow to predict which PPR protein may target *nad1-307*. It will be interesting to identify the PPR protein factor for site *nad1-307* which may have a similar set of PPR repeats recognizing similar nucleotides in the *cis*-element. Alternatively, the PPR repeats may be different from MEF25 and attach to a very different set of nucleotides in a shifted, possibly overlapping position. A Blast search using MEF25 as a query identified 80 PPR proteins with an overall identity of 30-35%, with AT3G12770 (MEF22) and AT2G29760 (OTP81) displaying the lowest e-values. A comparison between MEF25 and MEF22 or OTP81 is shown in

Supplementary Fig. 7. MEF 22 is required for editing at *nad3*-149 and not at *nad1*-307 [27], and OTP81 is a plastid protein involved in *rps12* intron editing [28]. No significant similarities are observed between these and the *nad1*-307/308 target sites. Thus, the PPR protein targeting *nad1*-307 remains to be identified. In conclusion, our results thus suggest that two specificity factors can share a *cis*-region in the RNA without competition in vivo.

5. Acknowledgments

We are very grateful to Axel Brennicke for his critical reading of the manuscript, discussion of results and constant encouragement, and to Franko Restovic for his valuable help and suggestions. This work was supported by grants from CONICYT (AT24100161), FONDECYT (1100601), Millenium Nucleus in Plant Functional Genomics (P10-062-F, Millennium Scientific Initiative Program, Mideplan, Chile) and the Deutsche Forschungsgemeinschaft. Mizuki Takenaka is a Heisenberg fellow.



Supplementary Fig. 7. (A) Similarity between MEF25 and MEF22. (B) Similarity between MEF25 and OTP81. Schematic structures of MEF25, MEF22 and OTP81 were shown. P represents the canonical PPR motif, L and S the variant “long” and “short” PPR motifs, and E, E+ and DYW the extension motifs at the C terminus. MEF25 and MEF22 proteins contain 13 PPR motifs, OTP81 14 PPR motifs. Below the structures, percentages of amino acid identities (grey bars) and similarities (black bars) between MEF25 and MEF22 or OTP81 are given for each of the repeats, the N-terminal region up to the first repeat, and the E and E+ elements.

6. References

- [1] Giegé, P. and Brennicke, A. (1999) RNA editing in Arabidopsis mitochondria affects 441 C to U changes in ORFs. *Proc. Natl. Acad. Sci. USA* 96, 15324-9.
- [2] Farré, J.C., León, G., Jordana, X. and Araya, A. (2001) Cis recognition elements in plant mitochondrion RNA editing. *Mol. Cell. Biol.* 21, 6731-37.
- [3] Neuwirt, J., Takenaka, M., van der Merwe, J.A. and Brennicke, A. (2005) An in vitro RNA editing system from cauliflower mitochondria: editing site recognition parameters can vary in different plant species. *RNA* 11, 1563-70.
- [4] Takenaka, M., Zehrmann, A., Verbitskiy, D., Kugelman, M., Härtel, 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. USA* 109, 5104-9.
- [5] Lurin, C. et al. (2004) Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell* 16, 2089-103.
- [6] Zehrmann, A. et al. (2009) A DYW domain-containing pentatricopeptide repeat protein is required for RNA editing at multiple sites in mitochondria of *Arabidopsis thaliana*. *Plant Cell* 21, 558-67.
- [7] Fujii, S. and Small, I. (2011) The evolution of RNA editing and pentatricopeptide repeat genes. *New Phytol.* 21, 37-47.
- [8] Schmitz-Linneweber, C. and Small, I. (2008) Pentatricopeptide repeat proteins: a socket set for organelle gene expression. *Trends Plant Sci.* 13, 663-70.
- [9] Pfalz, J., Bayraktar, O.A., Prikryl, J. and Barkan, A. (2009) Site-specific binding of a PPR protein defines and stabilizes 5' and 3' mRNA termini in chloroplasts. *EMBO J.* 28, 2042-2052.

- [10] Okuda, K., Nakamura, T., Sugita, M., Shimizu, T. and Shikanai, T. (2006) A pentatricopeptide repeat protein is a site recognition factor in chloroplast RNA editing. *J. Biol. Chem.* 281, 37661-7.
- [11] Hammani, K., Colas des Francs-Small, C., Takenaka, M., Tanz, S.K., Okuda, K., Shikanai, T., Brennicke, A. and Small, I. (2011) The pentatricopeptide repeat protein OTP87 is essential for RNA editing of *nad7* and *atp1* transcripts in *Arabidopsis* mitochondria. *J. Biol. Chem.* 286, 21361-71.
- [12] León, G., Holuigue, L. and Jordana, X. (2007) Mitochondrial complex II Is essential for gametophyte development in *Arabidopsis*. *Plant Physiol.* 143, 1534-46.
- [13] Zehrmann, A., Verbitskiy, D., Härtel, B., Brennicke, A. and Takenaka, M. (2011) RNA editing competence of trans-factor MEF1 is modulated by ecotype-specific differences but requires the DYW domain. *FEBS Lett.* 584, 4181-6.
- [14] Takenaka, M. and Brennicke, A. (2009) Multiplex single-base extension typing to identify nuclear genes required for RNA editing in plant organelles. *Nucleic Acids Res.* 37, e13.
- [15] Takenaka, M. and Brennicke, A. (2007) RNA editing in plant mitochondria: assays and biochemical approaches. *Methods Enzymol.* 424, 439-58.
- [16] Karimi, M., Depicker, A. and Hilson, P. (2007) Recombinational cloning with plant gateway vectors. *Plant Physiol.* 145, 1144-54.
- [17] Clough, S.J. and Bent, A.F. (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* 16, 735-43.
- [18] Wu, F.H., Shen, S.C., Lee, L.Y., Lee, S.H., Chan, M.T. and Lin, C.S. (2009) Tape-*Arabidopsis* Sandwich- a simpler *Arabidopsis* protoplast isolation method. *Plant Methods* 5, 16.
- [19] Elorza, A., Roschztardt, H., Gómez, I., Mouras, A., Holuigue, L., Araya, A. and Jordana, X. (2006) A nuclear gene for the iron-sulfur subunit of mitochondrial complex II

- is specifically expressed during *Arabidopsis* seed development and germination. *Plant Cell Physiol.* 47, 14-21.
- [20] Klodmann, J. and Braun, H.P. (2011) Proteomic approach to characterize mitochondrial complex I from plants. *Phytochemistry* 72, 1071-80.
- [21] 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.
- [22] Zhu, Q. et al. (2012) SLO2, a mitochondrial pentatricopeptide repeat protein affecting several RNA editing sites, is required for energy metabolism. *Plant J.* 71, 836-849.
- [23] Grewe, F., Herres, S., Viehover, P., Polsakiewicz, M., Weisshaar, B. and Knoop, V. (2012) A unique transcriptome: 1782 positions of RNA editing alter 1406 codon identities in mitochondrial mRNAs of the lycophyte *Isoetes engelmannii*. *Nucleic Acids Res.* 39, 2890-902
- [24] Tasaki, E., Hattori, M. and Sugita, M. (2010) The moss pentatricopeptide repeat protein with a DYW domain is responsible for RNA editing of mitochondrial *ccmFc* transcript. *Plant J.* 62, 560-70
- [25] Verbitskiy, D., Zehrmann, A., van der Merwe, J.A., Brennicke, A. and Takenaka, M. (2010) The PPR protein encoded by the *LOVASTATIN INSENSITIVE 1* gene is involved in RNA editing at three sites in mitochondria of *Arabidopsis thaliana*. *Plant J.* 61, 446-55
- [26] Barkan, A., Rojas, M., Fujii, S., Yap, A., Chong, Y.S., Bond, C.S. and Small, I. (2012) A combinatorial amino acid for RNA recognition by pentatricopeptide repeat proteins. *PLoS Genet.* 8: e1002910.
- [27] Takenaka, M., Verbitskiy, D., Zehrmann, A. and Brennicke, A. (2010) Reverse genetic screening identifies five E-class PPR proteins involved in RNA editing in mitochondria of *Arabidopsis thaliana*. *J. Biol. Chem.* 285, 27122-9

- [28] Hammani, K., Okuda, K., Tanz, S.K., Chateigner-Boutin, A.-L., Shikanai, T. and Small, I. (2009) A study of new *Arabidopsis* chloroplast RNA editing mutants reveals general features of editing factors and their target sites. *Plant Cell* 21, 3686-99

Supplementary Table 1. List of primers used in this work

Primer name	used for	Orientation	Primer (5'-3')
<i>LBb1-SALK</i>	<i>mef25-2</i> genotyping	-	GCGTGGACCGCTTGCTGCAACT
<i>LB1-SAIL</i>	<i>mef25-1</i> genotyping	-	GCCTTTTCAGAAATGGATAAATAGCCTTGCTTCC
<i>mef25-1-2F</i>	<i>mef25-1</i> and <i>mef25-2</i> genotyping	Forward	ATGAAGTGTGGGAAAATGGATGAA
<i>mef25-1R</i>	<i>mef25-1</i> genotyping	Reverse	TTTAATGACCCGACTTGCGAACA
<i>mef25-2R</i>	<i>mef25-2</i> genotyping	Reverse	TCCCCGACTGATAGATTCCTGTG
<i>Amino-mef25F</i>	localization	Forward	CACCATGGTACAGACAAAACATTTC
<i>Amino-mef25R</i>	localization	Reverse	GTCCACTGCTTTACACCACAC
<i>nad1ssamp-37F</i>	RT-PCR to verify the editing defect	Forward	TCCTTCAATATCATGATTGGGTC
<i>nad1ssampCA376R</i>	RT-PCR to verify the editing defect	Reverse	GGCATATTTTGAATTACTAGAACA
<i>nad1ssampAC580R</i>	RT-PCR to verify the editing defect	Reverse	GCTCGATTAGTTTCTGCTAGCCA
<i>nad1ssamp-37F</i>	Northern blot probe	Forward	TCCTTCAATATCATGATTGGGTC
<i>nad1-rev</i>	Northern blot probe	Reverse	TTAAGGAAGCCATTGAAAGG
<i>Pro1mef25-SalIF</i>	Promoter fusion to GUS	Forward	AGTCGACTTGGAACCCTTACCAAATTG
<i>Pro2mef25-SalIF</i>	Promoter fusion to GUS	Forward	ATGTCGACTTCGCCTTATCCGGCTGCAAC
<i>Promef25-NcoIR</i>	Promoter fusion to GUS	Reverse	CCCCATGGTGTTTTAATTCTTAGAGC
<i>intron1_nad1-R2B</i>	RT-PCR to obtain unspliced cDNA	Reverse	TTTAGTAGTCTATCAGTGGTGCGAAG
<i>intron1_nad1-R1</i>	RT-PCR to obtain unspliced cDNA	Reverse	GTAGAGTAGATCGTTCAACACCTG

II. 3 Caracterización Funcional del Factor de Edición Mitocondrial 26 (MEF26) ¹

¹ Modificado de: Arenas-M A., Zehrmann A., Moreno S., Takenaka M., Jordana X. (2014) The pentatricopeptide repeat protein MEF26 participates in RNA editing in mitochondrial *cox3* and *nad4* transcripts. *Mitochondrion* 19: 126-134.

Continuando con el trabajo de caracterización de miembros de la familia PPR seleccionados como genes candidatos en la presente tesis, se describe el factor MEF26 por “Mitochondrial Editing Factor 26” que es necesario para la edición completa de dos sitios en los transcritos mitocondriales *cox3* y *nad4*, respectivamente.

Los principales resultados concernientes a esta parte del trabajo de tesis son presentados y discutidos en la forma de un artículo científico, el cual ha sido recientemente publicado en la revista *Mitochondrion*.

TITLE:

The pentatricopeptide repeat protein MEF26 participates in RNA editing in mitochondrial *cox3* and *nad4* transcripts.

Authors:

Anita Arenas-M^a, Anja Zehrmann^b, Sebastian Moreno^a, Mizuki Takenaka^b, and Xavier Jordana^a

^aDepartamento de Genética Molecular y Microbiología, Facultad de Ciencias Biológicas, Pontificia Universidad Católica de Chile, Casilla 114-D, Santiago, Chile.

^bMolekulare Botanik, Universität Ulm, 89069 Ulm, Germany.

e-mail: anitamaribel@gmail.com, anja.zehrmann@uni-ulm.de, samoreno@uc.cl,
mizuki.takenaka@uni-ulm.de, xjordana@bio.puc.cl

Corresponding author: Xavier Jordana, Departamento de Genética Molecular y Microbiología, Facultad de Ciencias Biológicas, Pontificia Universidad Católica de Chile, Casilla 114-D, Santiago, Chile. Phone 56-2-26862668; e-mail xjordana@bio.puc.cl

ABSTRACT:

In angiosperms most members of the large nuclear-encoded family of pentatricopeptide repeat (PPR) proteins are predicted to play relevant roles in the maturation of organellar RNAs. Here we report the novel **Mitochondrial Editing Factor 26**, a DYW-PPR protein involved in RNA editing at two sites in different mitochondrial transcripts. These sites share similar *cis*-elements and application of the recently proposed amino acid code for RNA recognition by PPR proteins ranks them at first and second position of the most probable targets. The two amino acids in each PPR motif involved in RNA recognition are conserved in putative MEF26 orthologs.

Keywords: PPR protein, RNA editing factor, Plant mitochondria, Pentatricopeptide repeat protein, *cox3*, *nad4*

1. Introduction

In flowering plants, the pentatricopeptide repeat proteins (PPR) constitute a large family of proteins with more than 400 members. All PPR proteins investigated so far can form sequence-specific associations with organellar RNAs. These associations can have an effect on different processing and/or translation steps of the target RNA and thereby alter the abundance and variety of the respective gene product (Schmitz-Linneweber and Small, 2008). Several studies have found that mutants of individual PPR genes have compromised functions of specific mitochondrial or plastid genes. These phenotypes are manifested as result of the respective disturbed molecular processing step of organellar transcripts such as RNA editing, splicing, stabilization, and translation (Barkan and Small, 2014).

PPR proteins are characterized by the name giving degenerate motifs of 35 amino acids arranged as tandem repeats of 2-25 such elements. About half of these PPR proteins, the PLS subfamily, also contain shorter (S) or longer (L) motifs than the canonic (P) 35 amino acid elements, usually organized in alternating P-L-S arrangements. Moreover, one, two or three additional domains can be found in the C-terminal region, the E, the E+ and the DYW domains (Lurin et al., 2004). PPR proteins most likely bind to specific RNA sequences with their PPR motifs. A direct correlation of one PPR motif binding to one nucleotide in the RNA target has been suggested for both the canonic (P) and non-canonic (S, S2, L and L2) repeats (Barkan et al., 2012; Takenaka et al., 2013a; Yagi et al., 2013). These studies show that probably only few specific amino acids in each PPR motif interact with defined nucleotide identities in their target transcript, amino acid identities at positions 6 and 1' (the first amino acid of the next repeat) being the relevant ones for specific interaction with the RNA target. This model has been recently supported by the experimental determination of the structure of a complex between a canonic PPR protein (P subfamily) and its target RNA (Yin et al., 2013).

One of the major posttranscriptional RNA maturation processes in plant organelles is RNA editing. This process changes at the RNA level the information encoded by genomic DNA.

The mitochondrial transcriptome from land plants undergoes hundreds of specific C-to-U changes by RNA editing (Giegé and Brennicke, 1999; Notsu et al., 2002; Takenaka et al. 2013b). Many of these events are functionally necessary for the organelles and play important roles in gene expression by e.g. generating start codons, correcting codons to encode conserved amino acids or introducing stop codons (Covello and Gray, 1989; Gualberto et al., 1989; Giegé and Brennicke, 1999; Takenaka et al., 2008). Experiments with transformed plastids and with in organello and in vitro editing systems have shown that a *cis*-region between -20 or -25 and +6 nucleotides relative to the target site is generally necessary and sufficient for editing (e.g. Chaudhuri and Maliga, 1996; Farré et al., 2001; Miyamoto et al., 2002; Neuwirt et al., 2005). It is hypothesized that the plant “editosome” contains an as yet unknown cytidine deaminase and different *trans*-acting factors like MORF proteins and PPR proteins, which interact to edit single or several cytidine residues in their RNA targets (Takenaka et al., 2012 and 2013b; Bentolila et al., 2012). PPR proteins bind specifically to RNA with their PPR motifs, and in the case of RNA editing recognize the *cis*-element at the target site (e.g. Tasaki et al., 2010; Hammani et al., 2011; Okuda et al., 2012). When several chloroplast or mitochondrial editing sites are addressed by the same PPR protein, their upstream sequences often show some sequence similarity (e.g. Hammani et al., 2009; Sung et al., 2010; Verbitskiy et al., 2010).

Presently, all PPR proteins identified to be required for editing at specific sites contain an E domain and about half of them are further extended by E+ or by E+ and DYW domains. Here we report the identification of the *MEF26* gene, which encodes a PPR protein of the DYW subclass. This protein participates in RNA editing at two sites with similar upstream sequences in different mitochondrial transcripts.

2. Materials and methods

2.1. Genotyping of *mef26* mutant plants

Two mutant alleles of At3g03580 (TAIR) were identified, one in the Gene Trap (FGT)

collection generated at the John Innes Center by mobilization of a gene-trap Ds transposable element (line GT_5_22445, *mef26-1* in *Ler* background) and the other a T-DNA insertion mutant line in the GABI-Kat collection (line GK_895H11, *mef26-2* in Col-0 background). Seeds from these lines were obtained from the European Arabidopsis Stock Centre (uNASC, <http://Arabidopsis.info/>, ID: N176714 and N485919). The wild type *Arabidopsis thaliana* ecotypes Columbia-0 (Col-0) and Landsberg erecta (*Ler*) were used as references. Seeds were sown on half-concentrated Murashige and Skoog medium supplemented with 1% (w/v) sucrose and solidified with 0.8% (w/v) agar. After 2 weeks in a 16-h/8-h day/night cycle at 22°C, seedlings were transferred to soil and grown for 4-6 weeks under the same conditions. Alternatively, seeds were directly sown into soil and grown in these conditions. To isolate homozygous mutant plants, genotyping was performed by PCR as described (León et al., 2007), using the specific primers indicated in Supplementary Table 1.

2.2. Screening for affected RNA editing sites and evaluation of editing efficiency

Total RNA was isolated from leaves of six-week-old plants with the TRIzol reagent following the manufacturer's protocol (Invitrogen, <http://www.lifetechnologies>). RNAs from wild type plants and homozygous mutant plants (*mef26-1* and *mef26-2*) were screened by multiplexed single base extension "SNaPshot analysis" for altered RNA editing at specific sites (Takenaka and Brennicke, 2009). To confirm the observed editing defects by SNaPshot, specific cDNA fragments were analyzed. Two µg of RNA were used for cDNA synthesis with M-MLV Reverse Transcriptase (Promega), according to the supplier's instructions and with the specific primers *cox3*-R2 and *nad4*-R2 (Supplementary Table 1). PCR amplifications were carried out in 50 µl, with one-tenth of the cDNA, 0.5 units of Taq DNA polymerase (Invitrogen) and the following primer pairs: for *cox3*, *cox3*-F1 and *cox3*-R2; and for *nad4*, *nad4*-F1 and *nad4*-R2, respectively (Supplementary Table 1). RT-PCR products were purified and either directly sequenced, or cloned into the pGEM-T Easy plasmid (Promega). Twenty to 33 clones were analyzed for each

cDNA by DNA sequencing (Macrogen, Seoul, Korea). RNA editing efficiencies were calculated by the ratio of T and C peak heights in the direct sequencing of RT-PCR products (Figure 2B), using the DNA Dynamo software, or by the ratio of edited to total clones at each editing site (Figure 2C). Sequence logos (<http://weblogo.threeplusone.com>; Crooks et al., 2004) of several editing sites in each transcript were used for graphical representation of the data from clone sequences.

2.3. *MEF26* subcellular localization analysis

The DNA encoding the amino terminal 89 amino acids of MEF 26 was amplified by PCR with primers amino-mef26F and amino-mef26R (Supplementary Table 1) and cloned into pK7FGW2 vector (Karimi et al., 2007) fused to the GFP gene under the control of the CaMV35S promoter. The construct was verified by DNA sequencing, introduced into *Agrobacterium tumefaciens* GV3101 by electroporation and *Arabidopsis* plants were transformed by floral dip (Clough and Bent, 1998). Leaves of several independent transgenic lines with kanamycin resistance were used for preparation of protoplasts as described in Wu et al. (2009). GFP and Mitotracker Orange (Invitrogen) fluorescence in the protoplasts was analyzed with a confocal microscope (Nikon C2+).

2.4. *Complementation of the editing defects in mef26 mutant plants*

The *MEF26* coding region was amplified using *Arabidopsis* Col-0 genomic DNA as a template, with the forward primer amino-mef26F and the reverse primer carboxy-mef26R (Supplementary Table 1). This PCR product was first cloned into pENTR/D-TOPO (Invitrogen) and then recloned into the binary vector pMDC32 (<http://www.arabidopsis.org>), between the CaMV35S promoter and the NOS terminator. T1 generation seeds of the transgenic plants obtained as described above for the localization construct were selected for resistance to hygromycin. Transgene presences in several independent hygromycin resistant lines were verified

by genomic DNA PCR. RNA was prepared from leaves of individual plants, and RT-PCR products from *cox3* and *nad4* transcripts obtained as described above were directly sequenced.

2.5. Northern blot analysis

To characterize the *nad4* transcripts in wild type and *mef26-2* mutant plants, total RNA was isolated from leaves of six-week-old plants as described above. Northern blot analysis was performed as described (Zehrmann et al., 2011), using 7 µg of RNA per lane and a *nad4* ³²P-labeled probe obtained by RT-PCR with primers nad4-F1 and nad4-R2 (Supplementary table 1).

2.6. Prediction of potential MEF26 target editing sites

Potential target RNA editing sites of the MEF26 protein were predicted following the method of Takenaka et al. (2013a). All PPR motifs in MEF26 were aligned to the respective 5' nucleotide sequences of all known RNA editing sites in *Arabidopsis* with the fourth nucleotide upstream of each editing site (nucleotide -4) assigned to the S2 motif. The binding values of each PPR motif domain were calculated with the FIMO program in the MEME suite (<http://meme.nbcr.net/meme/fimo-intro.html>) with 30 nucleotides upstream sequence of 430 mitochondrial RNA editing sites in the respective coding regions. Predicted targets sites are ranked by their p-value.

3. Results

3.1. At3g03580 encodes a DYW-PPR protein targeted to mitochondria

We have initiated a screen to identify *trans*-factors of RNA editing in plant mitochondria by analysis of PPR candidate genes in *Arabidopsis*. This screen recently assigned the factor MEF25 to the *nad1*-308 site (Arenas-M et al., 2013). The candidates were selected based on four criteria: (i) PPR genes with EST or cDNA sequences in public databases; (ii) a mitochondrial destination predicted by subcellular sorting algorithms; (iii) members of the PLS subfamily with

E or DYW motifs; (iv) low sequence similarity with other *Arabidopsis* PPR proteins. We have now analyzed the candidate PPR gene At3g03580, which encodes an E-DYW- PPR protein of 882 amino acids and is predicted to be targeted to mitochondria by two algorithms, iPSORT and Mitoprot2 (Bannai et al., 2002; Claros and Vincens, 1996).

In order to confirm experimentally these targeting predictions, a DNA fragment encoding the N-terminal 89 amino acids was fused to the green-fluorescent protein (GFP) and transgenic lines were generated. GFP and Mitotracker-Orange fluorescence co-localized in protoplasts from these transgenic lines and were clearly distinct from chloroplast fluorescence (aminoAT3G03580-GFP in Figure 1), confirming that the AT3G03580 protein is targeted to mitochondria but not to chloroplasts.

3.2. *AT3G03580 (MEF26) is involved in RNA editing of two mitochondrial transcripts*

We named the protein encoded by At3g03580 MITOCHONDRIAL EDITING FACTOR 26 (MEF26) after the molecular phenotype of its mutants (see below). To determine the molecular function of MEF26, we obtained two independent At3g03580 disrupted mutants, bearing a Ds insertion (*mef26-1*) and a T-DNA insertion (*mef26-2*), respectively. Homozygous mutant plants of both lines were selected and the positions of insertions were confirmed by PCR and sequencing to be in the coding region, 18 bp (*mef26-1*) and 947 bp (*mef26-2*) downstream of the ATG initiation codon, respectively (Figure 2A, Supplementary Figure 1).

To evaluate the possibility that MEF26 protein is a RNA editing factor, we examined the RNA editing status of leaf mitochondrial and chloroplast transcripts from wild type and *mef26* mutant plants by the multiplexed SNaPShot method (Takenaka and Brennicke, 2009). With this assay the editing status of more than 350 mitochondrial sites and all plastid sites were monitored. No editing defects were found in plastid transcripts, which is consistent with our localization results. In contrast, we found defects at two sites in two mitochondrial transcripts, *cox3-311* and *nad4-166*, in *mef26* mutant plants. Other known editing sites that were not monitored by the

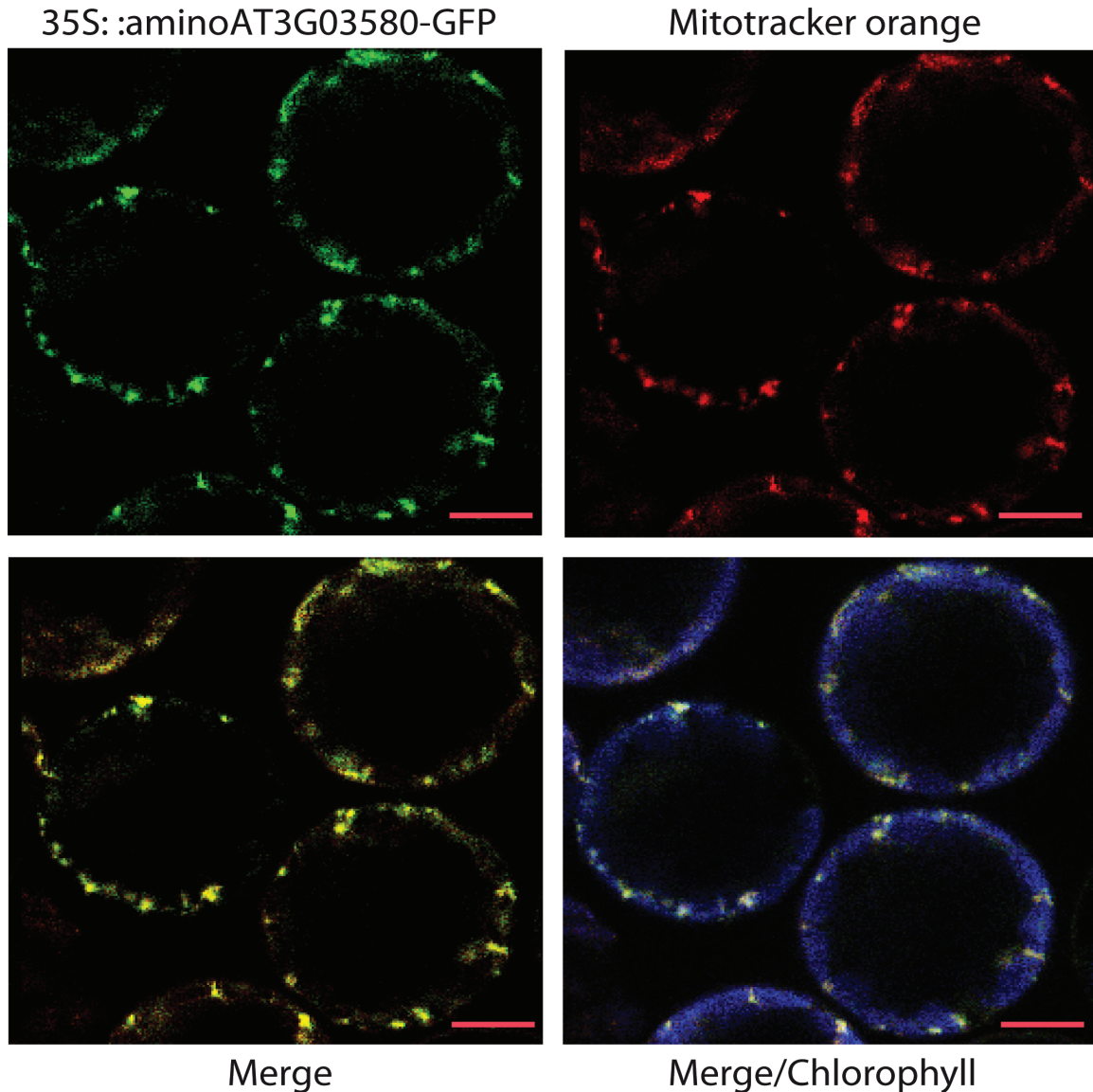


Figure 1. The AT3G03580 (MEF26) amino terminus targets GFP to mitochondria. The amino MEF26-GFP fusion protein (see Figure 2A) was expressed in transgenic *Arabidopsis* plants under the control of the CaMV35S promoter. Protoplasts isolated from leaves were incubated with Mitotracker Orange, and GFP fluorescence (in green), Mitotracker Orange fluorescence (in red) and chlorophyll autofluorescence (in blue) were analyzed by confocal microscopy. The yellow signals in the merge panel indicate that the MEF26-GFP proteins are located in mitochondria. Overlay of the merged image and the chlorophyll autofluorescence shows no chloroplast localization of the amino MEF26-GFP fusion proteins. Scale bar = 5 μ m.

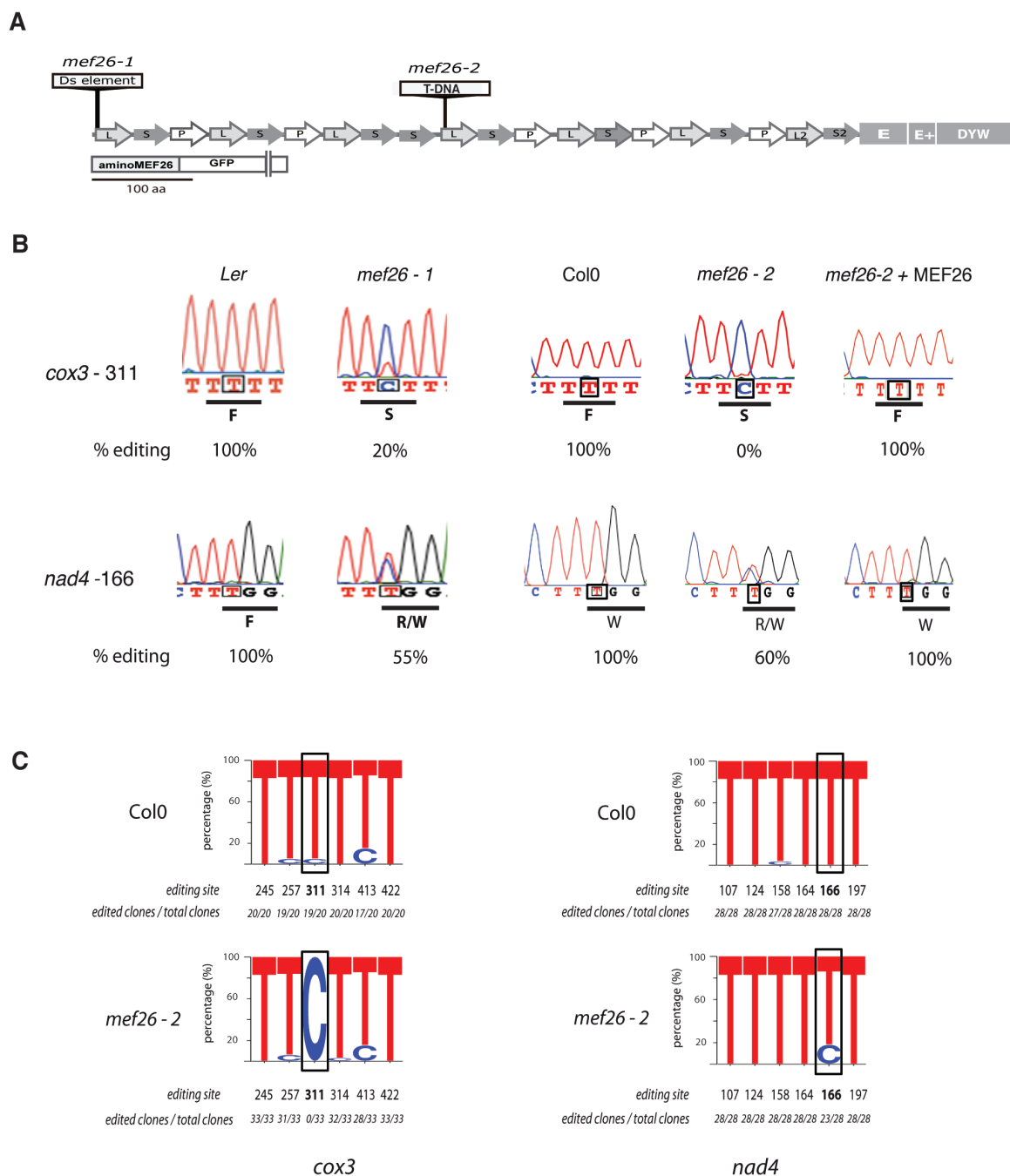
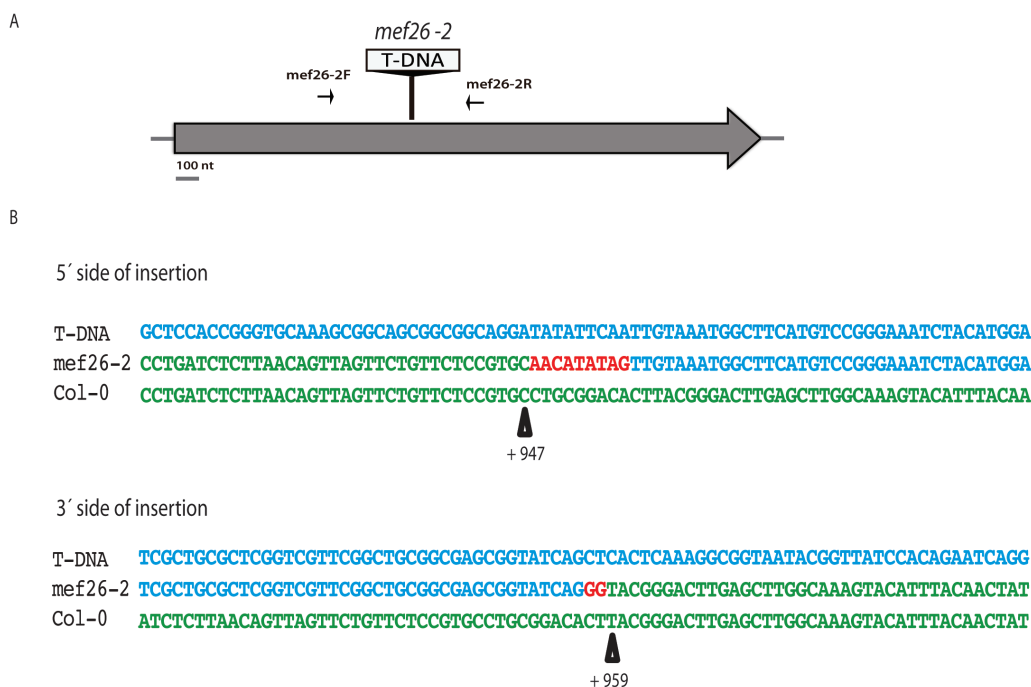


Figure 2. MEF26 protein is involved in two editing events in different mitochondrial transcripts. (A) MEF26 structure and the Ds and T-DNA insertion sites in the *mef26-1* and *mef26-2* mutants. Structural analysis of MEF26 (882 amino acids) predicts 20 pentatricopeptide repeats (PPR) (arrows). P represents the classic PPR motif, L and L2 the PPR-like longer motifs, S and S2 the

PPR-like shorter motifs. E and E+ denote the extension motifs and DYW is the last motif at the C-terminus (grey boxes). The insertions in the two mutant lines are located 18 bp (Ds element, *mef26-1*) and 947 bp (T-DNA, *mef26-2*, details given in Supplemental Fig.1) downstream of the ATG codon. The N-terminal 89 amino acids sequence was fused to GFP for the subcellular localization analysis (Figure 1) as shown below the MEF26 structure. (B) Sequence analysis of the bulk RT-PCR products for *cox3* and *nad4* from the mutant lines *mef26-1* and *mef26-2* and respective background wild type plants, Ler and Col-0, and the complemented plants. Nucleotides at *cox3*-311 and *nad4*-166 specifically affected in *mef26* mutants are boxed. The codon frames containing the editing sites are underlined, and the amino acid translated from the predominant mRNA is indicated below. RNA editing efficiencies were calculated by the ratio of T and C peak heights. (C) Detailed analysis of editing in *cox3* and *nad4* transcripts from wild type Col-0 and the *mef26-2* mutant plants. The editing status of seven *cox3* sites (6 are shown) and 15 *nad4* sites (6 shown) was determined by sequence analysis of cloned cDNAs. The number of edited/total clones at each site is indicated and the ratio is visualized by sequence logos (<http://weblogo.threeplusone.com>; Crooks et al., 2004). RNA editing is affected only at the *cox3*-311 and *nad4*-166 sites in the *mef26* mutant.



Supplementary Fig. 1. Characterization of the T-DNA insertion site in the *mef26-2* mutant. (A) PCR amplification of *MEF26*/T-DNA junctions in *mef26-2* was performed with either primers *mef26-2F* and the left border primer (LB GK08409) or *mef26-2R* and the right border primer (RB GK03144). (B) Sequencing of the PCR products established that no major deletions took place during the insertional event. A eleven bp deletion (between +948 and +958) and two insertions (9 and 2 bp, in red) were found in *mef26-2*. In both alignments the first line corresponds to the T-DNA sequence, the second line to the junction between *MEF26* and T-DNA in the *mef26-2* mutant and the third line to the wild type *MEF26* sequence. T-DNA and *MEF26* sequences are in blue and green, respectively. Numbers are in relation to the first nucleotide of the ATG start codon from *MEF26*.

SNaPshot approach were analyzed by direct sequencing of cDNAs. No additional site is affected in the *mef26-1* line, among 434 analyzed sites.

To confirm these changes in editing status, cDNA fragments encompassing several editing sites in *cox3* and *nad4* transcripts were directly sequenced (Figure 2B). Among 17 editing sites analyzed in *nad4* transcripts (all in the first exon), the only difference between wild type and mutant plants was found at site 166, which showed 40-45 % reduction of the editing extent in both mutant lines (Figure 2B). In the *cox3* transcript, only one site (*cox3*-311) out of seven analyzed sites was affected, editing being completely abolished in *mef26-2*, and reduced to less than 20 % in *mef26-1* plants (Figure 2B).

In both *cox3* and *nad4* transcripts, the nearby sites located three nucleotides downstream (*cox3*-314) and two nucleotides upstream (*nad4*-164) of the affected sites seemed to be normally edited in mutant plants. However, direct sequence analysis might not detect minor (less than 20%) defects at these sites. To detect minor defects at these neighboring sites, *cox3* and *nad4* cDNAs were cloned and individual clones were sequenced. The neighboring sites *nad4*-164 and *cox3*-314 are edited in 32 of 33 and 28 of 28 clones, respectively (Figure 2C), suggesting that the MEF26 protein is specifically involved in *cox3*-311 and *nad4*-166 sites. While the sequence analysis of cloned cDNAs shows totally abolished editing at site *cox3*-311 as observed by direct cDNA sequencing, only 20% of the cDNA clones are in the unedited form at *nad4*-166 in mutant *mef26-2* in comparison to the less than 20% seen by direct sequence analysis of bulk cDNAs. Different efficiencies observed between direct cDNA sequence and cloned cDNA analyses have also been reported in other publications (e.g. Kim et al., 2009; García-Andrade et al., 2013).

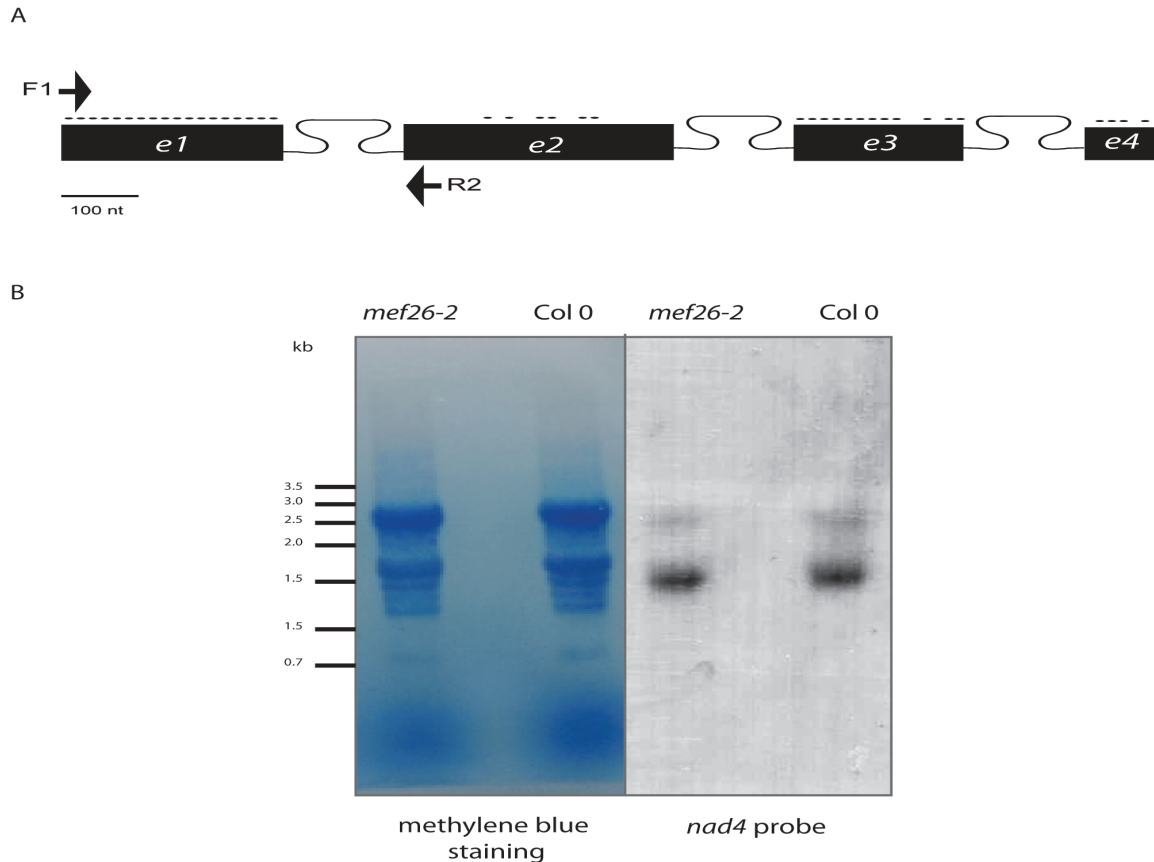
To confirm that the defect in RNA editing observed in the *mef26* mutant plants is indeed due to the identified *MEF26* disruption, we introduced the full-length *MEF26* gene, which does not contain introns, under the control of the CaMV35S promoter into homozygous *mef26-2* mutant plants. RNA editing in the transgenic lines was completely restored at both *cox3*-311 and *nad4*-166 sites (Figure 2B).

In contrast to the drastic reduction of editing at the *cox3*-311 site, editing at *nad4*-166 is only partially affected in both *mef26* mutant plant lines. Some RNA binding proteins including PPR proteins are involved in various RNA processing steps other than editing and indirectly change the RNA editing efficiencies at a given target site (e.g. Iwabuchi et al., 1993; Chateigner-Boutin et al., 2011). To analyze whether the partial editing defect at the *nad4*-166 site is due to the indirect influence of other RNA processing steps in the *nad4* transcript, we performed an RNA gel blot analysis of *nad4* transcripts. No significant differences in the *nad4* transcript profiles were observed in leaves of wild type and *mef26-2* mutant plants (Supplementary Figure 2), suggesting that reduced RNA editing at *nad4*-166 in *mef26* mutants is not a secondary effect of an unrelated RNA processing event in the *nad4* transcript.

In wild type plants, *cox3*-311 and *nad4*-166 sites are completely edited, changing codon identities from Ser (UCU) to Phe (UUU) and Arg (CGG) to Trp (UGG), respectively. These amino acids (Phe and Trp) are highly conserved in plants and Phe104 is present in COX3 also in other organisms (Figure 3). Thus the lack of editing at *cox3*-311, at least in the *mef26-2* mutant, would generate a COX3 polypeptide having serine at this otherwise conserved phenylalanine, and editing reduction at *nad4*-166 would decrease the amount or percentage of the NAD4 polypeptide having a wild type amino acid sequence. In spite of these effects, no changes were seen in the macroscopic phenotype or in the development of homozygous *mef26-1* and *mef26-2* plants, compared to wild type plants under standard growth conditions (data not shown). Neither was pollen viability affected. These results suggest that the lack of editing at *cox3*-311 and the editing reduction at *nad4*-166 have little effect on normal mitochondrial function.

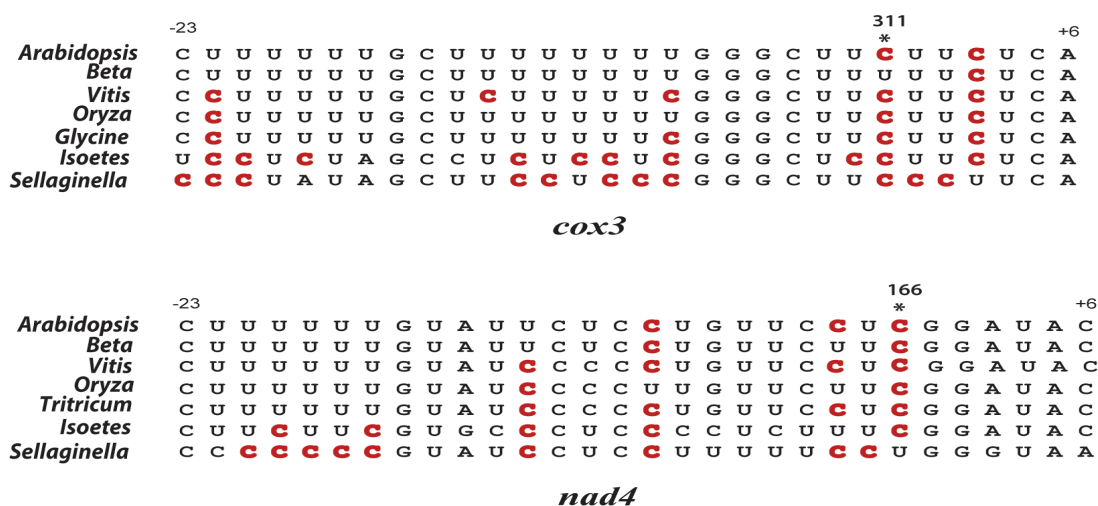
3.3. *Cis-elements of the two MEF26 target sites have similarity*

In plant mitochondria, several PPR proteins have been described to be RNA editing factors for more than one site (Takenaka et al., 2013b). In these cases, one *trans*-factor probably



Supplementary Fig. 2. Northern blot analysis of the *nad4* transcripts in wild type Col-0 and *mef26-2* mutant plants. (A) Exon-intron structure of the the *nad4* gene in *Arabidopsis*. Four exons (black boxes), three *cis*-spliced introns (closed lines) and forty-one editing sites (black dashes) are indicated. The defective editing site in *mef26* (*nad4*-166) locates in exon 1. (B) Seven μ g of total RNA from *mef26-2* and Col-0 leaves were separated by electrophoresis on a formaldehyde agarose gel and transferred onto a nylon membrane. The methylene blue stained membrane (left panel) indicates no degradation of RNA and serve as a loading and transfer control. Hybridization was performed with a probe obtained by RT-PCR with primers *nad4*-F1 and *nad4*-R2. Similar signal patterns of *nad4* between Col-0 and *mef26-2* (right panel) suggest the normal intron splicing of *nad4* in the *mef26-2* mutant (right panel). Positions of DNA size standards are indicated on the left in kilobase pairs.

A



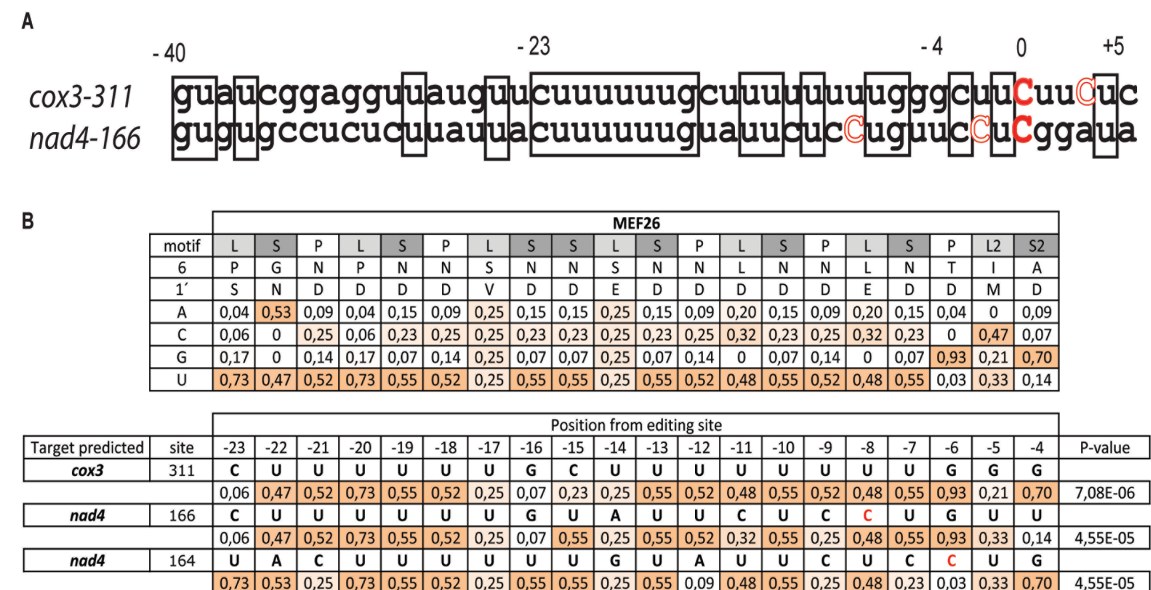
REDIdb (http://biologia.unical.it/py_script/REDIdb/). Lycophyte (*Isoetes engelmannii*) sequences are taken from Grewe et al. (2012), *Selaginella moellendorffii* sequences from Hecht et al. (2011), *Bos taurus* sequences from Shinzawa et al. (2007) and *Thermus thermophilus* amino acid sequences from Baradaran et al. (2013). Identical residues are shown with white letters on black background, and amino acids altered by RNA editing events in plants are boxed. Other species are *Arabidopsis thaliana*, *Beta vulgaris*, *Vitis vinifera*, *Oryza sativa*, *Triticum aestivum* and *Glycine max*.

recognizes the multiple *cis*-elements around the different cytidines to be edited. *In organello* and *in vitro* experiments revealed that normally a region of 20-25 nucleotides upstream and up to 6 nucleotides downstream is sufficient for RNA editing at specific cytidines and covers the essential elements (e.g. Choury et al., 2004; Takenaka et al., 2004). When we aligned the RNA sequences in a window from -40 to +5 of *nad4*-166 and *cox3*-311, we found that up to 64% of the nucleotides are shared within the 25 nucleotides upstream of the edited C (Figure 4A). Interestingly, this similarity increases to 72% after editing at *nad4*-158 and *nad4*-164. This is higher than the overall percentage of nucleotides shared between the sequences from -40 to -26 and those in the region downstream of the edited C, between +1 to +5. This observation suggests that the upstream 25 nt regions contain the recognition and binding region for the MEF26 protein.

According to the recently proposed hypothesis that one PPR module recognizes one nucleotide, MEF26, which has 20 PPR motifs, could bind to the -23 to -4 sequence regions of the target sites. The proposed PPR-nucleotide recognition codes (Takenaka et al., 2013a) of MEF26 clearly match the -23 to -4 upstream sequence of both *cox3*-311 and *nad4*-166 target sites (Figure 4B). Comparison of the binding probabilities of MEF26 to each of the 430 mitochondrial editing sites ranks the *cox3*-311 at 1st and the *nad4*-166 at 2nd position of the most probable targets, respectively (Figure 4B). Interestingly, the *nad4*-164 site located only two nucleotides apart from *nad4*-166 exhibits a binding probability similar to *nad4*-166. However, bulk sequencing of RT-PCR products and sequencing of cDNA clones (more than 50 in several experiments) showed that *nad4*-164 editing is not affected in *mef26* mutant plants (Figure 2B and C).

3.4. Putative MEF26 orthologs

A BLAST search using the whole length of MEF26 as a query identified putative orthologs with 95-54% identity in *Arabidopsis lyrata*, *Capsella rubella*, *Eutrema salsugineum*, *Vitis vinifera*, *Theobroma cacao*, *Cucumis sativus*, *Solanum tuberosum*, *Solanum lycopersicum* and *Fragaria vesca* (Supplementary Figure 3). MEF26 has similarities with these putative



(<http://meme.nbcr.net/meme/fimo-intro.html>) (far right column).

orthologs throughout the whole protein sequence, except that the *F. vesca* protein lacks the E' and DYW domains. Interestingly, amino acids at positions 6 and 1' in P and S motifs and at position 6 of the L motifs, which are proposed to be necessary for RNA recognition, are highly conserved in all species (Figure 5). For those species possessing a putative MEF26 protein and for which *cox3* and *nad4* genomic sequences are available (six species for *cox3*, four for *nad4*), cytidines are always found at positions *cox3*-311 and *nad4*-166, and are edited to U if RNA sequence data is available (four species for *cox3*, two for *nad4*). These data suggest that these putative MEF26 orthologs are involved in the same editing events in the respective other plant species.

However, this correlation between editing at *cox3*-311 and *nad4*-166, and the presence of MEF26-like proteins does not extend to all other plants. For instance, although *cox3*-311 (our data) and *nad4*-166 sites are edited in *Oryza sativa*, and the *cox3*-311 site is edited in *Glycine max* (data not available for *nad4*), no candidate orthologs of MEF26 could be identified in the complete genomes of these plants. Furthermore, we could not find candidates for MEF26 orthologs in the complete genomes of the monocots *Sorghum bicolor*, *Setaria italica* and *Brachypodium distachion* and the dicots *Cicer arietinum*, *Medicago truncatula* and *Populus trichocarpa*. In each one of these species several PPR proteins display overall identities of 35 to 40% to MEF26, but these proteins are still more similar to other *Arabidopsis* PPR proteins. For instance, the *Oryza sativa* and *G. max* PPR proteins closest to the MEF26 sequence (OS01G10800.1 and XP-006574752.1) are more similar to AT4G18750 and AT4G13650 respectively, and the amino acids located at positions 6,1' in its PPR motifs are different from those in MEF26 and the other putative orthologs (Figure 5). Thus, other specificity factors for editing at *cox3*-311 and *nad4*-166 sites must be encoded and employed in these plant species.

```

MEF26      1  MOTR-----VSSPFI SRALSSSSNLNELRRIHALVISLGLDGSDFSGKL
A.lyrata   1  MOTR-----VSSAFI SRALSSSSNLNELRRIHALVISLGLDGSDFSGKL
C.rubella  1  MOTGV-----SSSPFI SKALSSSSNLNELRRIHALVISLGLDGSDFSGKL
E.salsugineum 1  MQIR-----VSSAFI SKALSSSSNLKDLRIHALVISLGLDGSDFSGKL
T.cacao    1  MKTFNT-----LHEARHQILYSSITKALSSVSNSKOLHIHSIIITLGLNSVLFSGKL
F.vesca    1  MRTNVSS----FYGLTKQALYSLSFKTLSSITKTKTQLHKLHSLIITLGLHHSVLFSGKL
V.vinifera 1  MKILRLVH-----ECSRQTLFSSISRALASAATTTQLHKLHSLIITLGLHHSVLFSAKL
C.sativum  1  MKPPKFCNFNNTPEPSQEFRLRSSLTKTLSSAKNTPQLRTVHSLIITLGLSLSVLFSGKL
S.lycopersicum 1  MKTIKLGIVG----QRTEYWFHSLITLRLSSVINQTDLHKVHSLIVVSGQHQSDFCGKL

```

```

MEF26      46  IDKYSHFREPASSLSVFRRVSPAKNVYLWNSIIRAFSKNGLFPEALEFYGKLRRESKVSPD
A.lyrata   46  IDKYSHFREPASSLSVFRRVSPAKNVYLWNSIIRAFSKNGWFPKALEFYGKLRRESKVSPD
C.rubella  48  IDKYSHFREPASSLSVFRRVSPAKNVYLWNSIIRAFCKNGLYPKALEFYGKLRESKVSPD
E.salsugineum 46  IDKYSHFREPASSLSVFRRVSPAKNVYLWNSIIRALSNGLFPPKALEFYGKLRESKVSPD
T.cacao    55  ISKYAQFKDPTSSLSVFRVSTSNVYQWNSVIRALTHNGLFSEALGFYQMRKMDVLDPD
F.vesca    57  ISKYSNLKHPSSSLSVFRHFRPKHSPYLWNSIIRALTHNGLHSEALRHYNAMLHNVRPD
V.vinifera 55  IAKYAHFRDPTSSFSVFRLASPSNNVYLWNSIIRALTHNGLFSEALSLYSETQIRIOPD
C.sativum  61  ISKYAQVKDPTSSSVVFRSISPTNNVYLWNSIIRALTHNGLFTQALGYTEMREKKLOPD
S.lycopersicum 57  ISKYSQFKDPVSSLSIFRINSPTNNVYLWNTIIRALTHNGLMSKALGFYQMRKLNVPKPD

```

```

MEF26      106 KYTFPSVIKACAGLFD AEMGDLVYEQILDGMGFESDLVGNALVDMYSRMGLLRARQVFD
A.lyrata   106 KYTFPSVIKACAGLFD AEMGDLVYEQILEMGFESDLVGNALVDMYSRMGLLRARQVFD
C.rubella  108 KYTFPSVIKACAGLFD AETGDLVYEQILEMGFESDLVGNALVDMYSRMGLLRARQVFD
E.salsugineum 106 KETFPFVKACAGLFD AETGDLVYEQILEMGFESDLVGNALVDMYSRMGLLRARQVFD
T.cacao    115 KYTFPSVANS CAALVDIEMGKVHENVLDMGLGSDLYIGNALVDMYARFGCLAEALKVFN
F.vesca    117 SHTFPSVINACAALCDLEMGLVTHRRVSETGFGTDLVVCNALIDMYARIGELGHARQVFD
V.vinifera 115 TYTFPSVINACAGLDFEMAKSHDRVLDMGFGSDLYIGNALIDMYCRFNDLDRARKVFE
C.sativum  121 APTFPSPVINS CARILDELGCITVEHAMEMGFESDLYIGNALIDMYSRFVDLDNARYVFE
S.lycopersicum 117 NYTFPSIINSCGSLDLEMVKIVHNEVSEMGFGSDLYIGNALIDMYARMNELGRARVFD

```

```

MEF26      166 EMPVRDLVSWNSLISGYSSHGYYEEALEIYHELKNSWIVPDSFTVSSVLPAFANLLVVKQ
A.lyrata   166 EMPVRDLVSWNSLISGYSSHGYYEEALEIYHELKNSWIVPDSFTVSSVLPAFANLLVVKQ
C.rubella  168 AMPVRDLVSWNSLISGYSSHGYYEEALEIYNELKKYWIVPDSFTVSSVLPAFANLLVVKQ
E.salsugineum 166 EMPVRDLVSWNSLISGCSHGYYEEALGIYHELKNSWIVPDSFTFSSVLPAFANLLVVKQ
T.cacao    175 GMPERDVSWNSLISGYSANGYNEEALVYNMARMAGIIPDCYTVSSVLPACGGLVDVKE
F.vesca    177 EMRERDVSWNSLISGYSNGYNEEALMFWRLEMEGLVPDCFSISSVLPACGGLVDVKE
V.vinifera 175 EMPLRDVSWNSLISGYNANGYNEALEIYYRFRNLGVDPDSYTMSSVLRACGGLGSVEE
C.sativum  181 EMSNRDSVSWNSLISGYCSNGFWEDALDMYHKFRMTGMVPDCFTMSSVLLACGSLMAVKE
S.lycopersicum 177 EMPSRDVSWNSLVSGYSANGYNEEALVFRGRLSGVAADFTVSSVLPACGGLMEVEQ

```

```

MEF26      226 GQGLHGFALKSGVNSVVVNNGLVAMYLKFSRPTDARRVDEMVRDVSNTMTCGYLK
A.lyrata   226 GQGLHGFALKSGVNSVVVNNGLVAMYLKFSRPTDARRVDEMVRDVSNTMTCGYLK
C.rubella  228 GQGLHGFALKSGVNSVVVDNGLVAMYLKFSRPTDARRVDEMVRDVSNTMTCGYLN
E.salsugineum 226 GQGLHGFALKSGVNSVVVNNGLVAMYLKFSRPTDARRVDEMVRDVSNTMTCGYLN
T.cacao    235 GEVHCHLVEKIGLHRDVVVSNGLSMYFKFNRLVDARRIDEMVRDVTWNTMTCGYSQ
F.vesca    237 GRLVHGLVEKIGVNADVLVSNGLSMYFKFGWCRDARIFEEMVVRDVSVWNTMTCGYLQ
V.vinifera 235 GDIHGLIEKIGIKKDVVNNGLSMYCKFNGLIDGRRIFDKMVLRAVSWNTMTCGYSQ
C.sativum  241 GVAHVGVEKIGIAGDVIIIGNLSMYFKFERLRARRVFSKMAVSDVTWNTMTCGYAQ
S.lycopersicum 237 GQMVHGLVEKSGIKGDVAVSNGLSMYFKFERLLDCORIFDEMIYRDIVTWNITICGFSH

```

```

MEF26      286 LEMVEESVRMFLENLDQFKPDLLTVSSVLRACGHLRDLAKAYIYNMRLKAGFVLESTVR
A.lyrata   286 LEMVEESVRMFLENLDQFKPDILTVTSVLACGHLRDLAKAYIYNMRLRAGFVLESTVR
C.rubella  288 LEMHEASVRMFLENLDQFKPDILTASSILRACGHLRDLAKAYVHDYVLRAGFKLDTTVK
E.salsugineum 286 SEMVEESLRMFLENLDQFKPDILTSSILRACGHLRDLAKAYVHEYMLRSGFVLETTGS
T.cacao    295 MELFKESILFMQMNKFEPLDITITSVLACGHLRDLLEFGKFVHEYMKRSRYESDITAD
F.vesca    297 LCLFDEAISIFREMKEFTPDLLITITSVLRACSHLRDLAKAYVHGYMKRSRSGFEPTMAN
V.vinifera 295 VGLYEESIKLFMEMVNOFKPDLLITITSLQACGHLGLDLEFGKYVHDYMITSGFECDTTAS
C.sativum  301 LGRHEASVKLFMDMDGCFVPDMLSTITIRACGQSGDLOVGKVFVHKYLGSGFECDTVAC
S.lycopersicum 297 SGLYQESIKLFQEMVDEHKPDLLITITSVLQACGHMGDLRFGRFVHDYILENRECDTTAC

```

MEF26 346 NILIDVYAKCGDMITARDVFN SMECKDTVSWNSIISGYIQSGDLMEAMKLFK--MMIMEE
A. lyrata 346 NILIDVYAKCGDMITARDVFN SMECKDTVSWNSIISGYIQSGDLMEAMKLFK--MMIMEE
C. rubella 348 NILIDVYAKCADMTARDVFKSMECKDTVSWNSIISGYIQNGDLSEAMKLFRLMMIMEE
E. salsugineum 346 NILIDAYAKCGDMITARDVFKSMECKDTVSWNSIISGYIQSGDLLEAMKLFKMMIMEE
T. cacao 355 NILIDMYSKCGDLLASREVFDRIICRDSVSWNSIINGYFOYGYKYEAVKLFK--IMKIDS
F. vesca 357 NIATDMHAKCGSLIASREVFDYMECRDSVSWNSMINGYFLNGCFDECFNLFK--MMRNE
V. vinifera 355 NILINMYAKCGNLLASQEVFSGMCKDSVSWNSMINVYIQNGSFDEAMKLFK--MMKTDV
C. sativum 361 NILIDMYAKCGDLLAQEVFDITCKKDSVTWNSLINGYTQSGYYKEGLESEFK--MMKMER
S. lycopersicum 357 NILINMYAKCGDLVAARQVFDNMKRWDLVSWNSIISGYFENGLNKEAVDLLK--MMRIDL

MEF26 405 QADHITYLMLISVSTRDLADLKFGKGLHSNGIKSGICIDLSVSNALIDMYAKCGEVGDSLK
A. lyrata 405 QADHITYLMLISLSTRDLADLKFGKGLHSNGIKSGIYIDLVSNALIDMYAKCGEVGDSLK
C. rubella 408 QADHITYLMLISVSTRDLADLKFGKGLHSNVKSGINFDLVGNLIDMYAKCGEVGDSLK
E. salsugineum 406 QADHITYVMLISVSTRDLADLKFGKGLHSNATKSGIYSDLSVSNALIDMYAKCGEVGDSLQ
T. cacao 413 KVDSTITCVMLISASTOLADKDLGKKIHCDVTKLGFDSDIINNAMIDMYAKCGQINDSMK
F. vesca 415 KPDSVSCVMLISMTOLGQVDQGMHCDIVKLGFDSDLVNNVLDLYAKCGKQDALK
V. vinifera 413 KPDSVTYVMLISMSTOLGDLHLGKELHCDLAKMGFNSNVVSNLIDMYAKCGEMGDSLK
C. sativum 419 KPDSVTFVILLISIFSOADINOGAGIHCDVIKFGFEALIGNSLIDVYAKCGEMDILLK
S. lycopersicum 415 QPDSVTFVILLISMCTKIMDVDFTELEHCDIIRKGYDSTLIVGNALIDVYAKCGREHSVW

MEF26 465 IFSSMGTDITVTWNTVISACVRFGDFATGLOVITQMRKSEVVPDMATFLVTLPMCASLAA
A. lyrata 465 IFNSMGTDITVTWNTVISACVRFGDFATGLOVITQMRKNKVVPDMATFLVTLPMCASLAA
C. rubella 468 IFNSMETRDTVTWNTVISACVSSGDFATGLOVITQMRKSEVVPDMATFLVTLPMCASLAA
E. salsugineum 466 IFNCMDTRDTVTWNTVISSCVCFGDFATGLOITQMRITSGAVLDMATFLVILPMCASLAA
T. cacao 473 IPEYMKTHDRVSWNTIIACVQSGDFTLGLKIHOIRTEGLRADVATILGILPMCFLLAA
F. vesca 475 VEDSMTARDITWNSIIISACFHNECSLGLRVLRRMRNEGVMLDVATMLGILPVCSLLAA
V. vinifera 473 VFENMKARDIITWNTIIASCVHSEDCNLGLRISRMRTGCVTPDMATMLISILPVCSLLAA
C. sativum 479 VFSYMSAHDIIISWNTVIASSVHFDDCTVGQMINEMRTEGLMPDEATVLGILPMCSLLAV
S. lycopersicum 475 QFEIMTSRDITVTWNTIIAACSHEFSYLGKILSRMRTEGLMPDVATILGSLPLVCSLLAA

MEF26 525 KRIGKEIHCCLLRFGYESELOIGNALIEMYSKCGCLESSEFRVFERMSRRDVVTWTGMIYA
A. lyrata 525 KRLGKEIHCCLLRFGYESELOIGNALIEMYSKCGCLESSEFRVFERMSRRDVVTWTGMIYA
C. rubella 528 KRLGKEIHCCLLRFGYESELOVGNALIEMYSKCGCLESSEFRVFAHMSRRDIVTWTGMIYA
E. salsugineum 526 KRPGKEIHCCLLRFGYESELPIGNALIEMYSKCGSLESSEFSVFRHMSRRDVVTWTGMIYA
T. cacao 533 KROGQEIHCIFRLGFETDVPVGNALIEMYSKCSSLTISLEVFDSMKVRDVVTWTAMISA
F. vesca 535 KRKGQEIHCIFRLGFESDVPVGNALIEMYSCGSLETIRVFEQMCITKDVVTWTALISA
V. vinifera 533 KROGKEIHCCIFRLGFESDVPVGNVLIEMYSKCGSLRNSFQVFKLMTKDVVTWTALISA
C. sativum 539 RROGKEIHGYIFKSGFESNVPIGNALIEMYSKCGSLENCIKVFKYMKEDVVTWTALISA
S. lycopersicum 535 KROGKELHGFIIIRLKFESEQVPVGNALIEMYSKTGSCLKNAISVFEHMSIKDVVTWTAMISA

MEF26 585 YGMYGEGEKALETFADMEKSGIVPDSVVFIAIIYACSHSGLVDEGLACFEKMKTHYKIDP
A. lyrata 585 YGMYGEGEKALESFVMEKSGIVPDSVVFIAIIYACSHSGLVEKGLACFEKMKTHYKIDP
C. rubella 588 YGMYGEGEKALKTFADMEKSGIVPDNVVFIAIIYACSHSGLVEEGLACFEKMKTHYKIDP
E. salsugineum 586 YGMYGEGEKALEAFADMEKSGIVPDHVVFAIIYACSHSGLVEEGLACFEKMKTHYKIDP
T. cacao 593 YGMYGEGKALRAFAMEATGVIPDHVAFVAIIYACSHSGLVEEGLASFDRMKKVYNLEP
F. vesca 595 YGTYGEGEKALIAFOKMEATGVLPDHLAFLAIIYACSHSGLVEQGLAYFDRMKKHYNLEP
V. vinifera 593 CGMYGEGKALRAFGEEMAAAGIVPDHVAIIYACSHSGLVEEGLNYFHRMKKDYKIEP
C. sativum 599 FGMYGEGKALKAFQDMELSGVLPDSVAFIAFIYACSHSGMVKEGLRFFDRMKTDYNLEP
S. lycopersicum 595 YGMYGEGKALRSFOQMKETGTIPDHIVFAIIYACSHSGLVOEGRACFNOMRKYNIIEP

MEF26 645 MIEHYACVDLLSRSQISKAEETQAMPIKPDASIWASVLRACRTSGDMETAERVSRII
A. lyrata 645 MIEHYACVDLLSRSQISKAEETQAMPIEPDASIWASVLRACRTSGDMETAERVSRII
C. rubella 648 MIEHYACVDLLSRSQISKAEETQTMPIKPDASIWASVLRACRTSRDMETAERVSRII
E. salsugineum 646 VIEHYACVDLLSRSQITKAEETQAMPIKPDASIWASLLACRTSGDMETAERVSRII
T. cacao 653 RIEHYACVDLLSRSGLISKAEETYSMPMKPDASIWGSLLSACRSCGNIEVAORVSERI
F. vesca 655 QIEHYACVDLLSRSGLVQAEETIHSMPIKADASIWGSLLSACREAGEEMACRVSEOI
V. vinifera 653 RIEHYACVDLLSRSGLLDKAEETILSMPLKPDSSIWGSLLSACRMSGDTETIAERVSERI
C. sativum 659 RIEHYACVDLLSRSGLLAQAEETILSMPLKPDASLWGSLLSACRARGCNTNIAORVSKKI
S. lycopersicum 655 RIEHYACVDLLSRSGLLAEAEETILSMPLKPDASMWGSLLSACRMSGDTVTAERVVERL

MEF26	705	IELNPDDPGYSILASNAYAAALRKWDKVS	LI	RKSLKDKHITKKNPGYSWIEVGK	NVHVFSSG
<i>A. lyrata</i>	705	IELNPDDPGYSILASNAYAAALRKWDKVS	LI	RKSLKDKHITKKNPGYSWIEVGK	NVHVFCSG
<i>C. rubella</i>	708	IELNPDDPGYSILASNAYAAALRKWDKVS	LI	RKSLNDKLI	RKKNPGYSWIEGKIVHVF
<i>E. salsugineum</i>	706	IELNPDDPGYSILASNAYAAALRKWDKVS	LI	RKSLKDKHITKKNPGYSWIEGK	NVHVF
<i>T. cacao</i>	713	IELNSDITGYVVLASNVAALGKWDQVR	MI	RKSLKDKHITKKNPGYSWIEGK	NVHVF
<i>F. vesca</i>	715	VEFNSYDTGYHVLVSNVYAALGKWDKVR	MI	RKSLKDKHITKKNPGYSWIEGK	NVHVF
<i>V. vinifera</i>	713	IELNPDDITGYVVLVSNVAALGKWDQVR	MI	RKSLKDKHITKKNPGYSWIEGK	NVHVF
<i>C. sativum</i>	719	IELNSDDITGYVVLVSNVAALGKWDQVR	MI	RKSLKDKHITKKNPGYSWIEGK	NVHVF
<i>S. lycopersicum</i>	715	VELNSDDPGYNVLASNVAALGKWDQVR	MI	RKSLKDKHITKKNPGYSWIEGK	NVHVF

MEF26	765	DDSAPQSEATYKSLLEILYSLMAKEGYIPDP	REV	SQNL	EEEEKRRILICGHSERLAI
<i>A. lyrata</i>	765	DDSAPQSEATYKSLLEILYSLMAKEGYIPDP	REV	SQNL	EEEEKRRILICGHSERLAI
<i>C. rubella</i>	768	DISAPQSEATYKSLLEILYSLMAKEGYIPDP	REV	SQNL	EEEEKRRILICGHSERLAI
<i>E. salsugineum</i>	766	DDSAPQSEATYKSLLEILYSLMAKEGYIPDP	REV	SQNL	EEEEKRRILICGHSERLAI
<i>T. cacao</i>	773	DKFFEQFDEVTKLIGITSLMAKEGYVADLR	VLHD	EEDE	KRDLLICGHSERLAI
<i>F. vesca</i>					
<i>V. vinifera</i>	773	TKFFEQFDEVTKLIGITSLMAKEGYVADLR	VLHD	EEDE	KRDLLICGHSERLAI
<i>C. sativum</i>	779	DKSFEQYDKVDLLEILYSLMAKEGYVADLR	VLHD	EEDE	KRDMLICGHSERLAI
<i>S. lycopersicum</i>	775	DRSFOQFQVNELIEDLNRITMDKEGYVADLR	VLHD	EEDE	KINLLYGHSERLAI

MEF26	825	LNTEPGTLPQVMKNLRVCGDCHETVKLISKIV	GREILVRDANRFH	FKDGTCSCKDRW	882
<i>A. lyrata</i>	825	LNTEPGTLPQVMKNLRVCGDCHETVKLISKIV	GREILVRDANRFH	FKDGTCSCKDRW	882
<i>C. rubella</i>	827	LNTEPGTLPQVMKNLRVCGDCHETVKLISKIV	GREILVRDANRFH	FKDGTCSCKDRW	884
<i>E. salsugineum</i>	826	LNTEPGTLPQVMKNLRVCGDCHETVKLISKIV	GREILVRDANRFH	FKDGTCSCKDRW	883
<i>T. cacao</i>	832	LNTKPGTLPQVMKNLRVCGDCHETVKLISKIV	GREILVRDANRFH	FKDGTCSCKDRW	889
<i>F. vesca</i>					764
<i>V. vinifera</i>	832	LNTKPGTLPQVMKNLRVCGDCHETVKLISKIV	GREILVRDANRFH	FKDGTCSCKDRW	889
<i>C. sativum</i>	838	LNTKPGTLPQVMKNLRVCGDCHETVKLISKIV	GREILVRDANRFH	FKDGTCSCKDRW	895
<i>S. lycopersicum</i>	834	LNTKEGSPQVMKNLRVCGDCHETVKLISKIV	GREILVRDANRFH	FKDGTCSCKDRW	891

Supplementary Fig. 3. Comparison between amino acid sequences of MEF26 and MEF26 putative orthologs. The entire MEF26 amino acid sequence was used for a BLAST search at <http://blast.ncbi.nlm.nih.gov/Blast.cgi>, detected sequences were used for sequence alignment with CLUSTALW (<http://www.ebi.ac.uk/clustalw>), and residues identical (white letters on black background) and similar (white letters on grey background) to MEF26 amino acids were identified by BOXSHADE (http://www.ch.embnet.org/software/BOX_form.html).

MEF26 PPR motifs	L	S	P	L	S	P	L	S	S	L	S	P	L	S	P	L	S	P	L2	S2
Position	6	1'	6	1'	6	1'	6	1'	6	1'	6	1'	6	1'	6	1'	6	1'	6	1'
<i>A. thaliana</i> (MEF26)	P	S	G	N	N	D	P	D	N	D	N	S	V	N	D	N	S	E	N	D
<i>A. lyrata</i>	A	S	G	N	N	D	P	D	N	D	N	S	V	N	D	N	T	E	N	D
<i>V. vinifera</i>	S	S	A	N	N	D	P	D	N	D	N	S	V	N	D	N	D	T	N	D
<i>C. sativus</i>	S	S	G	N	N	D	P	D	N	D	N	S	D	N	D	N	D	T	N	D
<i>S. lycopersicum</i>	S	S	G	N	N	D	P	D	N	D	N	S	D	N	D	N	D	T	N	D
<i>S. tuberosum</i>	S	S	G	N	N	D	P	D	N	D	N	S	D	N	D	N	D	T	N	D
<i>F. vesca</i>	S	S	G	S	N	D	P	D	N	D	N	S	D	N	D	N	D	T	N	D
<i>O. sativa</i> (Os01g10800)	S	Q	N	D	N	N	L	D	S	N	N	S	Q	N	D	N	L	E	N	N
<i>A. thaliana</i> (AT4G18750)					N	D	C	D	S	K	N	S	R	N	D	N	D	V	E	N
<i>A. thaliana</i> (AT4G13650)	S	S	N	D	V	T	S	D	N	D	N	A	N	G	N	N	P	N	S	D
<i>G. max</i> (XP_006574752.1)	A	S	N	D	V	T	S	E	N	D	N	A	D	G	N	N	P	N	S	D

Figure 5. The amino acids at positions 6 and 1' in PPR motifs of putative MEF26 orthologs are highly conserved. The amino acids at positions 6 and 1' in PPR motifs of putative MEF26 orthologs from *Arabidopsis lyrata*, *Vitis vinifera*, *Cucumis sativus*, *Solanum lycopersicum*, *Solanum tuberosum* and *Fragaria vesca* are aligned. Most of these amino acids are identical in all potential orthologs suggesting that these MEF26-like proteins also perform the function as RNA editing factor for *cox3-311* and *nad4-166*. Low similarity at the 1' position of L and L2 domains are in line with the PPR code for these domains (Takenaka et al., 2013a), in which the amino acid at the 1' position is not involved in nucleotide recognition. *Oryza sativa* and *glycine max* do not have an obvious MEF26 ortholog. The PPR proteins most similar to MEF26 are the putative orthologs of other *Arabidopsis* PPR proteins (AT4G18750 and AT4G13650) and amino acids at positions 6 and 1' are different from the MEF26 orthologs suggesting that the two editing sites *cox3-311* and *nad4-166* are recognized by other PPR protein(s) in rice and soybean. The rice and soybean and AT4G13650 proteins have four additional domains to the N-terminus (not included) and AT4G18750 lacks two N-terminal PPR motifs when compared to MEF26.

4. Discussion

4.1. MEF26 is involved in editing two sites in different mitochondrial transcripts

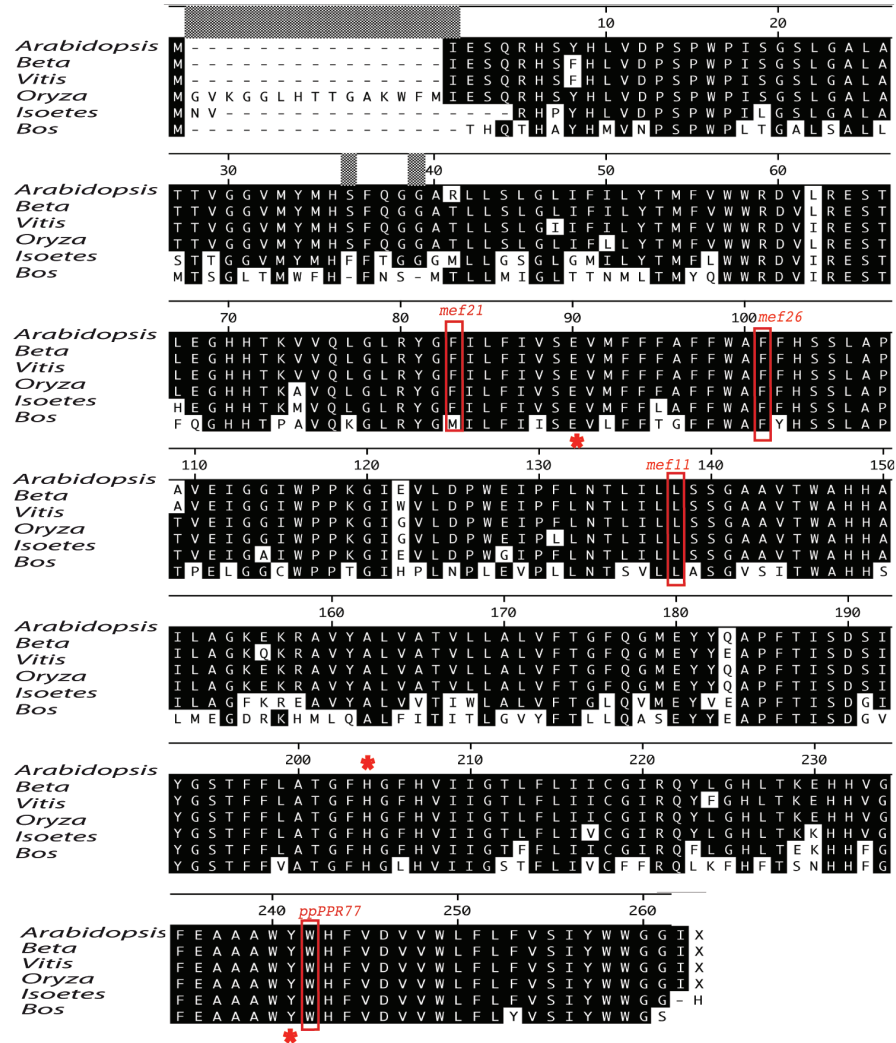
Here we identify MEF26 as a mitochondrial DYW-PPR protein required for RNA editing specifically at two sites in different transcripts. Editing at site *cox3-311* is completely abolished in one mutant line (*mef26-2*) and reduced to 20% in the other (*mef26-1*). Analogous residual partial RNA editing has been observed in other RNA editing mutant lines and may be explained by a mild mutant allele of the editing factor, by its lower expression, or by redundant targeting of this site by another, albeit less effective PPR editing factor. Since *cox3-311* editing is abolished in the *mef26-2* mutant, a redundant factor targeting *cox3-311* is unlikely. The insertion of the Ds element in *mef26-1* is only 18 nucleotides downstream of the ATG initiation codon so that most of the coding region remains intact. Thus, residual RNA editing observed in *mef26-1* might be due to a low expression of a truncated MEF26 translated from the next downstream in-frame ATG or to coincidentally translated products from an in-frame ATG in the Ds element.

In contrast to the complete loss of editing at *cox3-311*, at least in the *mef26-2* mutant, editing at *nad4-166* is only partially affected in both mutant lines. One possible explanation is that the *nad4-166* site is able to recruit another specific factor in addition to MEF26. This is not unprecedented, since two different PPR proteins are involved in editing at the same sites in several plastid and mitochondrial transcripts (Robbins et al., 2009; Tseng et al., 2010; Verbitskiy et al., 2012). An alternative hypothesis of an indirect effect of MEF26 on the *nad4-166* editing, for instance as a result of defects in other RNA processing steps like splicing is unlikely, since we did not observe differences in the *nad4* transcript profiles between wild type and *mef26-2* mutant plants (Supplementary Figure 2). Furthermore, the predicted direct interaction between MEF26 and the *nad4-166* site derived with the recently proposed amino acid code for RNA recognition by PPR proteins also supports the direct involvement of the MEF26 protein in the editing at this site (Figure 4; see below).

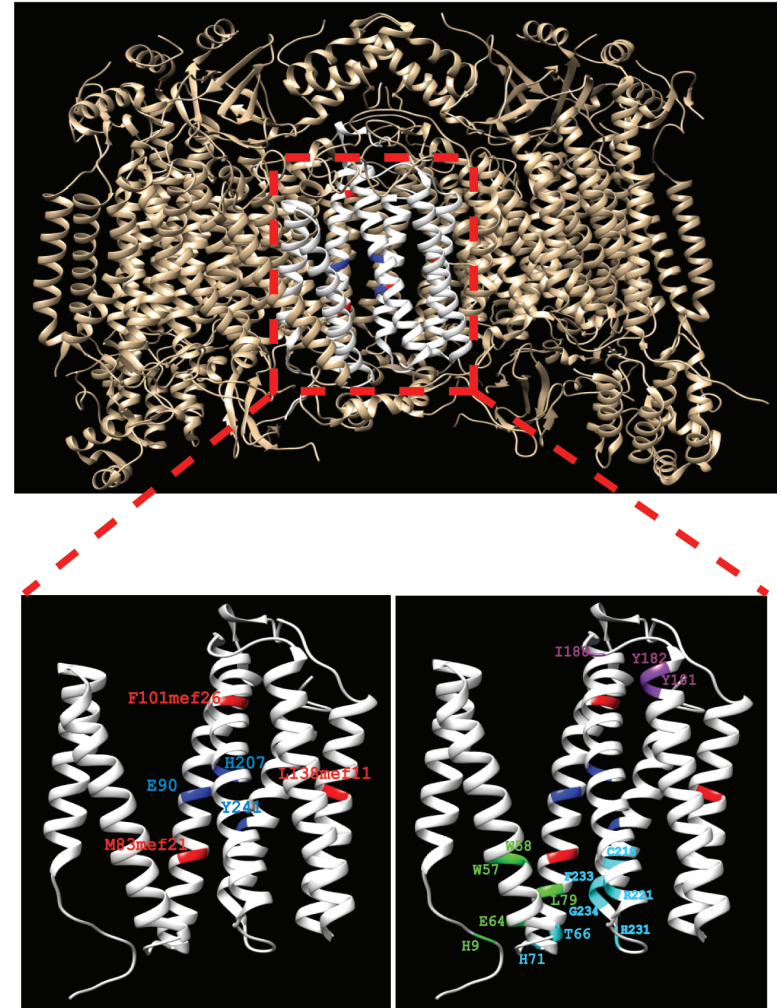
The unedited *cox3*-311 and *nad4*-166 cytidines in *mef26* result in Ser and Arg residues instead of highly conserved Phe104 and Trp56 moieties in COX3 and NAD4 polypeptides, respectively (Figure 3). Since editing at *cox3*-311 is abolished in *mef26-2* mutant plants, respiratory complex IV (cytochrome c oxydase) in these plants would contain COX3 subunits with a serine at amino acid position 104. COX3 is one of the three major subunits forming the functional core of complex IV in all organisms, contains bound phospholipids and contributes to complex stability, to control O₂ transfer to the active site and to proton transport (Gilderson et al., 2003; Shinzawa-Itoh et al., 2007). Despite the conservation of the Phe encoded by the edited codon in plants and other organisms (Figure 3), alteration to serine caused by the absence of editing at site *cox3*-311 is apparently tolerable at least under standard growth conditions.

Two additional PPR proteins have been identified to be involved in *cox3* editing in *Arabidopsis*. In *mef21* mutant plants a complete loss of editing was found at site *cox3*-257 (Takenaka et al., 2010). An UCU (Ser) codon is changed to UUU encoding a conserved phenylalanine by the RNA editing. However, like *mef26* mutant plants, *mef21* mutant plants display no gross disturbance in their growth or developmental pattern. Homozygous *mef11* mutant plants show a mild phenotype and a defect in editing at site *cox3*-422, which changes a Pro codon to a Leu codon (Verbitskiy et al., 2010). However this mutant has additional RNA editing defects leading to changes of codon identities in other transcripts. Therefore, phenotypic alterations cannot be unambiguously attributed to the change in COX3. The three amino acids in *Bos taurus* COX3 corresponding to *Arabidopsis* Phe104, Phe86 and Leu141, altered in *mef26*, *mef21* and *mef11* mutants respectively, are not known to be involved in either phospholipid binding, proton transport or the O₂ transfer channel (Supplementary Figure 4). It will be interesting to determine whether these highly conserved RNA editing sites affect structure or activity of the COX3 polypeptide and complex IV. Phenotypical analysis of these editing mutants under different conditions and additional biochemical and functional studies of the unedited form of COX3 protein may reveal subtle changes in structure or activity of complex IV.

A



B



Supplementary Fig. 4. Amino acid sequence alignment of COX3 proteins from several plant species and *Bos taurus*, and tridimensional representation of crystallographic structure of *Bos taurus* Complex IV. (A) Alignment of COX3 amino acid sequences from *A. thaliana*, *Beta vulgaris*, *Vitis vinifera*, *Oryza sativa*, *Isoetes engelmannii* and *Bos Taurus*. Numbering corresponds to the *B. taurus* sequence. Amino acids boxed in red represent modified sites in *Arabidopsis* PPR mutants (Ser replaces Phe in *mef21* and *mef26* mutants, Pro replaces Leu in the *mef11* mutant; Verbitskiy et al., 2010; Takenaka et al., 2010). Red stars indicate putative key residues in the O₂ transfer channel in *B. taurus* Complex IV. Residues involved in phospholipid binding are H9, W57, W58, E64, T66, H71, L79, Y181, Y182, I188, C218, R221, H231, F233 and G234 (Shinzawa-Itoh et al., 2007). Amino acids located in the N-terminus and known to be involved in proton transport regulation are not shown (Gilderson et al., 2003). (B) Dimeric structure of *Bos taurus* Complex IV, modified from PDB: 2DYR in the Protein Data Bank (<http://www.rcbs.org>; Shinzawa-Itoh et al., 2007). In the center, COX3 subunit is highlighted in white and a zoom (left panel) shows the characteristic seven α -helixes with the amino acids altered by defects in RNA editing in red color and the amino acids involved in O₂ channeling (E90, H207 and Y241) in blue color. Amino acids involved in phospholipid binding are shown (right panel) in green, light blue and purple colors.

Plants with significant defects in complex I (NADH dehydrogenase) show morphological phenotypes (e. g. De Longevialle et al., 2007; Dutilleul et al., 2003; Gutierrez et al., 1997; Meyer et al., 2009; Yamato and Newton, 1999). Since editing at *nad4*-166 is only partially affected in *mef26* mutant plants, one possibility is that sufficient wild type NAD4 polypeptide, with Trp at position 56, is translated from edited mRNA. An alternative explanation is that, like with COX3, the altered NAD4 bearing an Arg56 is translated from unedited mRNA, and assembled and tolerated in complex I. NAD4 is one of more than 40 subunits of respiratory complex I and is localized in the membrane arm. It is one of the anti-porter like subunits forming one out of the 4 proton-translocation channels (Baradaran et al., 2013). Four additional PPR proteins are known to be required for *nad4* editing in Arabidopsis. In *mef18* mutant plants loss of editing at *nad4*-1355 is complete, however the presence of Ser instead of Leu at position 452 in the NAD4 polypeptide does not result in phenotypic changes under standard growth conditions (Takenaka et al., 2010). *slo1*, *mef11* and *ahg11* mutants are also defective in *nad4* mRNA editing and show either a severe (*slo1*) or a mild phenotype (*mef11* and *ahg11*) (Sung et al., 2010; Verbitskiy et al. 2010). However, altered phenotype in *mef11* plants cannot be due to the editing defect at *nad4*-124 since this editing event is silent, i.e. does not change codon identity. *slo1* mutant has an additional RNA editing defect in *nad9* mRNA, so severe growth retardation may be the result of alterations in either NAD4, NAD9 or both. Finally, in *ahg11* mutant plants a single RNA editing defect was found among more than 400 analyzed sites, at *nad4*-376. Interestingly, the amino acid altered in the *slo1* mutant (Leu150 replaced by Pro) is highly conserved (Leu126 in *Thermus thermophilus*, Supplementary Figure 5) and close to a glutamic acid known to be involved in proton translocation (Glu147 in Arabidopsis, Glu123 in *T. thermophilus*; Supplementary Figure 5, Baradaran et al., 2013). Moreover, this Glu is part of a central axis of charged residues in the membrane arm which is involved in the mechanism of coupling between electron transfer and proton translocation. In the *ahg11* mutant an arginine replaced cysteine126 at a position not conserved in *T. thermophilus* (Met102), nevertheless this

amino acid is located near the proposed proton translocation pathway (Supplementary Figure 5). In contrast, amino acids changed by the lack of editing in *mef18* and *mef26* mutant plants do not appear to be involved in either proton translocation or coupling between electron transport and proton translocation (Supplementary Figure 5, Baradaran et al., 2013)

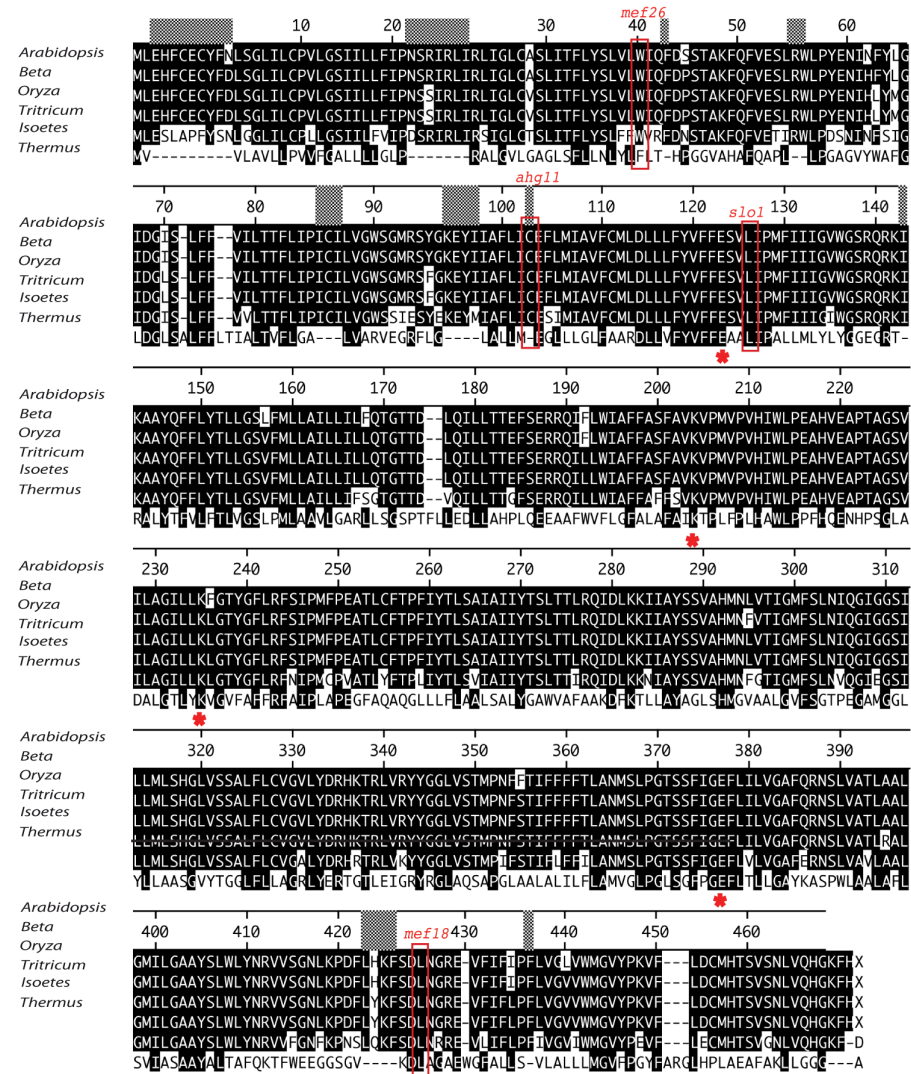
Thus, to date no correlation can be firmly established between editing defects at *cox3* or *nad4* and phenotypic alterations during plant development, at least under standard growth conditions. Alterations may become apparent under different conditions, and additional biochemical and functional studies may reveal subtle changes in complex structure or activity.

4.2. Edited sites targeted by MEF26 share similar cis-elements

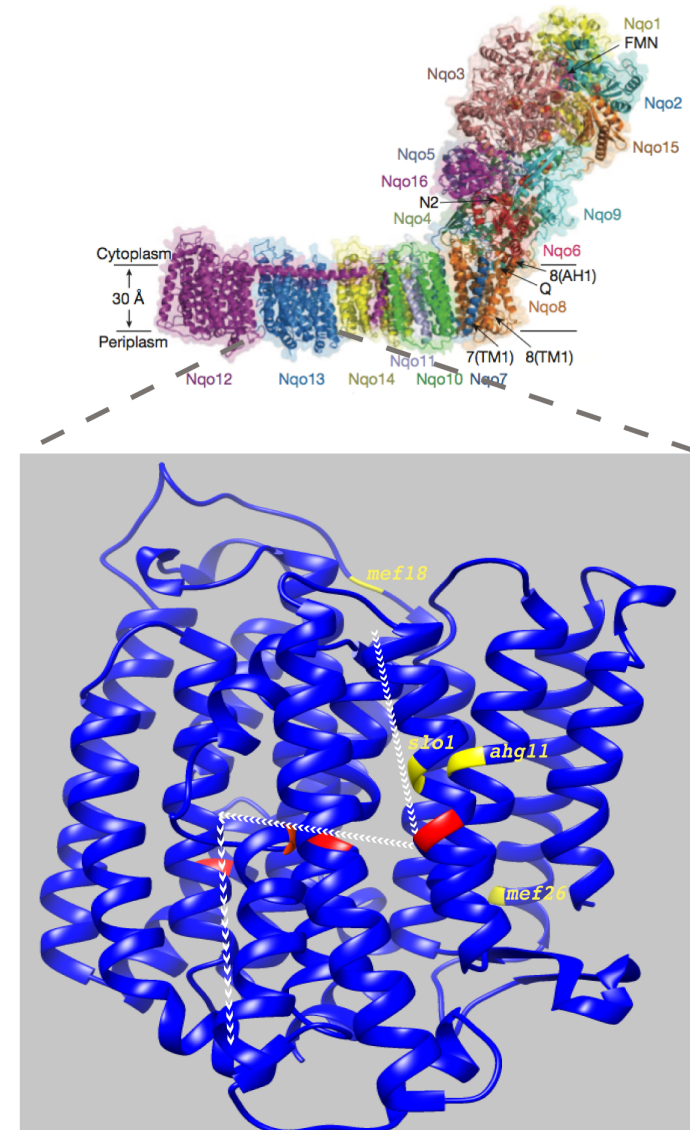
In vivo, *in vitro* and *in organello* investigations have shown that the *cis*-elements in the RNA for a number of editing sites reside in stretches of 20 to 25 nucleotides mostly upstream (5') of the editing site (Farré et al., 2001; Miyamoto et al., 2002; Takenaka et al., 2004). Within the window between -25 and -1 relative to the edited C, sixteen nucleotides are identical between the two editing sites targeted by MEF26 (Figure 4A). Interestingly, this similarity increases to eighteen shared nucleotides after editing at *nad4*-158 and *nad4*-164 sites, and thus these upstream editing sites in the *nad4* transcript may precede site recognition by MEF26. According to the recent PPR code predictions (Barkan et al., 2012; Takenaka et al., 2013a; Yagi et al., 2013), RNA editing PPR proteins recognize the sequence further upstream from nucleotide -4 via a modular one repeat:one nucleotide mechanism for about as many nucleotides as there are PPR motifs, thus up to 20 nucleotides for the 20 repeats in MEF26 which would cover the upstream *nad4*-158 site (Figure 4B). Editing at this site increases binding probability to MEF26 (Figure 4B, upper panel) and may thus be required for recognition of the *nad4*-166 editing site. On the other hand, current models do not assign any role to the -2 position of *nad4*-164 in *nad4*-166 editing.

Similarities in the upstream 20-25 nucleotides of editing sites are also observed between the targets of other mitochondrial PPR *trans*-factors, for instance MEF11, SLO1 and OTP87

A



B



Supplementary Fig. 5. NAD4 amino acid altered by editing defects in Arabidopsis *ppr* mutants and their localization in the *Thermus thermophilus* Complex I structure. (A) Alignment of NAD4 amino acid from *A. thaliana*, *B. vulgaris*, *O. sativa*, *Triticum aestivum*, *I. engelmannii*, and *T. thermophilus*. Numbering corresponds to the *T. thermophilus* sequence. Amino acid altered by defects in RNA editing in the *ppr* mutants are boxed in red (Arg replace Trp and Cys in *mef26* and *ahg11* mutants, respectively, and Pro replace Leu in *slo1* mutant). Amino acid known to be involved in proton traslocation in the bacterial complex are indicated by red stars. (B) Diagrame of *Thermus thermophilus* Complex I, PDB 4HE8 (<http://www.rcsb.org>) (Baradaran et al., 2013), with its characteristic L-shape. Nqo13 in the membrane arm is the Arabidopsis NAD4 ortholog. A tridimensional model of NAD4 subunit from Arabidopsis were built with Phyre2 software, a suite of tools available on the web to predict and analyze protein structure (<http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index>) (Kelley et al., 2015). Using the crystal structure of the membrane domain of respiratory complex I from *E. coli* at 3.0 angstrom of resolution. Key residues involved in proton translocation are labeled red. The white arrows show the movement to the proton translocating.

which are involved in editing at six, two and two sites respectively in *Arabidopsis* (Sung et al., 2010; Verbitskiy et al., 2010; Hammani et al., 2011). On the other hand, other PPR proteins such as MEF1, REME1, OGR1 and EMP5 show relatively low sequence similarity between their three, two, seven and five target sites, respectively (Kim et al., 2009; Zehrmann et al., 2009; Bentolila et al., 2010; Liu et al., 2013). These differences may be derived from the characteristics of the amino acid code for RNA recognition by PPR proteins. Two (or three) amino acids in each PPR motif constitute the code, identities of amino acid 6 in one repeat and amino acid 1 in the next repeat (called 1') being the primary determinants of the bound nucleotide. However different amino acid combinations show variable nucleotide specificities: for example the combination N, D at 6,1' positions strongly prefers U (high specificity), but N, N at 6,1' can recognize C and U, and S, S at 6,1' shows no preference for any nucleotide (non-specific). Thus, high similarity in the *cis*-elements between target sites could reflect a higher content of PPR domains with high specificity. This is indeed seen in MEF26 which possesses a high number (14 out of 20) of PPR motifs with high preferences for one nucleotide identity (Figure 4B, upper panel). Interestingly, a high number of N, D at 6, 1' positions of P and S motifs correlates well with the high U content of sequences upstream of *cox3*-311 and *nad4*-166. Such U-rich *cis*-elements have been found in other editing sites, including the target site of the longest RNA editing PPR protein characterized in *Arabidopsis*, MEF12 (Härtel et al., 2013). However, the PPR amino acid code ranks its target site only at position 14 of 430 editing sites in *Arabidopsis* mitochondria, and U stretches were proposed to act as spacers between the nucleotides bound by only some PPR motifs. In contrast, these uridine stretches in the MEF26 targets strongly contribute to the predicted binding with MEF26, and thus may not act as mere spacers (Figure 4).

The editing function of MEF26 is highly specific since the editing levels of the nearby sites *cox3*-314 and *nad4*-166, separated by 2 nucleotides from the *bona fide* targeted editing sites, are unaffected in *mef26* mutant plants. This can be easily explained by the amino acid code for *cox3*-314 (low binding probability) but not for *nad4*-164, which has a binding probability similar

to that of *nad4*-166 (Figure 4B). This can be rationalized by the PPR-nucleotide “mismatches” at positions -6 and -12 in the *nad4*-166 upstream region. These mismatches located very close to the target editing site could be more critical than mismatch(es) in the further upstream sequence, samples of which are observed in the -16 of *cox3*-311 and *nad4*-166 sites (Figure 4B). However these two neighboring potential MEF26 binding sites may enhance the RNA editing activity at *nad4*-166 by trapping the PPR protein around the position as observed in mitochondrial *in vitro* RNA editing (Verbitskiy et al., 2008). An analogous situation was previously described for *Arabidopsis* MEF25, which is required for editing at *nad1*-308 but not *nad1*-307 (Arenas-M. et al., 2013) in spite of a best match to the putative *cis*-element of *nad1*-307 (Takenaka et al., 2013a). From these and additional data, e.g. low ranks in prediction of genuine editing site recognition by some cognate PPR proteins (Takenaka et al., 2013a), it is clear that the PPR recognition code and target site prediction need to be refined to fully explain why certain cytidines are targeted by a PPR protein but not others.

4.3. Conclusions

Most members of the large nuclear-encoded family of pentatricopeptide repeat (PPR) proteins are predicted to be targeted to mitochondria and plastids and a number of these play relevant roles in the maturation of organellar RNAs. Here we report the novel Mitochondrial Editing Factor 26 (MEF26), which consists of twenty PPR domains and C-terminal E and DYW domains. Analysis of plastid and mitochondrial RNA editing in two homozygous mutant lines of this gene revealed that two sites are affected in different mitochondrial transcripts. While at one site, *cox3*-311, editing is completely abolished in the absence of MEF26, the other site, *nad4*-166, is still partially edited. Full editing at both sites is recovered in the mutants by complementation with a wild type *MEF26* gene. The different editing efficiencies at these sites suggest the presence of an additional RNA editing factor which can distinguish between the highly similar *cis*-elements of *nad4*-166 and *cox3*-311 and recognize only *nad4*-166.

The application of the recently proposed amino acid code for RNA recognition by PPR proteins ranks *cox3*-311 and *nad4*-166 sites at first and second position of the most probable targets of MEF26, respectively. Analysis of putative MEF26 orthologs in several other plant species reveal that the two amino acids involved in recognition of the nucleotides in the *cis*-element (6th amino acid in one PPR motif and 1st amino acid in the next PPR motif) are conserved, suggesting their functional involvement in editing the same sites.

5. Acknowledgments

We are very grateful to Axel Brennicke for his critical reading of the manuscript, discussion of results and constant encouragement. This work was supported by grants from CONICYT (AT24100161), FONDECYT (1100601 and 1141197), Millenium Nucleus in Plant Functional Genomics (P10-062-F, Millennium Scientific Initiative Program, Mideplan, Chile) and the Deutsche Forschungsgemeinschaft. Mizuki Takenaka is a Heisenberg fellow.

6. References

- Arenas, M.A., Takenaka, M., Moreno, S., Gomez, I., Jordana, X., 2013. Contiguous RNA editing sites in the mitochondrial *nad1* transcript of *Arabidopsis thaliana* are recognized by different proteins. *FEBS Lett* 587, 887-891.
- Bannai, H., Tamada, Y., Maruyama, O., Nakai, K., Miyano, S., 2002. Extensive feature detection of N-terminal protein sorting signals. *Bioinformatics* 18, 298-305.
- Baradaran, R., Berrisford, J.M., Minhas, G.S., Sazanov, L.A., 2013. Crystal structure of the entire respiratory complex I. *Nature* 494, 443-448.
- Barkan, A., Rojas, M., Fujii, S., Yap, A., Chong, Y.S., Bond, C.S., Small, I., 2012. A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. *PLoS Genet* 8, e1002910.
- Barkan, A., Small, I., 2014. Pentatricopeptide Repeat Proteins in Plants. *Annu Rev Plant Biol.* 65, 415-442.
- Bentolila, S., Knight, W., Hanson, M., 2010. Natural variation in *Arabidopsis* leads to the identification of REME1, a pentatricopeptide repeat-DYW protein controlling the editing of mitochondrial transcripts. *Plant Physiol* 154, 1966-1982.
- Bentolila, S., Heller, W.P., Sun, T., Babina, A.M., Friso, G., van Wijk, K.J., 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, E1453-1461.
- Chateigner-Boutin, A.L., des Francs-Small, C.C., Delannoy, E., Kahlau, S., Tanz, S.K., de Longevialle, A.F., Fujii, S., Small, I., 2011. OTP70 is a pentatricopeptide repeat protein of the E subgroup involved in splicing of the plastid transcript *rpoC1*. *Plant J* 65, 532-542.
- Chaudhuri, S., Maliga, P., 1996. Sequences directing C to U editing of the plastid *psbL* mRNA are located within 22 nucleotide segment spanning the editing site. *EMBO J* 15, 5958-5964.
- Choury, D., Farré, J.C., Jordana, X., Araya, A., 2004. Different patterns in the recognition of editing sites in plant mitochondria. *Nucleic Acids Res* 32, 6397-6406.
- Claros, M.G., Vincens, P., 1996. Computational method to predict mitochondrially imported proteins and their targeting sequences. *Eur J Biochem* 241, 779-786.
- Clough, S.J., Bent, A.F., 1998. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J* 16, 735-743.
- Covello, P.S., Gray, M.W., 1989. RNA editing in plant mitochondria. *Nature* 341, 662-666.
- Crooks, G.E., Hon, G., Chandonia, J.M., Brenner, S.E., 2004. WebLogo: a sequence logo generator. *Genome Res* 14, 1188-1190.
- De Longevialle, A.F., Meyer, E.H., Andres, C., Taylor, N.L., Lurin, C., Millar, A.H., Small, I.D., 2007. The pentatricopeptide repeat gene OTP43 is required for trans-splicing of the mitochondrial *nad1* Intron 1 in *Arabidopsis thaliana*. *Plant Cell* 19, 3256-3265.

- Dutilleul, C., Driscoll, S., Cornic, G., De Paepe, R., Foyer, C.H., Noctor, G., 2003. Functional mitochondrial complex I is required by tobacco leaves for optimal photosynthetic performance in photorespiratory conditions and during transients. *Plant Physiol* 131, 264-275.
- Farré, J.C., León, G., Jordana, X., Araya, A., 2001. Cis recognition elements in plant mitochondrion RNA editing. *Mol Cell Biol* 21, 6731-37.
- García-Andrade, J., Ramírez, V., López, A., Vera, P., 2013. Mediated plastid RNA editing in plant immunity. *PLoS Pathog* 9, e1003713.
- Giegé, P., Brennicke, A., 1999. RNA editing in Arabidopsis mitochondria effects 441 C to U changes in ORFs. *Proc Natl Acad Sci USA* 96, 15324-15329.
- Gilderson, G., Salomonsson, L., Aagaard, A., Gray, J., Brzezinski, P., Hosler, J., 2003. Subunit III of cytochrome c oxidase of *Rhodobacter sphaeroides* is required to maintain rapid proton uptake through the D pathway at physiologic pH. *Biochemistry* 42, 7400-7409.
- Grewe, F., Herres, S., Viehover, P., Polsakiewicz, M., Weisshaar, B., Knoop, V., 2012. A unique transcriptome: 1782 positions of RNA editing alter 1406 codon identities in mitochondrial mRNAs of the lycophyte *Isoetes engelmannii*. *Nucleic Acids Res* 39, 2890-2902.
- Gualberto, J.M., Lamattina, L., Bonnard, G., Weil, J.H., Grienemberger, J.M., 1989. RNA editing in wheat mitochondria results in the conservation of protein sequences. *Nature* 341, 660-662.
- Gutierrez, S., Sabar, M., Lelandais, C., Chetrit, P., Diolez, P., Degand, H., Boutry, M., Vedel, F., de Kouchkovsky, Y., de Paepe, R., 1997. Lack of mitochondrial and nuclear-encoded subunits of complex I and alteration of the respiratory chain in *Nicotiana sylvestris* mitochondrial deletion mutants. *Proc Natl Acad Sci USA* 94, 3436-3441.
- Hammani, K., des Francs-Small, C.C., Takenaka, M., Tanz, S.K., Okuda, K., Shikanai, T., Brennicke, A., Small, I., 2011. The pentatricopeptide repeat protein OTP87 is essential for RNA editing of *nad7* and *atp1* transcripts in Arabidopsis mitochondria. *J Biol Chem* 286, 21361-21371.
- Hammani, K., Okuda, K., Tanz, S.K., Chateigner-Boutin, A.L., Shikanai, T., Small, I., 2009. A study of new Arabidopsis chloroplast RNA editing mutants reveals general features of editing factors and their target sites. *Plant Cell* 21, 3686-3699.
- Härtel, B., Zehrmann, A., Verbitskiy, D., Takenaka, M., 2013. The longest mitochondrial RNA editing PPR protein MEF12 in Arabidopsis thaliana requires the full-length E domain. *RNA Biol* 10, 1543-1548.
- Hecht, J., Grewe, F., Knoop, V., 2011. Extreme RNA editing in coding islands and abundant microsatellites in repeat sequences of *Selaginella moellendorffii* mitochondria: the root of frequent plant mtDNA recombination in early tracheophytes. *Genome Biol Evol* 3, 344-358.
- Iwabuchi, M., Kyojuka, J., Shimamoto, K., 1993. processing followed by complete editing of an altered mitochondrial *atp6* RNA restores fertility of cytoplasmic male sterile rice. *EMBO J*, 12, 1437-1446.
- Karimi, M., Depicker, A., Hilson, P., 2007. Recombinational cloning with plant gateway vectors. *Plant Physiol* 145, 1144-1154.

Kim, S.R., Yang, J.I., Moon, S., Ryu, C.H., An, K., Kim, K.M., Yim, J., An, G., 2009. Rice OGR1 encodes a pentatricopeptide repeat-DYW protein and is essential for RNA editing in mitochondria. *Plant J.* 59, 738-749.

León, G., Holuigue, L., Jordana, X., 2007. Mitochondrial complex II Is essential for gametophyte development in Arabidopsis. *Plant Physiol* 143, 1534-1546.

Liu, Y.J., Xiu, Z.H., Meeley, R., Tan, B.C., 2013. Empty pericarp5 encodes a pentatricopeptide repeat protein that is required for mitochondrial RNA editing and seed development in maize. *Plant Cell* 25, 868-883.

Lurin, C., Andres, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyere, C., Caboche, M., Debast, C., Gualberto, J., Hoffmann, B., Lecharny, A., Le Ret, M., Martin-Magniette, M.L., Mireau, H., Peeters, N., Renou, J.P., Szurek, B., Taconnat, L., Small, I., 2004. Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell* 16, 2089-2103.

Meyer, E.H., Tomaz, T., Carroll, A.J., Estavillo, G., Delannoy, E., Tanz, S.K., Small, I.D., Pogson, B.J., Millar, A.H., 2009. Remodeled respiration in *ndufs4* with low phosphorylation efficiency suppresses Arabidopsis germination and growth and alters control of metabolism at night. *Plant Physiol* 151, 603-619.

Miyamoto, T., Obokata, J., Sugiura, M., 2002. Recognition of RNA editing sites is directed by unique proteins in chloroplasts: biochemical identification of *cis*-acting elements and *trans*-acting factors involved in RNA editing in tobacco and pea chloroplasts. *Mol Cell Biol* 22, 6726-6734.

Murayama M., Hayashi S., Nishimura N., Ishide M., Kobayashi K., Yagi Y., Asami T., Nakamura T., Shinozaki K. and Hirayama T. 2012. Isolation of Arabidopsis *ahg11*, a weak ABA hypersensitive mutant defective in *nad4* RNA editing. *J. Exp. Bot.* 63, 5301-5310.

Neuwirt, J., Takenaka, M., van der Merwe, J.A., Brennicke, A., 2005. An in vitro RNA editing system from cauliflower mitochondria: editing site recognition parameters can vary in different plant species. *RNA* 11, 1563-1570.

Notsu, Y., Masood, S., Nishikawa, T., Kubo, N., Akiduki, G., Nakazono, M., Hirai, A., Kadowaki, K., 2002. The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants. *Mol Genet Genomics* 268, 434-445.

Okuda, K., Shikanai, T., 2012. A pentatricopeptide repeat protein acts as a site-specificity factor at multiple RNA editing sites with unrelated *cis*-acting elements in plastids. *Nucleic Acids Res* 40, 5052-5064.

Robbins, J.C., Heller, W.P., Hanson, M.R., 2009. A comparative genomics approach identifies a PPR-DYW protein that is essential for C-to-U editing of the Arabidopsis chloroplast *accD* transcript. *RNA* 15, 1142-1153.

Schmitz-Linneweber, C., Small, I., 2008. Pentatricopeptide repeat proteins: a socket set for organelle gene expression. *Trends Plant Sci* 13, 663-670.

Shinzawa-Itoh, K., Aoyama, H., Muramoto, K., Terada, H., Kurauchi, T., Tadehara, Y., Yamasaki, A., Sugimura, T., Kurono, S., Tsujimoto, K., Mizushima, T., Yamashita, E., Tsukihara, T., Yoshikawa, S., 2007. Structures and physiological roles of 13 integral lipids of bovine heart cytochrome c oxidase. *EMBO J* 26, 1713-1725.

Sung, T.Y., Tseng, C.C., Hsieh, M.H., 2010. The SLO1 PPR protein is required for RNA editing at multiple sites with similar upstream sequences in *Arabidopsis* mitochondria. *Plant J* 63, 499-511.

Takenaka, M., Brennicke, A., 2009. Multiplex single-base extension typing to identify nuclear genes required for RNA editing in plant organelles. *Nucleic Acids Res* 37, e13.

Takenaka, M., Neuwirt, J., Brennicke, A., 2004. Complex cis-elements determine an RNA editing site in pea mitochondria. *Nucleic Acids Res* 32, 4137-4144.

Takenaka, M., Verbitskiy, D., van der Merwe, J.A., Zehrmann, A., Brennicke, A., 2008. The process of RNA editing in plant mitochondria. *Mitochondrion* 8, 35-46.

Takenaka, M., Verbitskiy, D., Zehrmann, A., Brennicke, A., 2010. Reverse genetic screening identifies five E-class PPR proteins involved in RNA editing in mitochondria of *Arabidopsis thaliana*. *J Biol Chem* 285, 27122-27129.

Takenaka, M., Zehrmann, A., Brennicke, A., Graichen, K., 2013a. Improved computational target site prediction for pentatricopeptide repeat RNA editing factors. *PLoS One* 8, e65343.

Takenaka, M., Zehrmann, A., Verbitskiy, D., Hartel, B., Brennicke, A., 2013b. RNA editing in plants and its evolution. *Annu Rev Genet* 47, 335-352.

Takenaka, M., Zehrmann, A., Verbitskiy, D., Kugelmann, M., Hartel, B., 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.

Tasaki, E., Hattori, M., Sugita, M., 2010. The moss pentatricopeptide repeat protein with a DYW domain is responsible for RNA editing of mitochondrial *ccmFc* transcript. *Plant J* 62, 560-570.

Tseng, C.C., Sung, T.Y., Li, Y.C., Hsu, S.J., Lin, C.L., Hsieh, M.H., 2010. Editing of *accD* and *ndhF* chloroplast transcripts is partially affected in the *Arabidopsis vanilla cream1* mutant. *Plant Mol. Biol.* 73, 309-323.

Verbitskiy, D., van der Merwe, J.A., Zehrmann, A., Brennicke, A., Takenaka, M., 2008. Multiple specificity recognition motifs enhance plant mitochondrial RNA editing *in vitro*. *J Biol Chem* 283, 24374-24381.

Verbitskiy, D., Zehrmann, A., van der Merwe, J.A., Brennicke, A., Takenaka, M., 2010. The PPR protein encoded by the LOVASTATIN INSENSITIVE 1 gene is involved in RNA editing at three sites in mitochondria of *Arabidopsis thaliana*. *Plant J* 61, 446-455.

- Verbitskiy, D., Zehrmann, A., Hartel, B., Brennicke, A., Takenaka, M., 2012. Two related RNA-editing proteins target the same sites in mitochondria of *Arabidopsis thaliana*. *J Biol Chem* 287, 38064-38072.
- Wu, F.H., Shen, S.C., Lee, L.Y., Lee, S.H., Chan, M.T., Lin, C.S., 2009. Tape-Arabidopsis Sandwich - a simpler Arabidopsis protoplast isolation method. *Plant Methods* 5, 16.
- Yagi, Y., Hayashi, S., Kobayashi, K., Hirayama, T., Nakamura, T., 2013. Elucidation of the RNA recognition code for pentatricopeptide repeat proteins involved in organelle RNA editing in plants. *PLoS One* 8, e57286.
- Yagi, Y., Hayashi, S., Kobayashi, K., Hirayama, T., Nakamura, T., 2013. Elucidation of the RNA recognition code for pentatricopeptide repeat proteins involved in organelle RNA editing in plants. *PLoS One* 8, e57286.
- Yamato, K.T. and Newton, K.J., 1999. Heteroplasmy and homoplasmy for maize mitochondrial mutants: isolation of a rare homoplasmic nad4 deletion plant. *J. Hered.* 90, 369-373.
- Yin, P., Li, Q., Yan, C., Liu, Y., Liu, J., Yu, F., Wang, Z., Long, J., He, J., Wang, H.W., Wang, J., Zhu, J.K., Shi, Y., Yan, N., 2013. Structural basis for the modular recognition of single-stranded RNA by PPR proteins. *Nature* 504, 168-171.
- Zehrmann, A., Verbitskiy, D., Hartel, B., Brennicke, A., Takenaka, M., 2011. RNA editing competence of trans-factor MEF1 is modulated by ecotype-specific differences but requires the DYW domain. *FEBS Lett* 584, 4181-4186.
- Zehrmann, A., Verbitskiy, D., van der Merwe, J.A., Brennicke, A., Takenaka, M., 2009. A DYW Domain-Containing Pentatricopeptide Repeat Protein Is Required for RNA Editing at Multiple Sites in Mitochondria of *Arabidopsis thaliana*. *Plant Cell* 21, 558-567.

Supplementary Table 1: List of primers used in this work

Primer name	Used for	Orientation	Primer (5'-3')
RB-GK	<i>mef26-2</i> genotyping	-	GTGGATTGATGTGATATCTCC
DsSo2	<i>mef26-1</i> genotyping	-	AAATCAGATCTACGATAACGGTCGG
<i>mef26-2F</i>	<i>mef26-2</i> genotyping	Forward	TTGGGAATTTGCTGGTTGTC
<i>mef26-2R</i>	<i>mef26-2</i> genotyping	Reverse	GGCTTCCATAAGATCTCCGC
<i>mef26-1F</i>	<i>mef26-1</i> genotyping	Forward	CGTGGAAAAGAAATAAGTTTCTCGG
<i>mef26-1R</i>	<i>mef26-1</i> genotyping	Reverse	CTCGTCTTGCATCCGTTG
amino-mef26F	Localization and complementation	Forward	CACCATGCAGACTAGAGTGTCATCAC
amino-mef26R	Localization	Reverse	TTCAGGGAACAAACCATTCTTAGAAAA
carboxy-mef26R	Complementation	Reverse	CCACCGATCTTTGCAGCTAC
cox3-F1	RT-PCR to verify the editing defect	Forward	ACCAAAGTCGTACAATTAGGAC
cox3-R2	RT-PCR to verify the editing defect	Reverse	CGTCAGATGACCAAGATATTGC
nad4-F1	RT-PCR to verify the editing defect	Forward	GTGGTCTTATTCTGTGTCCTG
nad4-R2	RT-PCR to verify the editing defect	Reverse	CCTTGATCTTTCTTTGTCTCGAA

II. 4 Caracterización Funcional del Factor de Edición Mitocondrial 31 (MEF31)

En este capítulo se presenta una descripción detallada de la metodología y de los resultados obtenidos en la caracterización del gen At2g46050 que codifica para el factor de edición mitocondrial 31 (MEF31). Ensayos de localización subcelular muestran que MEF31 tendría localización mitocondrial. Paralelamente, en una línea mutante *mef31* se observa la nula edición de un sitio en el transcrito *orfX*, lo que genera un cambio de aminoácido en la putativa proteína mitocondrial tatC (twin arginine translocase C). Plantas mutantes *mef31* son indistinguibles de las plantas silvestres. Por otra parte, la utilización del “código de las PPR” predice que los residuos en las posiciones 6 y 1' de los motivos PPR de MEF31 serían capaces de unir a los nucleótidos del putativo elemento en *cis* de la citosina 581 en el transcrito *orfX*. En conjunto estos resultados sugieren que MEF31 es necesaria para la edición eficiente del sitio *orfX*-581 en mitocondrias de *A. thaliana*.

Metodología

1. - Crecimiento de *Arabidopsis thaliana*.

Se esterilizaban las semillas por el método del gas cloro (Sew y cols. 2013). Para ello se colocaban las semillas en tubos Eppendorf de 1,5 ml (máximo 0,1 ml de semillas por tubo) que eran puestos en un recipiente cerrado (deseCADOR) junto a dos vasos de precipitado que contenían 50 ml de NaClO 5% [v/v] y a los que se les agregaban 2 ml de HCl al 37% [v/v]. Después de esterilizar por 12 horas, rápidamente se cerraban los tubos. Una vez esterilizadas las semillas, se sembraban en una campana de flujo laminar, en placas Petri que contenían medio MS 0,5X (2,2 g/L de medio basal Murashige & Skoog con vitaminas, 1% [p/v] de sacarosa, 0,8% [p/v] de agar-agar y el antibiótico necesario: kanamicina 50mg/L, higromicina 25 mg/L o Basta (glufosinato de amonio) 7,25 mg/L). Se estratificaba por 48 horas en oscuridad a 4°C y posteriormente se incubaba en una cámara de crecimiento Percival® a una temperatura de 22°C, intensidad de luz entre 120-150 $\mu\text{E}^2\text{m}^{-2}\text{s}^{-1}$ y fotoperíodo de 16 horas de luz - 8 horas de oscuridad. Luego de tres semanas de crecimiento *in vitro*, se traspasaban las plántulas a macetas con mezcla de tierra: vermiculita 1:3 y se mantenían a temperaturas de 20-25°C con el mismo fotoperíodo. Las plantas eran regadas con agua una o dos veces a la semana.

2. - Extracción de ácidos nucleicos de *Arabidopsis thaliana*.

2.1- Extracción de DNA genómico.

Se cortaban trozos de tejido de aproximadamente 5 mm² y se les colocaba en un tubo Eppendorf con 200 μl de una solución TNE/SDS (Tris-HCl 200 mM pH 7,5; NaCl 250 mM, EDTA 25 mM y SDS 0,5% [p/v]). Se homogeneizaba el tejido con un DremelTM Minimate 763 equipado con un vástago plástico hasta tener una solución homogénea. El homogeneizado era incubado a 65°C por 5 minutos y se centrifugaba a 16000 g por 5 minutos a temperatura ambiente. Luego, se transfería el sobrenadante a un tubo con 200 μl de isopropanol, se agitaba y

se centrifugaba en las mismas condiciones anteriores. Posteriormente se descartaba el sobrenadante, se lavaba el precipitado con 200 µl de etanol 70%, y se centrifugaba nuevamente a 16000 g por 5 minutos a temperatura ambiente. Finalmente se descartaba el sobrenadante, se secaba el precipitado a temperatura ambiente y se le resuspendía en 35 µl de agua libre de nucleasas. Se guardaba el DNA a -20°C para su posterior uso.

2.2- Extracción de RNA total

Se preparaba RNA utilizando 50-100 mg de tejido que se congelaban en nitrógeno líquido en tubos Fastprep de 2 ml con esferas de óxido de zirconio. Luego se pulverizaba el tejido en un homogeneizador FastPrep®-24 a 4 m/sec por 20 segundos. El tejido pulverizado era resuspendido en 1 ml de TRIzol®, se agitaba en un vortex, se añadían 200 µl de cloroformo y se centrifugaba a 12500 g por 15 minutos a 4°C. Se traspasaba la fase acuosa a un tubo Eppendorf de 1,5 ml, se adicionaban 500 µl de isopropanol y se dejaba una hora a -20°C para precipitar el RNA. Se centrifugaban los tubos a 12500 g por 15 minutos a 4°C, se eliminaba el sobrenadante, se añadía etanol 70% [v/v] para lavar el precipitado y se centrifugaba a 7500 g por 10 minutos a 4°C. Se eliminaba el sobrenadante y se dejaba secar el precipitado a temperatura ambiente durante 15-30 minutos, para finalmente resuspenderlo en 35 µl de agua libre de nucleasas.

Para determinar la concentración de RNA se medía la razón de absorbancia a 260 y 280 nm en un espectrofotómetro NanoDrop 100 Spectrophotometer V3.7. Se guardaba el RNA a -20°C hasta su utilización.

3 - Técnicas de DNA recombinante

3.1- Reacción en cadena de la Polimerasa (PCR)

Se amplificaban los fragmentos a partir de DNA genómico (50 - 200 ng), DNA plasmidial (10 -50 ng) o de cDNA en un volumen final de 25 µl, utilizando la *Taq* DNA polimerasa. Se utilizaba el termociclador “Applied Biosystems® 2720 Thermal Cycler” con el

siguiente programa: 5-10 minutos a 94°C; 35-40 ciclos con las siguientes condiciones: 30 segundos a 94°C, 30 segundos entre 55-60°C y 35-90 segundos a 72°C; y finalmente 10 minutos a 72°C y se enfriaba a 4°C.

3.2- Síntesis de cDNA

Se realizaba la síntesis de cDNA por RT-PCR, utilizando entre 1 y 2 µg de RNA. Previamente, se trataba el RNA con 2U de DNasa RQ1 (Promega) por 20 minutos a 37°C en un volumen total de 10 µl, luego se adicionaba 1 µl de solución “Stop” (EGTA 20 mM, pH 8,0) e incubaba a 65°C por 10 minutos. Posteriormente, dependiendo del caso se adicionaba 1 µl de oligo-dT (500 ng/µl), 2 µl de “Random Hexamer” (125 ng/µl) o 1 µl de partidor específico (10 µM), se incubaba a 65°C por 5 minutos y luego se dejaba a temperatura ambiente por 5 minutos. La transcripción inversa era realizada con la enzima de M-MLV (Promega), adicionando a la mezcla anterior 4 µl de “M-MLV RT 5X buffer”, 2 µl de dNTPs (10 mM), 2 U del inhibidor de RNasa Ribolock (Fermentas), 200 U de transcriptasa inversa de M-MLV y H₂O hasta completar 25 µl. Finalmente, se incubaba la reacción por 90 minutos a 42°C (oligo-dT o partidor específico) o a 37°C (Random Hexamers), luego por 15 minutos a 65°C, se dejaba a 4°C para su utilización inmediata o se guardaba a -20°C.

3.3 - Electroforesis de DNA en geles de agarosa.

Se analizaban los fragmentos de DNA por electroforesis en geles de agarosa al 1-2% [p/v], preparados en una solución TAE 1X (Tris base 40mM, Ácido acético 20 mM, EDTA 1 mM, pH 8,0) y para su tinción se utilizaba bromuro de etidio a una concentración de 0,5 mg/L. Se mezclaban 5 µl de la solución con el DNA con 1 µl de amortiguador de carga azul – celeste 6X (glicerol 30% [v/v], azul de bromofenol 0,25% [p/v] y Xilen-Cianol FF 0,25% [p/v]), se cargaban los geles y se procedía a realizar la electroforesis por 20-30 minutos a 100 V (160 mA) en una solución TAE 1X. Finalmente, se visualizaban los fragmentos de DNA en un transiluminador “DigiDoc-It Imaging System” y se fotografiaba usando una cámara digital Canon 1130.

3.4 - Clonamiento de fragmentos de DNA en plásmidos.

Los fragmentos de DNA sintetizados por PCR eran clonados con los sistemas comerciales “pGEM-T[®] Easy Vector System” o “pENTRTM/SD/D-TOPO[®]” Cloning Kit”. La reacción de ligación en el vector pGEM-T era realizada en un volumen final de 10 µl utilizando 5 µl de “2X Rapid ligation buffer”, 1 µl del producto de PCR fresco (diluido 1/5 o 1/10), 50 ng del vector pGEM-T, 3U de T4 DNA ligasa. Se incubaba la reacción a 4°C durante toda la noche. Se transformaba *E.coli* y se seleccionaban las colonias positivas (métodos 3.5.1), a las cuales se les realizaba la extracción y purificación de DNA plasmidial descrita en Métodos 3.6.

Paralelamente, el clonamiento en el vector pENTRTM/SD/D-TOPO (Invitrogen) era realizado en un volumen total de 1,5 µl con 1 µl del DNA amplificado por PCR a ~ 1,25 ng/ µl, 0,25 µl de “Salt Solution” y 5 ng de vector pENTR. Se incubaba la reacción a 4°C toda la noche. Posteriormente se transformaba *E.coli* y se seleccionaban las colonias positivas (métodos 3.5.1), a las cuales se les realizaba la extracción y purificación de DNA plasmidial descrita en métodos 3.6. Luego, se realizaba la recombinación en el vector de destino (pK7FWG2.0) en un volumen final de 5 µl utilizando 25-75 ng del vector de entrada, 75 ng del vector de destino, 1 µl de 5X “LR clonase buffer”, 1 µl de amortiguador TE (Tris-HCl 10 mM pH 8,0 y EDTA 1mM) y 1 U de la enzima LR clonasa, y se incubaba la reacción de recombinación a temperatura ambiente durante la noche.

3.5 Transformación bacteriana.

3.5.1- Preparación y transformación de células químico-competentes

En la preparación de células competentes de *Escherichia coli* se inoculaban 100 ml de medio LB (peptona 10g/L, extracto de levadura 5g/L y NaCl 5g/L) con 5 ml de un cultivo de bacterias en fase estacionaria crecido durante toda la noche. Se incubaban las células a 37°C con agitación hasta alcanzar una D.O_{600 nm} de 0,6 aproximadamente, se centrifugaba a 3000 g por

15 minutos a 4°C en tubos de 50 ml y se descartaba el sobrenadante. El sedimento era resuspendido en una solución estéril fría de CaCl₂ 75 mM hasta alcanzar un décimo del volumen inicial (10 ml). Finalmente, se adicionaba glicerol hasta un 15% [v/v] y se alicuotaba de a 100 µl, se congelaban las alícuotas en nitrógeno líquido y se las guardaba a -80°C hasta su utilización.

Para realizar la transformación se utilizaban alrededor de 100 ng de DNA plasmidial, incubados con 100 µl de bacterias *E.coli* competentes en hielo por 30 minutos, para luego aplicar un shock térmico a 42°C por 90 segundos y enfriar rápidamente en hielo. Luego se adicionaba 1 ml de medio LB, se agitaba a 37°C por una hora, y se sembraba en placas Petri con medio LB con agar 1,5% [p/v] y el marcador de selección respectivo (ampicilina 50 mg/L, kanamicina 25 mg/L o espectinomicina 25 mg/L). Se incubaban las placas Petri a 37°C por 12- 16 horas y se analizaban las colonias por PCR para determinar si tenían el plasmido con la construcción deseada. Para esto se utilizaban los oligonucleótidos M13-F y M13-R específicos del vector o combinaciones de oligonucleótidos específicos del gen de interés. Finalmente, se crecía un inóculo de 5 ml de las colonias positivas y se realizaba una purificación de DNA plasmidial.

3.5.2 - Preparación y transformación de células *electro-competentes*

Para la preparación de células *electro-competentes* se inoculaban 500 ml de medio LB con 5 ml de un cultivo saturado de *Agrobacterium tumefaciens* GV3101. Se crecía el cultivo a 28°C con agitación hasta que alcanzara una D.O_{600 nm} de 0,5 - 0,8, se enfriaban las células en hielo y se mantenían a 4°C durante los pasos siguientes. Las células eran centrifugadas a 4000 g por 10 minutos y se descartaba el sobrenadante. Posteriormente, se resuspendía el precipitado muy suavemente en 10 ml de agua fría estéril y luego se ajustaba el volumen de suspensión a 100 ml con agua fría estéril. Se repetían la centrifugación y la resuspensión dos veces, ajustando la primera resuspensión a 50 ml y la segunda a 10 ml de volumen final. Se centrifugaba a 4000 g por 10 minutos y se resuspendían las células en 5 ml de 10% [v/v] de glicerol frío estéril.

Finalmente, las células eran alicuotadas en volúmenes de 100 μ l, para luego ser congeladas en nitrógeno líquido y ser guardadas a -80°C hasta su utilización.

Para transformar *Agrobacterium tumefaciens* (GV3101) se utilizó el electroporador “MicroPulserTM” de Biorad y las cubetas “Gen Pulser Cuvette” que poseen un electrodo de 0,2 mm. Se mantenían las cubetas por 10 minutos en hielo y luego se agregaban 100 ng de plásmido y 100 μ l de células electrocompetentes de *Agrobacterium*. Se aplicaba un pulso eléctrico bajo las siguientes condiciones: 25 μ F de capacitancia; 2,4 mV de voltaje; 200 Ω de resistencia y 5 milisegundos de duración. Se adicionaba 1 ml de medio LB a la cubeta, se homogeneizaba y luego se traspasaban las bacterias a un tubo Eppendorf de 1,5 ml. Se incubaba a 28°C por 2 horas en agitación constante, para luego centrifugar a 1000 g por un minuto y descartar la mayor parte del sobrenadante. Finalmente, una vez resuspendido el sedimento bacteriano, las células eran distribuidas homogéneamente en placas Petri con medio LB-agar que contenía los antibióticos rifampicina (50 mg/L), gentamicina (50 mg/L) y el antibiótico de resistencia del plásmido, y se dejaban creciendo a 28°C por 48 horas.

3.6 Extracción de DNA plasmidial.

Para la extracción de DNA plasmidial a partir de *E. coli*, se tomaba una colonia de la cepa de interés y se crecía durante toda la noche a 37°C con agitación, en 5 mL de medio LB. Posteriormente se utilizaba el sistema comercial “Axyprep[®] Plasmid Miniprep kit”. En breve, se centrifugaba el cultivo para obtener un sedimento de bacterias, se agregaban las soluciones I, II y III que permiten la ruptura de las células, la liberación y desnaturalización del DNA y la posterior renaturalización de éste, lo que separa el DNA cromosómico del plasmidial. A continuación se centrifugaba nuevamente para obtener el plásmido en solución y se empleaba una columna de afinidad para purificarlo de manera más eficiente y limpia. Con este método se obtienen entre 5 y 20 μ g de plásmido aproximadamente.

3.7 Secuenciación y alineamiento de secuencias.

El DNA plasmidial y los fragmentos amplificados por PCR eran secuenciados utilizando ya sea el servicio de secuenciación automática proporcionado por el Departamento de Ecología de la PUC o a través de Macrogen Korea (<http://www.macrogen.com/eng/sequencing/automatic.jsp>). Posteriormente se analizaban las secuencias utilizando la aplicación SeqMan del programa Lasergene-DNAstar (Burland, 2000) o DNADynamo (<http://www.bluetractorsoftware.co.uk/>). Los alineamientos de secuencias fueron realizados con las aplicaciones “SeqBuilder” y “MegAlign” contenidas en el programa Lasergene-DNAstar (Burland, 2000).

4.- Generación de plantas transgénicas de *Arabidopsis thaliana*.

4.1 Generación de construcciones en vectores binarios.

Para determinar la localización sub-celular de MEF31 se utilizó el vector pK7FWG2.0, el cual permite expresar una proteína de fusión entre la secuencia de interés y la proteína GFP bajo el control del promotor de RNA 35S del Virus del Mosaico de la Coliflor (CaMV35S). Se amplificó el DNA que codifica para el amino terminal de la proteína PPR desde DNA genómico de *Arabidopsis*, ya que este gen carece de intrones, utilizando los oligonucleótidos pENT_mef31_6050F y amino_mef31_R (Tabla 2). El oligonucleótido pENT_mef31_6050F tiene la secuencia 5'-CACC-3' en el extremo 5' que permite el clonamiento direccional en el vector pENTR/SD/D-TOPO. Se transformó *E.coli* y se seleccionaron las colonias positivas (Métodos 3.5.1). A partir de ellas se realizó la purificación de DNA plasmidial descrita en Métodos 3.6. Luego se procedió a la recombinación en el vector pK7FWG2.0 usando el sistema Gateway® como se indica en métodos 3.4.

Tabla 2: Lista de oligonucleótidos utilizados:

Oligonucleótido	Usado en	Orientación	Secuencia (5'-3')
M13-F	Secuenciación	Directo	GTAAAACGACGGCCAGT
M13-R	Secuenciación	Reverso	GCGGATAACAATTTCACACAGG
pENT_mef31_6050F	Localización	Directo	CACCATGCGTTTTACATTCTCCGATCA
pENT_mef31_STOP_R	Secuenciación	Reverso	CTAGATTGAGTAGTCCCCTAGCCAAC
amino_mef31_R	Localización	Reverso	ACCGTCTCGTTGAATGACCCCGTG
interno6050_E1_F	Genotipificación y secuenciación	Directo	TCATTCAACGAGACGGTGAC
6050-E1_LP	Genotipificación	Directo	TACATTCTCCGATCAACTCG
6050-E1_RP	Genotipificación	Reverso	TCCCAAATTGCAGAAAATTTG
LB1-SAIL	Genotipificación	Directo	GCCTTTTCAGAAATGGATAAATAGCCTTGCTTCC
orfX_4F	RT-PCR para verificar defectos en la edición	Directo	CTTCACTTTTAGCTTTGAATTAC
orfX_417_F	RT-PCR para verificar defectos en la edición	Directo	CGTTCCCAATGTTTGGCAC
orfX_775_R	RT-PCR para verificar defectos en la edición	Reverso	GCTCTCCCTCATTCCACTCGT

4.2 Transformación de Arabidopsis por el método de la inmersión floral.

Se realizaba la transformación estable de plantas por *Agrobacterium tumefaciens* con el método de la inmersión floral (Clough y Bent, 1998). Se preparaba un inóculo con 5 ml cultivo de *Agrobacterium* transformados en 100 ml de medio LB que contenía los antibióticos rifampicina 50mg/L, gentamicina 50mg/L y el antibiótico de selección específico. Se haría crecer el cultivo bacteriano a 28°C con agitación por toda la noche, se medía la D.O_{600 nm}, se centrifugaban las bacterias a 4000 g por 10 minutos y se las resuspendía en una solución fresca de sacarosa 5% [p/v], llevándolas a una D.O._{600 nm} de entre 0,5 y 0,8. Posteriormente, se traspasaba la solución a una placa Petri grande. Mientras se homogeneizaba la suspensión con un agitador magnético, se agregaba el surfactante Silwet L-77 a una concentración final de 0,025% [v/v] para luego sumergir las inflorescencias de Arabidopsis por 5 segundos en la suspensión bacteriana. Finalmente, se dejaban las plantas en oscuridad por una noche, en un pequeño invernadero cerrado para mantener la humedad, y al día siguiente se traspasaban a la cámara de crecimiento,

en donde eran crecidas hasta la obtención de las semillas T₁.

4.3 Selección de líneas transgénicas.

El plásmido pK7FWG2.0 presenta el gen de resistencia a kanamicina para la selección de plantas transformadas. Una vez obtenidas las semillas T₁, se esterilizaban y sembraban en medio MS 0,5X con kanamicina 50 mg/L. Después de aproximadamente 10 días de crecimiento en la cámara de cultivo era posible identificar líneas transgénicas ya que las plantas que no expresan la resistencia al antibiótico no continúan creciendo y no alcanzan a generar hojas verdaderas. Las plantas resistentes en T₁ son líneas independientes. Luego de dos semanas de crecimiento *in vitro* se trasapaba a tierra cada línea independiente, y eran crecidas hasta obtener semillas T₂. Para obtener plantas homocigotas mutantes con una sola inserción de T-DNA se seleccionaron las líneas en que las semillas T₂ dieran un 75% de plántulas resistentes al ser sembradas en un medio con el antibiótico (indicativo de una inserción). Finalmente se seleccionaron como homocigotas aquellas plantas en que la generación T₃, que fue totalmente resistente al antibiotico

5.- *Análisis de la localización sub-celular de MEF31*

5.1 Obtención de protoplastos.

Se aislaban protoplastos desde hojas de plantas de entre 4 y 6 semanas, homocigotas para la construcción 35S::aminoMEF31::eGFP (generación T₂ en adelante). Se utilizaba el método descrito por Wu y cols. (2009) llamado “Tape-Sandwich”, para preparar los protoplastos. Una vez cortadas las hojas se les colocaba cinta adhesiva (“tape”) por ambos lados, luego se retiraba rápidamente la cinta de la cara basal de la hoja, removiendo la epidermis, y se incubaba en una placa Petri con la cara basal embebida en la solución de digestión (MES 20 mM pH 5,7; Manitol 0,4 M; KCl 20 mM, Celulasa 1,5% [p/v] (Yakult), Macerozima 0,4% [p/v] (Sigma), CaCl₂ 10 mM, β- mercaptoetanol 2,5 mM y BSA 0,1% [p/v]), permitiendo así la digestión de la pared de las células del mesófilo. En cada preparación se utilizaban entre 7 y 10 hojas por 10 ml de

solución de digestión. Se agitaban las placas suavemente en luz por 3 horas, y luego se concentraban los protoplastos por centrifugación a 100 g por 3 minutos. Se lavaba el sedimento con 25 ml de la solución W5 (MES 2 mM pH5,7; NaCl 150mM, CaCl₂ 125mM, KCl 5mM y glucosa 5mM) y se centrifugaba a 100 g por 3 minutos. Después de repetir el lavado se resuspendían los protoplastos en 2 ml de la solución MMG (MES 4 mM pH 5,7; Manitol 0,4 M y MgCl₂ 15 mM), se incubaba 30 minutos en hielo y se contaban los protoplastos en una cámara de Neubauer. Aproximadamente, 100 µl de una solución que contenía $2-5 \times 10^4$ protoplastos por ml eran incubados con 100 µl de la solución W5 que contenía MitoTracker® Orange 0,1 µM por 5 minutos a temperatura ambiente, para luego proceder con los análisis de microscopía.

5.2 Análisis de microscopía

Primero se seleccionaban las líneas transgénicas que tuvieran señal de eGFP. Para esto se cortaban trozos de hojas de la generación T1, se los montaba en un porta objeto, se les cubría con un cubre-objeto y se los observaba en un microscopio de epifluorescencia “Nikon Eclipse 80i”. Se excitaban las muestras a 488 nm y solo se continuaba trabajando con las líneas que emitieran fluorescencia de la proteína fluorescente verde. Luego, para un mejor análisis de la localización sub-celular, se analizaban protoplastos o raíces de las líneas positivas en el microscopio confocal-espectral NIKON. Se preparaban protoplastos como se menciona en métodos 5.1, o se utilizaban plántulas de 15 días. Se incubaban las muestras con el marcador mitocondrial MitoTracker® Orange 0,1 µM, y luego las raíces de las plántulas o alícuotas de 100 µl de protoplastos eran montadas en porta objetos. Se cubrían las muestras con cubre-objetos y eran observadas en el microscopio. Las longitudes de onda para la excitación de eGFP y de MitoTracker fueron de 488nm y 561 nm, respectivamente. Las fluorescencias de eGFP y de MitoTracker® Orange eran detectadas a 510 nm y 595 nm, respectivamente.

6. Genotipificación de plantas mutantes *mef31*.

Para identificar y luego aislar plantas homocigotas mutantes se genotipificaban plantas de la línea mutante SAIL-451. Una vez obtenido el tejido adecuado para la extracción de DNA (métodos 2.1), se utilizaban dos reacciones de PCR para determinar el genotipo de las plantas. En una reacción se amplificaba el alelo silvestre utilizando los oligonucleótidos interno6050_E1_F y 6050-E1_RP, y en la otra el alelo mutado utilizando los oligonucleótidos LB1-SAIL y 6050-E1_RP (Tabla 2). Se clasificaban las plantas de acuerdo a los fragmentos amplificados: plantas silvestres eran aquellas en que solamente se amplificaba el alelo silvestre, heterocigotas aquellas en que se amplificaban ambos alelos y homocigotas mutantes aquellas en que se amplificaba solamente el alelo mutado.

7.- *Análisis de la edición del RNA.*

7.1 Secuenciación directa de productos de RT-PCR.

Los oligonucleótidos utilizados para la transcripción inversa y posterior PCR están indicados en la Tabla 2. Se purificaban los productos de PCR con el kit “NucleoSpin® Gel and PCR Clean-up” y posteriormente se mandaban a secuenciar a MacroGen Korea (Métodos 3.7). Finalmente, el porcentaje de edición del RNA era estimado analizando la altura de las señales del cromatograma utilizando el software “DNADynamo” (<http://www.bluetractorsoftware.co.uk/>) C y T si la secuencia era la de la hebra con-sentido, G y A si la secuencia era la de la hebra anti-sentido.

7.2 Secuenciación de clones.

Además de la secuenciación directa de los productos de RT-PCR, éstos eran clonados en el vector pGEM-T Easy y luego se transformaba *E.coli* DH5 α de acuerdo a métodos 3.4. Posteriormente, se realizaba un PCR a las colonias blancas, utilizando los oligonucleótidos M13F y M13R (Tabla 2), se aislaban los plásmidos de las colonias positivas y finalmente se mandaban a

secuenciar. Se analizaron entre 10 y 30 clones del transcrito *orfX* para cada genotipo, ya sea desde cDNA preparado desde hoja adulta o plántulas de 15 días.

Resultados

1.- El gen At2g46050 codifica para una proteína PPR mitocondrial.

El *locus* At2g46050 es uno de los 5 genes candidatos seleccionados de acuerdo a los criterios descritos en Resultados II.1. Este gen codificaría para una proteína de 590 aminoácidos con 11 motivos PPR y un dominio adicional en el extremo carboxilo terminal de tipo E (Figura 6A). Cuatro programas predictores de localización subcelular (MitoProtII, TargetP, Predotar e iPSORT) señalan que esta proteína presenta un péptido de destinación a mitocondrias.

Para determinar experimentalmente si el amino terminal de esta proteína PPR contiene la información para una destinación mitocondrial, se clonó la región que codifica para los primeros 117 aminoácidos en el vector pK7FWG2.0, tal como se describe en métodos 4.1. Con esta construcción se transformaron plantas de Arabidopsis (Métodos 4.2) y se analizó la localización de la proteína de fusión entre el péptido de destinación de la proteína codificada por el gen At2g46050 y GFP tanto en protoplastos como en raíces de plantas establemente transformadas con esta construcción (Métodos 5). Los resultados muestran que la proteína de fusión con GFP presenta el característico patrón puntiforme de las mitocondrias tanto en protoplastos como en raíces de plántulas de 15 días (Figura 6, paneles B y D). Paralelamente, la observación de células de la epidermis de hojas de plantas adultas (Figura 6 panel C), específicamente de un estoma, muestra un patrón similar al anterior y se observa a las mitocondrias rodeando a los cloroplastos. En ambos casos la señal de GFP co-localiza con el marcador mitocondrial MitoTracker (paneles B, C y D) y no parece en el caso de los protoplastos existir sobreposición entre la señal de GFP y la autofluorescencia de la clorofila (paneles B y C). Estos resultados sugieren que el gen At2g46050 codifica para una proteína PPR mitocondrial.

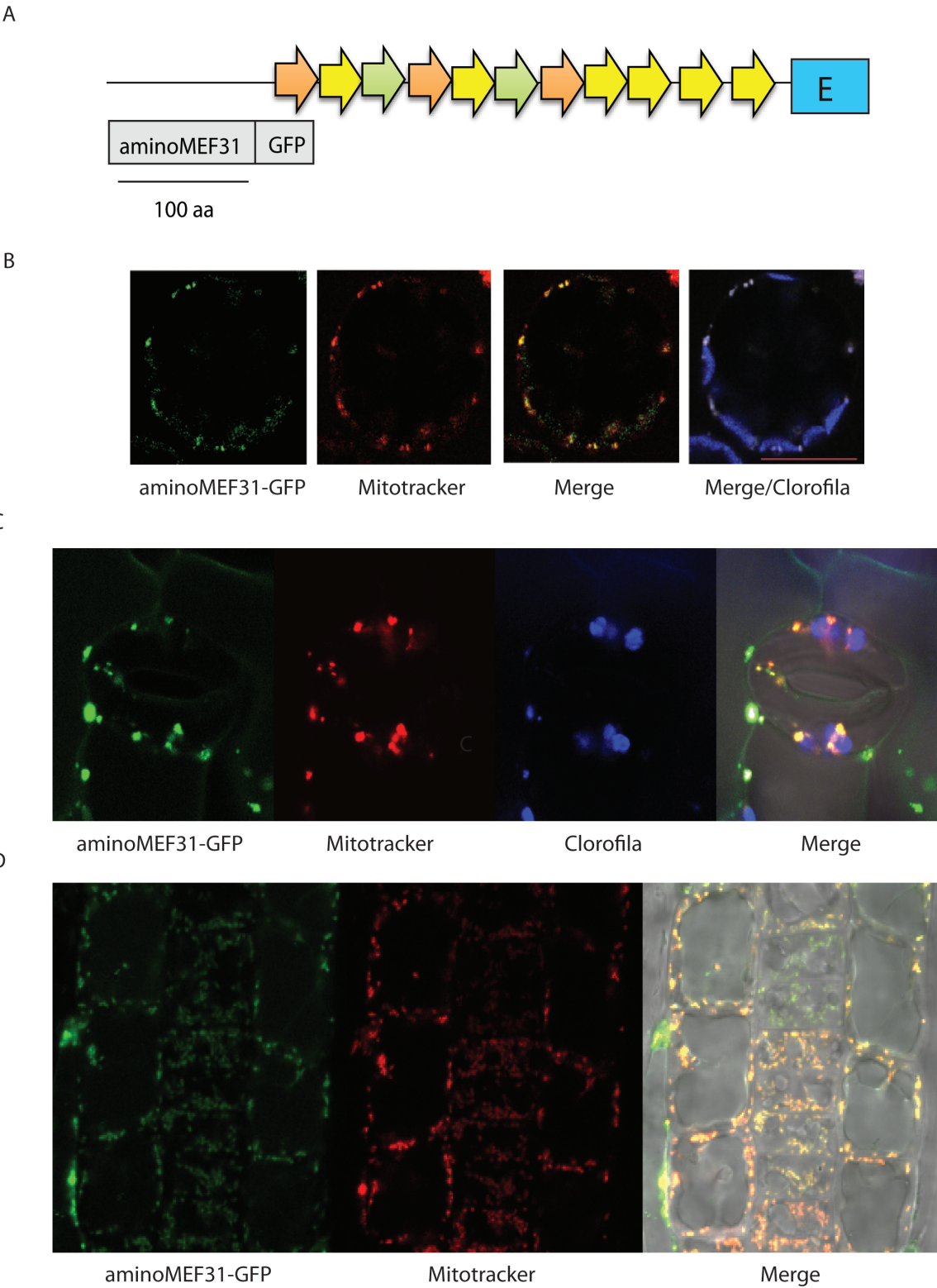


Figura 6. MEF31 presenta un péptido de destinación mitocondrial. (A) Organización de los 11 motivos PPR y del motivo E adicional en el extremo carboxilo terminal de la proteína MEF31 (Flechas naranjas: motivo PPR tipo L; flechas verdes: motivo PPR tipo P; flechas amarillas: motivo PPR tipo S; cuadro azul: motivo E; cuadro gris: segmento utilizado para la proteína de fusión a GFP). (B) A partir de hojas de plantas adultas 35S::aminoMEF31-GFP se prepararon protoplastos los que fueron analizados para las señales de fluorescencia de GFP y de Mitotracker Orange, para la superposición de ambos canales, y para la autofluorescencia de la clorofila. Barra = 10 μ m. (C) Estoma y (D) raíces de plántulas transgénicas fueron analizadas bajo las mismas condiciones que los protoplastos. Las señales de GFP y del marcador mitocondrial Mitotracker muestran mayoritariamente un solapamiento, con un característico patrón punctiforme de las mitocondrias.

2.- Caracterización de una línea mutante en el gen At2g46050

Para contar con una herramienta que permita analizar la función de la proteína PPR, se realizó una búsqueda *in silico* de mutantes en diferentes bases de datos. Se identificó una línea mutante para este gen en la colección Syngenta Arabidopsis Insertional Lines (SAIL-451-B04) y se solicitaron semillas de esta línea mutante (ecotipo Col0) al ABRC. Luego se realizó la caracterización molecular de la inserción. Mediante experimentos de PCR utilizando como molde DNA genómico de plantas y como partidores los oligonucleótidos 2 (6050-E1_RP) y LB (LB1-SAIL) (Figura 7A y Tabla 2), se identificaron plantas homocigotas y heterocigotas para la inserción (Figura 7B). La secuenciación del producto obtenido de plantas homocigotas mutantes (de aproximadamente 500 pb) confirmó que efectivamente se trataba de una inserción en la región codificante del gen At2g46050 y se estableció la posición exacta de la inserción en el 4° dominio PPR (Figura 7 panel A). Al menos en el lado 3' de la inserción, la secuenciación muestra que no hubo rearrreglos cromosómicos importantes, detectándose sólo la inserción de 10 nucleótidos de origen desconocido entre el borde izquierdo del T-DNA y el nucleótido 681 de la región codificante del gen At2g46050 (Figura 7 panel C, en rojo). Desafortunadamente no fue posible caracterizar el lado 5' de la inserción ya que los intentos de amplificarlo utilizando como partidores oligonucleótidos del gen y de los bordes derecho e izquierdo del T-DNA no dieron resultado. De todas formas, los resultados sugieren que en las plantas homocigotas mutantes no se podría sintetizar una proteína PPR funcional ya que la inserción está localizada antes del codon 227 (de un total de 590), por lo que de traducirse el mRNA el producto carecería de la mayoría de los dominios PPR. Se puede concluir que este mutante es una buena herramienta para analizar el fenotipo asociado a la falta de esta proteína PPR.

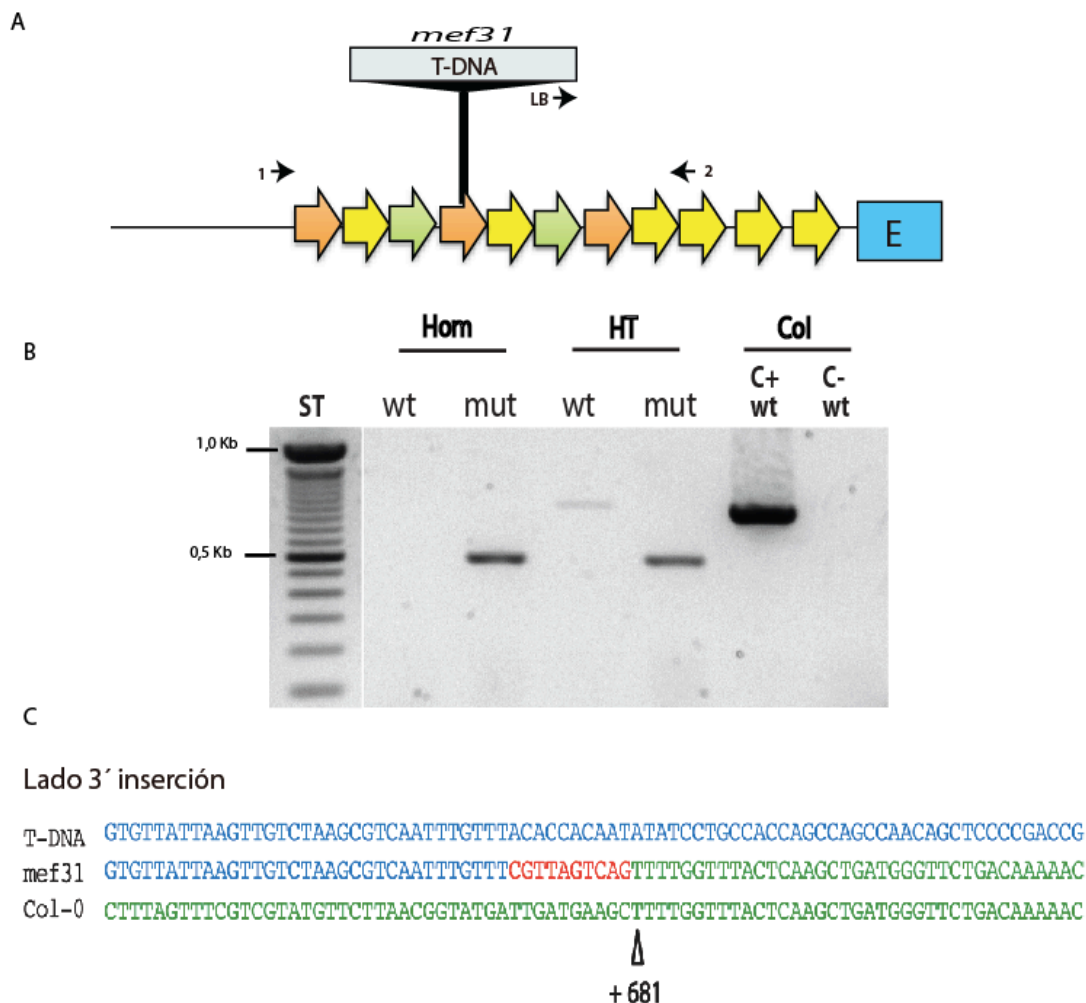


Figura 7. Estructura de la proteína PPR codificada por el gen At2g46050 y caracterización de una línea mutante insersional. (A) La proteína PPR contiene 11 motivos PPR (Las flechas naranjas, amarillas y verdes corresponden a los motivos L, S y P, respectivamente) y un dominio adicional de tipo E en el carboxilo terminal (cuadro azul). La línea mutante insersional SAIL-451 fue genotipificada utilizando dos pares de partidores: LB (específico líneas SAIL) y 2 (6050-E1-RP) para el alelo mutante, y los oligonucleótidos 1 (interno6050_E1_F) y 2 para el alelo silvestre. (B) Plantas homocigotas mutantes y heterocigotas fueron identificando amplificando por PCR con los oligonucleótidos especificados en (A) para: El alelo mutante (mut) se amplificó un fragmento de 500 pb y para el alelo silvestre (wt) uno de 700 pb. (C) Caracterización del lado 3' de la inserción de T-DNA en el locus At2g46059 (gen de *MEF31*). En el alineamiento se muestra

en la primera línea la secuencia del T-DNA (T-DNA, en azul), en la segunda línea se muestra el empalme entre la secuencia del T-DNA y la secuencia de *MEF31* en la línea mutante (*mef31*), y en la tercera línea la secuencia genómica del gen de *MEF31* silvestre (Col-0, en verde). En la secuencia del producto de PCR del alelo mutante se observa que no existieron reordenamientos cromosómicos importantes en el momento del establecimiento de la inserción de T-DNA en el lado 3' de la inserción, y que se insertaron de 10 pb (en rojo) de origen desconocido. Los números están en relación al primer nucleótido del codón de inicio ATG en el gen *MEF31*.

3.- Fenotipo molecular de las plantas mutantes *mef31*: defectos en la edición de un transcrito mitocondrial.

Dado que este gen codifica para una proteína PPR que posee el dominio carboxilo terminal de tipo E es probable que esta proteína pueda actuar como un factor en “*trans*” necesario para la edición de transcritos organelares. En colaboración con el grupo del Dr. Mizuki Takenaka en Alemania a quien se le enviaron semillas y RNA preparado desde plantas homocigotas mutantes, se evaluaron 269 sitios de edición descritos en mitocondrias y en cloroplastos de *Arabidopsis* mediante el método denominado SNaPshot (Takenaka y Brennicke, 2009). Se encontró un defecto en la edición del sitio 581 en el transcrito mitocondrial *orfX*, en una mezcla de RNAs provenientes de varias plantas homocigotas mutantes en distintos genes de proteínas PPRs (Figura 8A). Este sitio está completamente editado en plantas silvestres (Figura 8A, panel inferior). No se encontraron defectos en la edición de sitios cloroplásticos. A continuación se analizó la edición del sitio *orfX-581* en las mutantes individuales a partir de las cuales se había preparado la mezcla de los RNAs. Para ello se preparó cDNA por RT-PCR y éste fue secuenciado (Figura 8B). En el mutante homocigoto *At2g46050* el sitio *orfX-581* no es editado.

En plantas silvestres la edición del sitio *orfX-581* cambia un codón de prolina (CCA) por uno de leucina (CTA), aminoácido muy conservado en esa posición en plantas (Figura 8C). La ausencia de edición en las plantas mutantes *mef31* tendrá como consecuencia la síntesis de un polipéptido *orfX* con una prolina en lugar de la leucina conservada. Debido a esto se pensó que en la mutante podría verse afectada la función de *orfX* y quizás el desarrollo y crecimiento de las plantas. Sin embargo, las plantas mutantes no presentaron alteraciones fenotípicas detectables a simple vista, ni en el estadio de plántulas de 15 días (Figura 9A) ni en el estadio de plantas adultas (Figura 9B). En estos preliminares algunas plantas mutantes tenían un menor número de frutos (Figura 9B). Sin embargo experimentos adicionales son necesarios para validar este fenotipo.

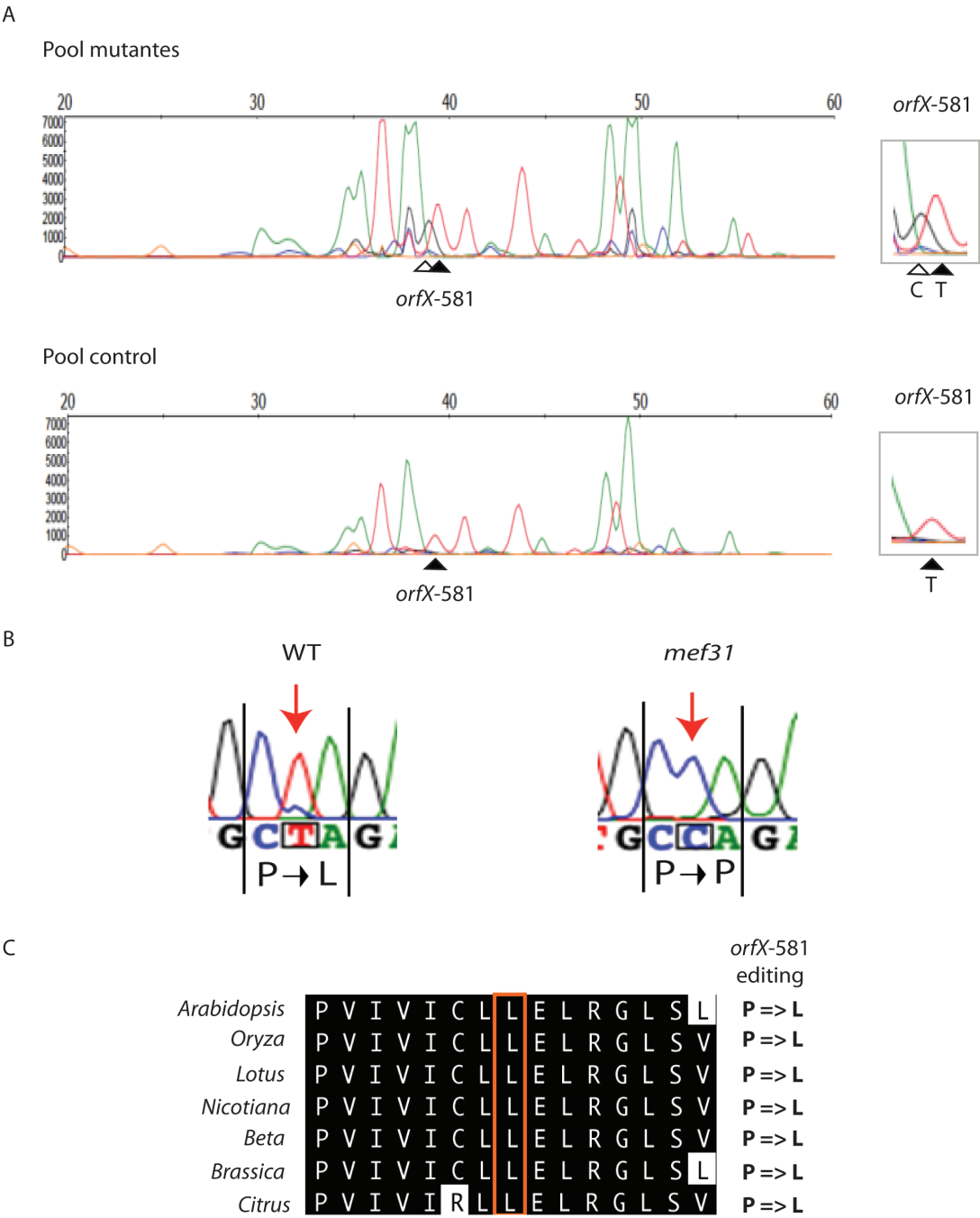


Figura 8. Análisis de SNaPshot identifica un defecto específico en la edición de un transcrito mitocondrial. Se utilizó en el ensayo un pool de RNAs de mutantes en diferentes proteínas PPR. (A) Análisis de RNA preparado desde una mezcla de hojas adultas de plantas homocigotas mutantes en ocho genes de proteínas PPRs (pool mutantes) y de RNA preparado desde plantas

silvestres (pool control). A la derecha se muestra una amplificación de la región del cromatograma en que se observa que el partidor para analizar el sitio *orfX-581* es elongado tanto con ddC como con ddT en el RNA del pool de mutantes y sólo con ddT en el RNA control. La detección de “C” (citósina) indica una disminución total o parcial de la edición en una de las plantas que forman el pool mutante (punta de flecha abierta). (B) MEF31 es esencial para la edición del sitio 581 en *orfX*. En la secuenciación directa de productos de RT-PCR de *orfX* desde RNA preparado de plántulas Col-0 (wt) de 15 días se observa un porcentaje de edición de aproximadamente un 90% para el sitio 581. Este sitio no es editado en el mutante *mef31*, lo que impide la conversión de un codón de Prolina en uno de Leucina. Los nucleótidos están representados por los colores azul para C, rojo para T, negro para G y verde para A. (C) El codón editado codifica para un aminoácido altamente conservado. Alineamiento parcial de la secuencia aminoacídica de proteínas ORFX de distintas plantas. Se muestra el alineamiento de 15 aminoácidos y en el recuadro rojo se destaca el aminoácido alterado en *mef31* en *Arabidopsis* (leucina). En todas las especies mostradas la Leucina es codificada por un codón editado en la posición 581 (P→L).

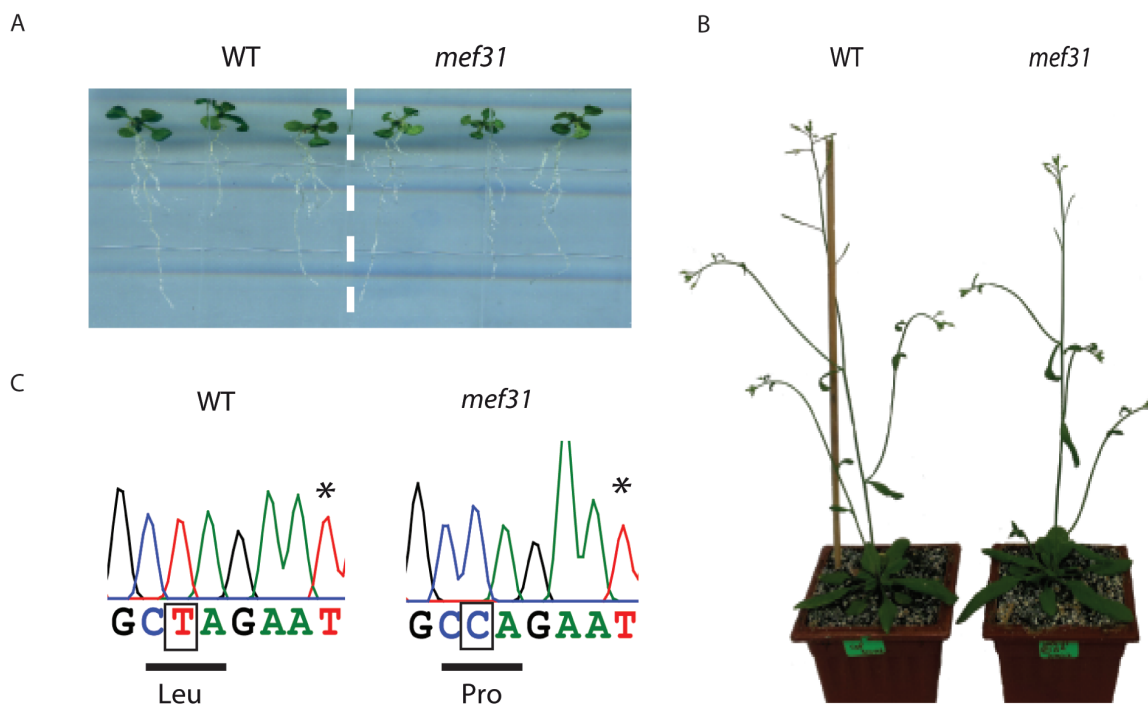


Figura 9. Plantas mutantes *mef31* presentan un desarrollo similar al de las plantas silvestres. A nivel de crecimiento vegetativo, no se observan diferencias significativas entre los genotipos silvestre (WT) y mutante (*mef31*) tanto en (A) plántulas de 15 días como en (B) plantas adultas de aproximadamente 8 semanas. (C) El análisis de la secuenciación de los productos de RT-PCR de RNA de hojas de plantas adultas para los transcritos de *orfX*, utilizando los oligonucleótidos *orfX*_4F y *orfX*_775R (Tabla 2), muestra la completa edición de los sitios 581 y 586 de *orfX* en plantas silvestres (WT). En la mutante *mef31* no es editado *orfX*-581 y el sitio 586 es editado normalmente. Se indicant los sitios 581 y 586 con un cuadro de color negro y con un asterisco, respectivamente.

En su conjunto, estos resultados sugieren que la proteína PPR codificada por el gen At2g46050 es mitocondrial y que es un factor de edición (MEF31) necesario para la edición del sitio 581 en el transcrito *orfX*. No obstante, dado que hasta ahora sólo se ha aislado una línea mutante homocigota de *mef31*, no se puede descartar que la ausencia de la edición del sitio *orfX*-581 se deba por ejemplo a una inserción en otro gen no identificado. Se está trabajando en la realización de un experimento de complementación, para lo cual actualmente se tiene la secuencia codificante del gen At2g46050 clonada en un vector de entrada del sistema Gateway. Paralelamente se está caracterizando una segunda línea mutante insercional en dicho gen.

Es interesante señalar la detección de una pequeña diferencia en cuanto al nivel de edición del sitio *orfX*-581 en plántulas de 15 días (Figura 8B) y en hojas de plantas adultas (Figura 9C). En hojas de plantas silvestres adultas se observa un 100% edición de este sitio, en cambio en plántulas el porcentaje de edición es de un 90%. Se ha confirmado esta edición parcial en plántulas por análisis de secuencia de clones obtenidos a partir del producto de RT-PCR (ver mas adelante).

4.- Los aminoácidos en las posiciones 6 y 1' de los motivos PPR de MEF31 predicen de acuerdo al código de las PPRs que el transcrito orfX es su RNA blanco.

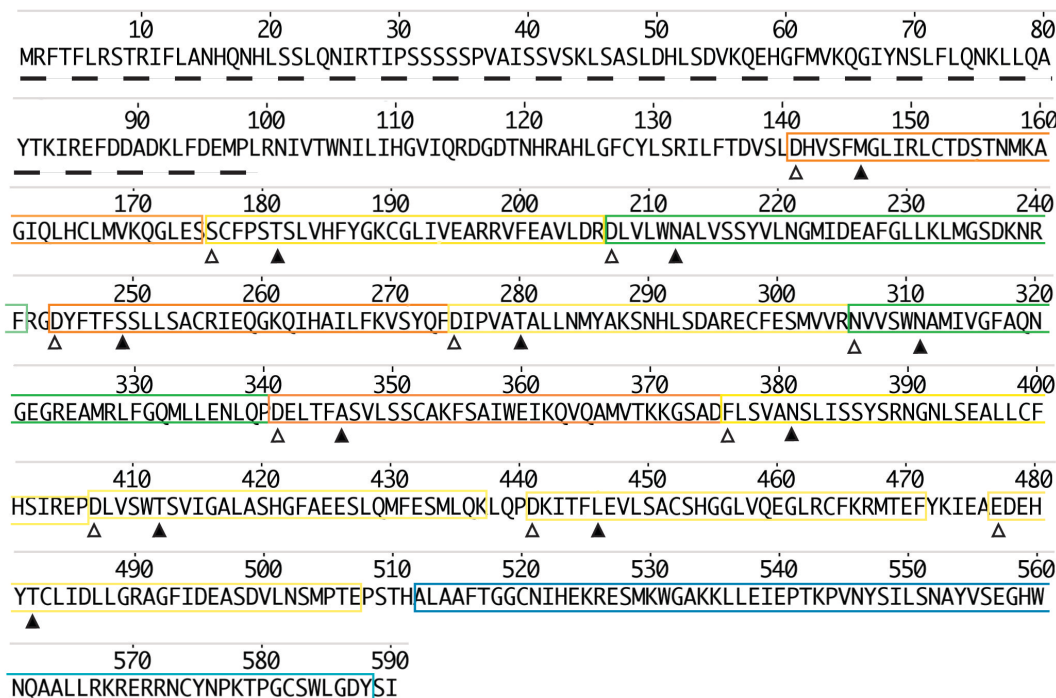
El código de las proteínas PPR surge del análisis de la frecuencia aminoacídica en los motivos PPR de proteínas a las cuales se les había identificado experimentalmente el transcrito diana y se había determinado mediante ensayos *in vitro* o *in vivo* la zona específica de interacción “proteína – RNA” (Barkan y cols., 2012; Ruwe y Schmitz-Linneweber, 2012; Zhelyazkova y cols., 2012; Okuda y Shikanai, 2012; Okuda y cols., 2006), para luego correlacionar combinaciones de aminoácidos de los motivos PPR con nucleótidos específicos de los transcritos. Este conjunto de datos implicó pares (proteína – transcrito) tanto mitocondriales como cloroplásticos, en un esfuerzo por determinar las bases del reconocimiento secuencia específico “proteína PPR – RNA”. El código propuesto en primera instancia por Barkan y colaboradores en el 2012, señala a dos residuos en la posiciones 6 y 1' (definido como el primer aminoácido de la siguiente repetición) de los motivos PPR de tipo P y S como los participantes en la interacción directa con el RNA y a los motivos L como “espaciadores” entre los elementos de tipo P y S generando así una organización de tipo PLS en las proteínas (Rivals y cols., 2006; Barkan y cols., 2012).

Cuando se analiza la secuencia aminoacídica de MEF31 compuesta por 590 aminoácidos se identifica una región amino terminal que contiene un péptido de destinación a mitocondrias, además de una región carboxilo terminal con un dominio extra de tipo E con 77 residuos. En la región central de la secuencia aminoacídica se identifican 11 repeticiones de tipo PPR organizadas como LS-PLS-PLS-SSS (Figura 10A). La identidad específica de los aminoácidos en las posiciones 6 y 1 de cada motivo de MEF31 es mostrada en la figura 10A, en donde predominan los residuos de treonina (T) y asparagina (N) en la posición 6 en las repeticiones S y P (en los motivos S cuatro de seis aminoácidos en esta posición corresponden a treoninas y en los dos motivos P sólo se identifican asparaginas). Por otra parte, el primer aminoácido de los motivos S y P es mayoritariamente un residuo cargado negativamente, como el aspartato (D).

Paralelamente, en los motivos L de MEF31 no se detecta una preferencia por un residuo específico en la primera y sexta posiciones.

Las reglas para el alineamiento propuesto por Barkan y cols. 2012, entre una proteína PPR implicada en la edición de un sitio en un transcrito organelar, toma como referencia la citosina (C) a editar y desde allí se consideran 4 nucleótidos río arriba para luego alinear el cuarto nucleótido (-4) con el aminoácido en la posición 6 de la primera repetición desde el extremo carboxilo terminal de la proteína. Se continúa sucesivamente con el nucleótido (-5), (-6), (-7), etc. alineado con los aminoácidos en las posiciones 6 y 1' del segundo, tercer, cuarto, etc. motivo PPR respectivamente desde el extremo carboxilo hasta incluir todas las repeticiones. En general, el número de nucleótidos del transcrito incluidos en el alineamiento dependerá del número total de motivos PPR existentes en la proteína. En el caso de MEF31 el alineamiento entre la proteína PPR y el RNA de *orfX* parte desde el nucleótido 581, que es la citosina a editar, y es mostrado en la figura 10 B: involucra 11 nucleótidos del transcrito alineados con 11 combinaciones de aminoácidos en las posiciones 6 y 1' pertenecientes a cada uno de los motivos PPR identificados. De acuerdo a Barkan y cols., 2012, los principales pares de aminoácidos que se encuentran en las posiciones 6 y 1' corresponden a treonina (T), asparagina (N), aspartato (D) y serina (S) generando combinaciones de tipo T6N1', T6D1', N6D1', N6N1', N6S1', N6T1', S6N1' y S6D1' que en conjunto están presentes en el 85% de todos los motivos P y S (Barkan y cols., 2012). La forma en que los motivos PPR reconocerían los nucleótidos estaría dada por las siguientes correspondencias aminoácidos – nucleótidos: T6D1' = G, T/S6N1' = A, N6D1' = U, N6N/S1' = C (Barkan y cols., 2012; Takenaka y cols., 2013). En nuestro caso, encontramos en los dos motivos P de MEF31 la combinación N6D1' que tendría una correspondencia con nucleótidos de tipo pirimidinas específicamente uracilos (U) en las posiciones -9 y -12 de *orfX*. De igual modo, en los motivos S se identifican los pares T6D1' y T6N1', los cuales tendrían correspondencia con purinas, guaninas (G) en las posiciones -6 y -13 y adenina (A) en -10. Finalmente, la combinación N6D1' se correlaciona positivamente con una pirimidina uracilo (U)

A

MEF31

B

motivo PPR	L	S	P	L	S	P	L	S	S	S	S
pos. 6	M	T	N	S	T	N	A	N	T	L	T
pos. 1'	S	D	D	D	N	D	F	D	D	E	
<i>orfX</i>	u	u	g	u	g	a	u	c	u	g	u
	-14	-13	-12	-11	-10	-9	-8	-7	-6	-5	-4
											-3
											-2
											-1
											*

Figura 10. Predicción de los residuos de MEF31 implicados en el reconocimiento de la secuencia del RNA de *orfX*. (A) Se muestra la secuencia aminoacídica de MEF31 compuesta por 590 residuos, el péptido de destinación (línea segmentada), las 11 repeticiones PPR de tipo L, S y P en recuadros naranjos, amarillos y verdes respectivamente, y el dominio adicional tipo E en el carboxilo terminal (recuadro azul) (<http://www.uniprot.org/uniprot/O82363>). Las puntas de flecha abierta señalan al aminoácido en la posición 1 de cada motivo PPR; Puntas de flecha cerrada señalan al residuo en la posición 6 de cada repetición. (B) Las combinaciones de aminoácidos en las posiciones 6 y 1' (primer aminoácido de la siguiente repetición) de cada uno de los motivos de la proteína PPR MEF31 han sido alineadas con la secuencia de *orfX* cuatro nucleótidos río arriba del sitio 581 (asterisco). Los números bajo los nucleótidos indican la posición en *orfX* en relación a la citosina 581. Los recuadros abiertos señalan a aquellos nucleótidos de la secuencia del RNA que tienen la correlación esperada para dicha combinación de aminoácidos de acuerdo al “código de las PPRs” propuesto por Barkan y cols., 2012. El recuadro gris cerrado indica que la correlación no es esperada.

en la posición -7 del transcrito. El par aminoacídico L6E1' en el penúltimo motivo S de MEF31 no muestra una correlación directa con la identidad del nucleótido en la posición -5 de acuerdo al código de las PPRs (Figura 10 B). Podemos proponer entonces que la identidad de los nucleótidos río arriba del sitio 581 en el transcrito mitocondrial *orfX* está correlacionado con los aminoácidos en la posiciones 6 y 1' de las repeticiones P y S de la proteína MEF31 de acuerdo al código de las PPRs propuesto por Barkan y cols., 2012. Esto fortalece la idea de que la proteína PPR MEF31 reconoce al transcrito *orfX* uniéndose a la region río arriba de la C581 a través de sus motivos PPR, y es esencial para la edición de este sitio.

5.- Análisis de la edición de sitios cercanos a orfX-581 en plantas mutantes *mef31*

La edición del sitio *orfX-581* está conservada en monocotiledóneas (ej. *Oryza sativa*) y dicotiledóneas (ej. *Brassica napus*) (Figura 8, panel C y Figura 11, panel A). En *Arabidopsis* y arroz también existe edición del sitio *orfX-552*, pero este evento es “silencioso”, es decir no cambia la identidad del codón (TCC y TCT codifican para una serina). En *Arabidopsis* existe edición de los sitios *orfX-586* y *587*, ambos en el mismo codón, de manera que el codón CCA (prolina) es modificado a TTA (leucina). Sin embargo, si el sitio *587* es editado (2ª posición del codón) y la edición en el sitio *586* (1ª posición del codón) pasa a ser silenciosa, porque CTA y TTA codifican ambos para leucina. En este contexto es interesante señalar que en *O. sativa* y *B. napus* sólo el sitio *orfX-587* es editado, lo que basta para cambiar la identidad del codón (Pro→Leu).

En este trabajo de tesis se analizó la edición en *Arabidopsis* de los 4 sitios, *orfX-552*, *581*, *586* y *587*, en hojas de plantas adultas y plántulas de 15 días de los genotipos silvestre y mutantes *mef31* (Figura 11B). En este caso se obtuvieron los cDNAs por RT-PCR, se les clonó y se secuenciaron varios clones para cuantificar el grado de edición en cada sitio. Aunque los resultados están sujetos a confirmación, especialmente en el caso de hojas adultas debido al reducido número de clones secuenciados (10 clones v/s 30 clones en el caso de plántulas), su análisis es interesante. En primer lugar se observa que los sitios de edición silenciosos (*552* y *586*) están parcialmente editados, un hecho conocido que guardaría relación con que no tienen efecto en la proteína. En el caso del sitio *552* (previamente descrito por Bentolila y cols., 2010 como un diana de la proteína PPR REME1) no parece haber diferencias significativas entre plantas silvestres (wt) y *mef31*, pero sí parece ser la edición más eficiente en hojas que en plántulas (un 73-56% editado en hojas respecto de un 18-16% de edición en plántulas). En segundo lugar, se aprecia que el sitio *587* es completamente editado en todos los casos, de lo que se puede concluir que pese a su cercanía no depende ni de la edición previa en los sitios *581* y *586*, ni de MEF31. En tercer lugar se confirma que la edición del sitio *581* requiere de MEF31

tanto en hojas como en plántulas. La observación de una edición residual en la mutante *mef31*, de confirmarse (por ahora basado en un clon de 30), podría indicar que en ausencia de MEF31 podría otra proteína PPR permitir una misma edición del sitio. Posiblemente lo más interesante es que la edición del sitio 586 está disminuida en las plantas *mef31* tanto en hojas como en plántulas. Ello es consistente con que el sitio 581 forma parte del elemento en *cis* necesario para la edición del sitio 586, y que debería estar editado para que el sitio 586 sea reconocido por la maquinaria en *trans* (se supone otra proteína PPR). Aunque el sitio 581 también podría formar parte del elemento *cis* para la edición del sitio 587, es posible que la identidad del nucleótido en esa posición (C o U) no sea un determinante.



Figura 11. Patrón de conservación y edición de sitios adyacentes al nucleótido 581 en el transcrito de *orfX*. (A) Fragmento del alineamiento de la secuencia nucleotídica del transcrito *orfX* alrededor del sitio de edición diana de MEF31 (estrella: *orfX*-581) en *Brassica napus* (*Bn*), *Oryza sativa* (*Os*) y *Arabidopsis thaliana* (*At*). La posición 581 (desde el primer AUG en *Arabidopsis*) y los otros sitios a editar (*orfX*-552, *orfX*-586 and *orfX*-587) están indicados en rojo con una “C” (no editados en esa especie) o una “U” (editados en esa especie). Se indica con corchetes grises debajo de la secuencia los codones correspondientes a este segmento de *orfX*. (B) Los RNAs fueron preparados desde hojas adultas y plántulas de 15 días de plantas *Col-0* (wt) y

mef31. Los cDNAs fueron obtenidos por RT-PCR utilizando los oligonucleotides orfX_4F y orfX_775R (Tabla 2). Treinta clones en el caso de plántulas y diez clones en el caso de hojas fueron secuenciados para cada genotipo. Se muestra el resultado del análisis de los clones esquemáticamente, utilizando el sistema “logo” en donde la altura de la letra representa el porcentaje de clones editados o no-editados en esa posición específica. T: en rojo, sitio editado, C: en azul, sitio no-editado. Los números bajo cada logo representan el porcentaje de edición.

Discusión

El gen At2g46050 codificaría para un nuevo factor de edición mitocondrial (MEF31) en Arabidopsis thaliana

El gen At2g46050 codifica para una proteína PPR de tipo PLS, es decir no solo contiene repeticiones canónicas de tipo P en su secuencia, sino que también posee motivos de 31 aminoácidos de tipo S y motivos L de 35 ó 36 residuos. A esto se suma la existencia de un dominio adicional en el carboxilo terminal de tipo E, por lo que es probable que esta proteína actúe como un factor en “*trans*” necesario para la edición de transcritos organelares ya que aproximadamente el 50% de las proteínas PPR mitocondriales descritas como implicadas en el proceso de edición presentan estas características en su secuencia aminoacídica (Barkan y Small, 2014). Para dilucidar entonces cuál o cuáles podrían ser los transcritos diana de esta proteína PPR, con la aproximación de SNaPshot se evaluaron cerca de 300 sitios de edición tanto de mitocondrias como de cloroplastos de *Arabidopsis* (Takenaka and Brennicke, 2009). Se encontró que el sitio 581 en el transcrito *orfX* no estaba editado en plantas homocigotas mutantes para este gen. No podemos descartar que otros sitios de edición no evaluados en el ensayo anterior, pudiesen también ser blanco de MEF31. Por otro lado, no debería sorprendernos el hecho que sólo el sitio 581 se viera afectado en esta mutante, si consideramos que de la totalidad de los factores PPR mitocondriales de *Arabidopsis* de tipo E descritos en literatura sólo dos de ellos, OTP87 y SLO2, son necesarios para la edición de más de una citosina, específicamente dos y cinco sitios en distintos transcritos respectivamente (Hammani y cols., 2011; Zhu. y cols., 2012). La capacidad de algunos factores PPR para participar o al menos ser necesarios para la eficiente edición de múltiples sitios en distintos transcritos podría estar dada por una elevada identidad nucleotídica alrededor de la citosina a editar. Se podría esperar entonces que sitios contiguos o muy cercanos en la secuencia de RNA, como es el caso de los sitios 581 y 586 de *orfX* estuviesen influenciados por un único factor PPR. No obstante se ha descrito que dos sitios contiguos a

editar pueden ser reconocidos por factores PPR diferentes (Arenas-M y cols., 2013). Estos resultados en su conjunto, tanto el fenotipo molecular detectado en la mutante de *mef31* como la demostración experimental de la existencia de un péptido de destinación mitocondrial en la proteína, sugieren fuertemente que MEF31 es un factor mitocondrial necesario para la edición de un transcrito de este organelo de Arabidopsis. Es necesario, sin embargo validar el fenotipo molecular complementando plantas mutantes *mef31* con el gen silvestre y/o caracterizando una segunda línea mutante independiente.

El código de las proteínas PPR correlaciona aminoácidos en posiciones específicas de los motivos PPR de MEF31 con la identidad de los nucleótidos adyacentes al sitio orfX-581

El código de las proteínas PPR hipotetiza sobre la capacidad que tendrían aminoácidos específicos en cada uno de los motivos PPR para distinguir purinas de pirimidinas en la secuencia nucleotídica de su o sus transcritos diana. La mayoría de esas correspondencias se basan en correlaciones estadísticas de alineamientos descritos entre factores PPR y sus RNAs diana (Barkan y cols., 2012; Okuda y cols., 2014). Es así como en el análisis de la identidad de los nucleótidos río arriba del sitio de edición 581 en el transcrito mitocondrial *orfX* vemos que existe una correlación con los aminoácidos en las posiciones 6 y 1' de las repeticiones P y S de la proteína MEF31 de acuerdo al código de las proteínas PPR propuesto por Barkan y cols., 2012 (Figura 10). Ello sustenta la hipótesis de que la proteína PPR MEF31 reconocería al transcrito *orfX* y sería necesaria para la edición del sitio 581. Sin embargo, aún es reducido el número de pares proteína PPR-RNA para los que se ha demostrado experimentalmente la unión. Un avance promisorio que fortalece la validez del código está constituido por los resultados recientemente presentados por Kindgren y cols., (2015), quienes mutaron sistemáticamente los aminoácidos en las posiciones 6 y 1' de los motivos P, L y S del factor de edición cloroplástico CLB19. Demostraron con ello la capacidad de modificar *in vivo* y en forma predecible la especificidad de un factor PPR (Kindgren y cols., 2015). CLB19 es un factor de edición de dos sitios en los

transcritos cloroplásticos *rpoA* y *clpP*. Esta proteína PPR tiene 10 motivos PPR en su secuencia y cuando se realiza el alineamiento de CLB19 con cada uno de sus transcritos diana (Figura 12 panel A) se observa que cuatro de las combinaciones de aminoácidos en las posiciones 6 y 1' de los motivos P y S son compartidas con MEF31. Se trata específicamente las combinaciones T6N1' (dos en CLB19, una en MEF31) y N6D1' (dos en CLB19, tres en MEF31), que son capaces de discriminar *in vivo* nucleótidos de tipo purinas y pirimidinas, respectivamente (Kindgren y cols., 2015). También ambas proteínas PPR presentan la combinación T6D1' que une preferentemente a G. El análisis del alineamiento entre MEF31 y su transcrito Diana (Figura 10B) sugiere que mutando las posiciones 6 y 1' de los motivos S y P con las combinaciones T6N1', N6D1' y T6D1' podría alterar la especificidad de MEF31.

La edición del sitio 581 podría influenciar la edición de un sitio adyacente en el transcrito orfX

En el análisis de clones de cDNA del transcrito *orfX* se observó una notoria disminución del porcentaje de edición de un sitio adyacente al sitio diana de MEF31, el sitio *orfX*-586 ubicado 4 nucleótidos río abajo del sitio 581 (Figura 11). Tanto en hojas como en plántulas el porcentaje de edición disminuye significativamente en la mutante *mef31* respecto de Col 0. Una explicación plausible para este interesante efecto es que el sitio 581 forme parte del elemento *cis* de reconocimiento de *orfX*-586 por otra proteína PPR, y que esta tenga una mayor afinidad cuando el sitio 581 está editado (U en lugar de C). Dos argumentos sugieren que es otra proteína PPR y no MEF31, la que reconoce al sitio 586: por un lado en la mutante *mef31* no se pierde completamente la edición del sitio 586, sugiriendo la existencia de otro factor, y por otro lado los pares 6, 1' de los motivos PPR de MEF31 alinean mal con el transcrito para editar el sitio 586, de acuerdo al código PPR (Figura 12 B). Es importante señalar que un trabajo previo Kindgren y cols., 2015 resalta la importancia que tiene la estructura secundaria de la molécula de RNA en la afinidad de unión de un específico factor PPR, por lo que también se podría pensar en la posibilidad de que la edición previa del sitio 581 podría generar una estructura secundaria mas

Figura 12: Alineamiento del factor de edición cloroplástico CLB19 con sus transcritos diana y del factor MEF31 con un sitio de edición cercano a *orfX-581*. (A) Se muestra para la proteína PPR CLB19 la organización de los tipos de motivos PPR en la proteína (primera línea), debajo se muestran los aminoácidos en las posiciones 6 (2° línea) y 1' (3° línea) de cada repetición, y se indica el grado de correspondencia entre cada par de aminoácidos y el nucleótido en el RNA diana de acuerdo al código de las PPR. (:) señala correspondencia entre la combinación de aminoácidos y nucleótido, validada experimentalmente. (.) correspondencia predicha computacionalmente y (X) correspondencia no esperada (Modificado desde Barkan y cols., 2012). (B) Las combinaciones de aminoácidos en las posiciones 6 y 1' de cada uno de los motivos de la proteína PPR MEF31 son alineadas respecto del sitio *orfX-586* (asterisco). Recuadros abiertos señalan aquellos nucleótidos de la secuencia del RNA que tienen la correlación esperada para dicha combinación de aminoácidos de acuerdo al “código de las PPRs” propuesto por Barkan y cols., 2012. En gris se muestran nucleótidos que no serían reconocidos por el par aminoacídico de acuerdo al código.

El comportamiento antes propuesto para el factor MEF31 sobre estos dos sitios podría denominarse “intermedio” si se le compara con otros factores PPR previamente descritos tales como SLO2 y MEF25 y su efecto sobre dos sitios de edición contiguos. En la mutante *slo2* se pierde completamente la edición de los sitios *mttB-144* y *145* (*orfX-144* y *145*), sugiriendo que SLO2 es esencial para la edición de ambos sitios (Zhu y cols., 2012). Por otra parte en la mutante *mef25* la edición del sitio *nad1-307* se ve inalterada mientras que la edición del sitio contiguo *nad1-308* es abolida en dos líneas mutantes independientes, sugiriendo con ello que la edición de ambos sitios contiguos depende de dos factores PPR independientes (Arenas-M y cols., 2013). En resumen, los antecedentes sugieren que MEF31 tendría como diana directo al sitio 581 y que el efecto de esta proteína PPR sobre *orfX-586* sería indirecto y podría estar dado por la modificación del elemento en *cis* adyacente a la citosina 586, facilitando así la unión de un segundo factor de edición.

Putativa función de la proteína codificada por orfX (mttB o tatC) en mitocondrias de plantas

El factor de edición mitocondrial MEF31 descrito en este capítulo tendría una influencia directa en la edición de un sitio en el transcrito *orfX* de Arabidopsis. La proteína ORFX también denominada *mttB* por “**m**embrane **t**argeting and **t**ranslocation”, correspondería a un homólogo de la proteína bacteriana *tatC* (**t**win **a**rginine **t**ranslocation) que forma parte de un sistema de transporte y translocación de proteínas plegadas a través de la membrana (Bogsch y cols., 1998). Ha sido propuesto que la existencia de homólogos en mitocondrias de plantas de este sistema de translocación bacteriano permitiría por ejemplo la correcta localización y ensamblamiento de complejos proteicos que contengan cofactores del tipo FeS en la membrana interna de la mitocondria (Marienfield y cols., 1999). Para realizar esta función la proteína *tatC* contiene en su secuencia numerosos dominios transmembrana. En el caso de la proteína codificada por el gen mitocondrial *orfX* de Arabidopsis se identificaron, en su secuencia de 287 aminoácidos, utilizando el programa TMHMM (<http://www.cbs.dtu.dk/services/TMHMM/>), 6 dominios

transmembrana con un promedio de 23 residuos en cada dominio (datos no mostrados). Si se analiza la posición de cada uno de los 32 sitios de edición encontrados en el transcrito *orfX*, queda en evidencia que la mayoría de ellos contribuye a la modificación de aminoácidos ubicados en los dominios transmembrana. Es así como la edición del sitio *orfX-581* modifica el último residuo del 5° dominio transmembrana, cambiando una prolina por una leucina altamente conservada. Sin embargo, la ausencia de edición de este sitio no sería esencial para la función de esta putativa proteína tatC en mitocondrias dado que las plantas mutantes en *mef31* no presentan diferencias significativas respecto de las plantas silvestres (Figura 9) al menos bajo condiciones de crecimiento estándar. No obstante, ha sido descrito un aumento en la expresión del transcrito *orfX* (tatC) en células de tabaco en presencia de ácido salicílico (SA) (van der Merwe y Dubery, 2007) sugiriendo con ello la posibilidad de que esta proteína tenga un rol preponderante en la translocación de proteínas a la membrana interna mitocondrial bajo condiciones de estrés oxidativo o daño celular (van der Merwe y Dubery, 2007). Con este antecedente se podría pensar que las plantas mutantes *mef31* podrían presentar un fenotipo alterado bajo condiciones de estrés celular.

Conclusiones

- I. La proteína MEF31 está localizada en la mitocondria.
- II. MEF31 sería esencial para la edición del sitio 581 del transcrito mitocondrial *orfX*. En ausencia de edición de este sitio la proteína ORFX contendría una Pro194 en lugar de una Leu194 conservada en distintas especies, pero ello no afecta el crecimiento y desarrollo en las plantas mutantes al menos bajo condiciones estándar
- III. La confirmación del fenotipo molecular de la mutante de *mef31* necesita del análisis de una segunda línea mutante independiente o de un ensayo de complementación.
- IV. La aplicación del recientemente propuesto “código de las PPR” predice que la proteína MEF31 identificaría específicamente al sitio *orfX-581*, sustentando la hipótesis de que el transcrito *orfX* es una diana directa de MEF31.

CAPÍTULO III

DISCUSIÓN Y CONCLUSIONES GENERALES

La edición del RNA es un proceso que causa alteraciones en la secuencia de moléculas de RNA, generando así transcritos maduros que difieren de la información contenida en el genoma. Es posible que la edición de RNA sea un tipo de procesamiento post-transcripcional utilizado como un punto de control que permita restaurar funciones de proteínas codificadas en el genoma, o que pueda servir como un reservorio para crear proteínas diferentes con nuevas funciones. Desde los estudios iniciales, la edición de RNA llamó la atención de muchos investigadores por sus características y particularidades. La modificación de citosinas específicas en uracilos ocurre en transcritos mitocondriales y plastidiales de la mayoría de las plantas terrestres, sin embargo hay una diferencia respecto del número de sitios editados en ambos organelos: más de 400 sitios en mitocondrias y entre 30 y 40 sitios en cloroplastos. El modelo actual para la edición de RNAs organelares incluye un elemento en *cis* situado río arriba de la citosina a editar, a un factor en *trans* que reconoce al elemento en *cis* y que guiaría la actividad de edición, y a la enzima que realiza la reacción de desaminación. Mientras se ha avanzado en la caracterización de algunos de los elementos en *cis* y los factores PPR que los reconocen, la identidad de la enzima que cataliza la desaminación es desconocida. Es claro eso sí, que el genoma mitocondrial no codifica para ninguno de los factores de especificidad, ni tampoco para la desaminasa (Hammani y Giege, 2014), por lo que toda la maquinaria que actúa en *trans* está codificada en el núcleo.

Este trabajo de tesis ha tenido como objetivo la identificación de factores de especificidad necesarios para la edición de transcritos mitocondriales de *Arabidopsis*. En el año 2009, fecha en que se inició este trabajo de tesis, sólo 6 factores habían sido descritos como implicados en la edición de RNAs organelares (Kotera y cols., 2005; Okuda y cols., 2006; Chateigner-Boutin y cols., 2008; Okuda y cols., 2009), y de ellos sólo MEF1 correspondía a un factor de edición mitocondrial (Zehrmann y cols., 2009). Todos estos factores tenían en común estar codificados en el núcleo y pertenecer a la numerosa y, en ese momento, poco conocida familia de proteínas PPR. Es así como utilizando una aproximación que incluyó algunos criterios de selección de

genes PPR candidatos, el uso de herramientas genéticas (mutantes, plantas complementadas), y de un método de análisis masivo de los sitios de edición organelar, se identificaron y caracterizaron tres nuevos Factores de Edición Mitochondrial (MEFs): MEF25, MEF26 y MEF31.

1.- Criterios de selección de proteínas PPR candidatas.

En el año 2009 las principales estrategias utilizadas para la búsqueda de factores implicados en la edición de RNAs organelares evaluaban variantes naturales ecotipo específicas en la edición de sitios mitocondriales de *Arabidopsis*, utilizando metodologías tales como “extension de un partidador” (Takenaka y cols., 2008; Zehrmann y cols., 2009), mapeo genético asociado a polimorfismos en la edición mitochondrial (Bentolila y cols., 2008 y 2010) o el screening de mutantes en genes de proteínas PPR que presentasen defectos en la edición de sitios plastidiales (Hammani y cols., 2009). En esta tesis se planteó realizar una búsqueda “dirigida” de factores proteicos importantes para la edición de transcritos mitocondriales pensando en cuales debían ser la características básicas de estos factores. Naturalmente, las proteínas PPR eran las candidatas ideales ya que luego de la descripción de MEF1 (Zehrmann y cols., 2009) se consolidaba la hipótesis de que podría existir una relación directa entre los numerosos sitios de edición mitochondrial y la explosiva expansión de genes de proteínas PPR en plantas terrestres, especialmente de la subfamilia PLS (Rüdinger y cols., 2008; Fuji y Small, 2011).

Los cuatro criterios de búsqueda utilizados fueron: i) una primera selección de genes con motivos PPR que tuvieran secuencias de ESTs o cDNAs en las bases de datos, para tener una prueba de que estos genes eran funcionales y se estaban transcribiendo. Con este primer criterio se redujo el número de genes de los 441 iniciales a 329. ii) El segundo criterio fue la predicción de la localización subcelular de las proteínas PPR candidatas. Existen múltiples algoritmos disponibles que realizan estas predicciones, y son mayoritariamente de dos tipos: basados en el reconocimiento de motivos característicos en la región amino-terminal de las proteínas o basados

en la composición aminoacídica (Nakai, 2000). La plataforma SUBA (The Arabidopsis Subcellular Database) permitió el análisis simultáneo de las secuencias aminoácídicas candidatas con 4 programas predictores pertenecientes a ambos tipos de algoritmos, buscando en ello una forma de predicción “integrada” de la destinación subcelular. Esto permitió reducir a 207 los candidatos con una localización mitocondrial predicha por al menos dos algoritmos predictivos.

iii) El tercer criterio fue dejar sólo a aquellas proteínas PPRs pertenecientes a la subfamilia PLS, de las subclases E, E+ o DYW, dado que hasta ese momento las 6 proteínas PPR implicadas en la edición de transcritos organelares pertenecían a esta subfamilia. Con ellos se redujeron los candidatos a 73. Trabajos posteriores han avalado la importancia de este criterio, ya que se ha demostrado que dominio E es esencial para el papel en la edición de algunas proteínas PPR (Okuda y cols., 2007; Härtel y cols., 2013).

iv) Finalmente la última selección fue realizada considerando el porcentaje de identidad de secuencia de las PPR candidatas con otras proteínas PPR de Arabidopsis con lo cual se jerarquizó las proteínas candidatas. Se escogieron aquellas con la menor identidad posible para reducir las posibilidades de redundancia funcional y aumentar las posibilidades de observar alteraciones en la edición en mutantes insercionales en dichos genes. Dado el elevado número de genes de proteínas PPR en el genoma de *Arabidopsis* (Lurin y cols., 2004), y la hipótesis de que la expansión de la familia se debe a eventos masivos de duplicaciones génicas (O’Toole y cols., 2008), la posibilidad de redundancia funcional es real. Los cinco candidatos finalmente seleccionados tenían una identidad igual o menor al 40% con la mas parecida de las otras proteínas PPR de Arabidopsis.

Es pertinente señalar respecto del último criterio que el umbral utilizado de (40%) no es muy bajo. Sin embargo, las proteínas PPR estan constituidas por repeticiones de 35 aminoácidos (y de sus variantes), lo que hace que a menudo presenten identidades del 30% aunque no sean ortólogos. De todas maneras, el hecho de que en 3 de 5 candidatos haya sido posible detectar alteraciones en la edición de transcritos mitocondriales en mutantes homocigotas del gen de la

proteína PPR indica que es un buen criterio. No podemos descartar que en el caso de los otros dos genes candidatos (At1g26900 y At1g77170), para los que no se detectaron defectos en la edición, haya redundancia funcional con otras proteínas PPR. Alternativamente las proteínas PPR AT1G26900 y AT1G77170 podrían ser necesarias para la edición de sitios no analizados por SNaPShot.

2.- Identificación de los factores de edición MEF25, MEF26 y MEF31

Una vez finalizada la etapa de selección de los genes candidatos, una etapa no menos importante fue la evaluación del nivel de edición organelar en mutantes homocigotas de ellos. Para llevar a cabo esta tarea se contó con el apoyo de los Dres. Mizuki Takenaka y Axel Brennicke del laboratorio de Botánica Molecular de la Universidad de Ulm en Alemania, quienes desarrollaron la aproximación de SNaPshot para analizar más de 300 sitios de edición (Takenaka y Brennicke, 2009). Esta metodología permite evaluar en un solo ensayo la edición de cientos de sitios a partir de mezclas complejas de RNAs provenientes de plantas mutantes en distintos genes PPR. Además no presenta dificultad para analizar secuencias de RNA con elevada densidad de sitios de edición, lo que representa una ventaja respecto al otro método de ensayo masivo de sitios de edición propuesto por Chateigner-Boutin y Small (2007). Este último, denominado “high-resolution melting”, demostró ser sólo efectivo para analizar sitios presentes en el transcriptoma cloroplástico, ya que este organelo presenta una reducida densidad de sitios de edición (31 sitios) (Chateigner-Boutin y cols., 2013).

Los defectos de edición en las mutantes *mef25*, *mef26* y *mef31* fueron confirmados por secuenciación de productos de RT-PCR. En el caso de MEF25 se caracterizaron dos mutantes insercionales, y se analizaron casi todos los sitios de edición mitocondriales (412) y todos los cloroplásticos, encontrándose que MEF25 es esencial solo para la edición del sitio 308 en el transcrito que codifica para la subunidad NAD1 del complejo I mitocondrial. En el caso de MEF26 se caracterizaron dos mutantes insercionales y plantas de una de las mutantes

complementadas con el gen *mef26* silvestre, y se analizaron 434 sitios mitocondriales y todos los cloroplásticos, encontrándose que la ausencia de MEF26 anula la edición del sitio 311 en el transcrito de la subunidad 3 (*cox3*) del complejo IV mitocondrial o citocromo oxidasa, y afecta parcialmente a la edición del sitio 166 del transcrito de la subunidad NAD4 del complejo (NADH deshidrogenasa). Finalmente, en el caso de MEF31 sólo se ha caracterizado una mutante insercional y se analizaron 269 sitios de edición, encontrándose que MEF31 es esencial para la edición del sitio 581 en el transcrito *orfX* y que podría estar implicado indirectamente en la edición del sitio 586 del mismo transcrito.

En resumen, se han identificado inequívocamente los factores de edición mitocondriales MEF25 y MEF26 y sus sitios diana, aunque no podemos descartar que participen en la edición de los pocos sitios no analizados. En el caso de MEF31 es necesario caracterizar otra mutante y/o complementar el mutante ya caracterizado con MEF31 silvestre para establecer fehacientemente su implicación en la edición del sitio *orfX*-581 (y 586), y es posible que participe en la edición de otros sitios no analizados (numerosos en este caso).

3.- La edición de los sitios diana de MEF25, MEF26 y MEF31 es necesaria para la incorporación de aminoácidos conservados en las proteínas pero no para el desarrollo de las plantas, al menos en condiciones estándar de crecimiento.

Los factores PPR descritos en este trabajo de tesis (MEF25, MEF26 y MEF31) son necesarios para la edición de uno o más sitios en distintos transcritos mitocondriales, y comparten el hecho de que la edición de sus sitios diana (*nad1-308*, *cox3-311*, *nad4-166* y *orfX-581*) genera un cambio en la naturaleza del aminoácido que finalmente estará presente en la proteína funcional. En el caso de *nad1-308*, la completa ausencia de edición en este sitio lleva a la incorporación de una serina en lugar de una leucina altamente conservada en la subunidad NAD1 del complejo respiratorio NADH deshidrogenasa. Ordenamientos tridimensionales propuestos

para este complejo sitúan a la subunidad NAD1 en la interfase entre el dominio hidrofóbico embebido en la membrana y el dominio expuesto directamente a la matriz mitocondrial del complejo I (Klodmann y Braun., 2011). No obstante, líneas *mef25* mutantes son fenotípicamente indistinguibles de las plantas silvestres. Esto podría sugerir que la presencia del residuo de serina en la proteína NAD1 es eficientemente compensada por el sistema, en otras palabras la edición del sitio *nad1-308* no es esencial para la función del complejo I mitocondrial, al menos bajo condiciones óptimas de crecimiento. Paralelamente, la ausencia de edición del sitio *cox3-311* y la reducción del nivel de edición del sitio *nad4-166* en las mutantes *mef26* generan por una parte una proteína COX3 que tiene una serina en lugar de una leucina altamente conservada, y una mezcla del polipéptido NAD4 con *trp56* (secuencia aminoacídica silvestre) y con *arg56* también, las mutantes *mef26* tienen un fenotipo de crecimiento normal, lo que sugiere que la falta de edición de *cox3-311* y la disminución en el estatus de edición de *nad4-166* no tendrían efecto en la función mitocondrial normal. La edición de ambos sitios no sería esencial al menos bajo condiciones de crecimiento estándares. Finalmente, la falta de edición del sitio *orfX-581* en la mutante *mef31* no genera alteraciones en el fenotipo de crecimiento de las plantas mutantes respecto de las plantas silvestres, sugiriendo con ello que la putativa función de transporte y translocación de polipéptidos por parte de la proteína tatC (ORFX o MTTB) no se ve afectada por el cambio de una leucina conservada. La edición del sitio *orfX-581* no es entonces esencial al menos en condiciones de crecimiento normales. Sin embargo, el hecho de que ninguna de las plantas mutantes en los factores MEFs anteriormente descritos presenten alteraciones en el crecimiento o desarrollo respecto de las plantas silvestres, al menos bajo condiciones de crecimiento óptimas, no descarta la posibilidad de que bajo ciertas condiciones de crecimiento o bajo condiciones de estrés ambiental, estas plantas mutantes puedan presentar un fenotipo de crecimiento alterado (Takenaka y cols., 2010).

Es posible que nuestra aproximación experimental haya introducido un sesgo, descartando al menos proteínas cuya ausencia sea letal (fenotipo extremo). En efecto, el método

de “SNaPshot” debe usarse con plantas homocigotas mutantes porque los defectos en la edición no son perceptibles en plantas heterocigotas.

La falta de fenotipos visibles ha sido observada en otras mutantes de genes de proteínas PPR descritas recientemente. En efecto, muy pocas de las 34 plantas mutantes en factores de edición de tipo PPR de Arabidopsis, presentan defectos severos en el crecimiento o letalidad prematura, aunque la falta de edición en estas mutantes sea completa y se trate de la modificación de un aminoácido altamente conservado (Hammani y Giegé, 2014). Por el contrario, mutantes de genes PPR en las que se ven afectados otros tipos de procesamiento post-transcripcional, tales como splicing, estabilización de transcritos o iniciación de la traducción, presentan alteraciones fenotípicas mayores. También plantas mutantes en las que se observa ausencia de la edición de varios sitios simultáneamente están mas afectadas o son inviables (Takenaka y cols., 2012; Bentolila y cols., 2012). Esto sugiere que la completa edición de un sitio o la conservación filogenética no son buenos indicadores de la importancia fisiológica de un sitio de edición (Chateigner-Boutin y cols., 2013). Por otra parte es posible pensar que en las mutantes los transcritos no editados son traducidos y que las proteínas generadas a partir de ellos serían estables. De esta forma, la falta de edición en cada uno de estos sitios sólo promovería sutiles efectos en la estructura o la actividad de los complejos mitocondriales. Futuros estudios serían necesarios para determinar fehacientemente bajo que condiciones las mutantes *mef25*, *mef26* y *mef31* podrían presentar fenotipos alterados de crecimiento. En este sentido sería interesante evaluar si la acumulación de alteraciones en la edición a través de la generación de plantas mutantes dobles en genes PPR específicos podrían evidenciar efectos fenotípicos visibles por combinaciones anómalas de aminoácidos en la misma proteína de un complejo mitocondrial. Algunos de los posibles candidatos para la generación de plantas mutantes dobles podrían ser: la mutante *mef25* con la mutante PPR *mef32*, necesaria para la edición de los transcritos *nad1*, *ccmB*, *cox2* (Takenaka y cols., 2013). En el caso de *mef26* un buen candidato serían las plantas

mutantes en *mef21* que tampoco presentan fenotipos de crecimiento alterado a pesar de que en esta mutante se ve afectada la edición de un sitio en el transcrito de *cox3* (Takenaka y cols., 2010)

Conclusiones Generales

- I. El uso de criterios de selección, herramientas genéticas y de un método de análisis masivo de sitios de edición permitió identificar tres nuevos Factores de Edición Mitochondrial: MEF25, MEF26 y MEF31.
- II. La edición de los sitios diana de MEF25, MEF26 y MEF31 es necesaria para la incorporación de aminoácidos conservados en subunidades de los complejos mitocondriales pero NO para el desarrollo de las plantas, al menos bajo condiciones estándares de crecimiento. Es posible que en condiciones menos ideales de crecimiento se observen fenotipos alterados
- III. No se puede descartar la posibilidad de que la falta de edición en sitios específicos genere alteraciones en la funcionalidad de algunos de los complejos mitocondriales y que paralelamente mecanismos compensatorios esten siendo activados, lo que impediría observar fenotipos severos en las plantas mutantes.
- IV. Los criterios de selección utilizados fueron efectivos para la identificación de nuevos “factores de especificidad” necesarios para la edición de transcritos mitocondriales los que mostraron ciertas particularidades:

MEF25 ➔ Fue el primer factor PPR descrito como necesario para la edición de nad1. Este factor edita sólo uno de dos sitios contiguos en un transcrito mitocondrial.

MEF26 ➔ Es factor representa a uno de los muy pocos ejemplos descritos de dos PPR que tienen el mismo sitio a editar como blanco.

MEF31 ➔ Es el segundo factor descrito hasta ahora como necesario para la edición de un sitio en el transcrito orfX.

- V. El “código de las PPR” propuesto para el reconocimiento de RNA necesita de refinamiento y de futuras mejoras, para predecir adecuadamente los dianas de los factores PPRs.

CAPÍTULO IV

REFERENCIAS BIBLIOGRÁFICAS

- Barkan A., Rojas M., Fuji S., Yap A., Chong Y.S., Bond C.S., Small I., 2012. A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. *PLoS Genet* 8, e1002910.
- Barkan A., Small I., 2014. Pentatricopeptide Repeat Proteins in Plants. *Annu Rev Plant Biol.* 65, 415-442.
- Bentolila S., Elliott L.E., Hanson M.R., 2008. Genetic architecture of mitochondrial editing in *Arabidopsis thaliana*. *Genetics* 178, 1693-1708.
- Bentolila S., Knight W., Hanson M., 2010. Natural variation in *Arabidopsis* leads to the identification of REME1, a pentatricopeptide repeat-DYW protein controlling the editing of mitochondrial transcripts. *Plant Physiol* 154, 1966-1982.
- Bentolila S., Heller W.P., Sun T., Babina A.M., Friso G., Van Wijk K.J., 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, E1453-1461.
- Binder S., Brennicke A., 2003. Gene expression in plant mitochondria: transcriptional and post-transcriptional control. *Philos Trans R Soc Lond B Biol Sci* 358, 181-188; discussion 188-189.
- Blanc V., Litvak S., Araya A., 1995. RNA editing in wheat mitochondria proceeds by a deamination mechanism. *FEBS Letters* 373, 56-60.
- Bogsch EG., Sargent F., Stanley NR., Berks BC., Robinson C., Palmer T. 1998. An essential component of a novel bacterial protein export system with homologues in plastids and mitochondria. *J Biol Chem* 273, 18003-18006.
- Cai W., Ji D., Peng L., Guo J., Ma J., Zou M., Lu C., Zhang L., 2009. LPA66 Is Required for Editing psbF Chloroplast Transcripts in *Arabidopsis*. *Plant Physiol* 150, 1260-1271.
- Chateigner-Boutin AL., Small I., 2007. A rapid high-throughput method for the detection and quantification of RNA editing based on high-resolution melting of amplicons. *Nucleic Acids Res* 35, e114
- Chateigner-Boutin, A.L., Ramos-Vega, M., Guevara-García, A., Andres, C., de la Luz Gutierrez-Nava, M., Cantero, A., Delannoy, E., Jimenez, L.F., Lurin, C., Small, I., Leon, P., 2008. CLB19, a pentatricopeptide repeat protein required for editing of rpoA and clpP chloroplast transcripts. *Plant J* 56, 590-602.
- Chateigner-Boutin AL., Colas des Francs-Small C., Fujii S., Okuda K., Tanz SK., Small I., 2013. The E domains of pentatricopeptide repeat proteins from different organelles are not functionally equivalent for RNA editing. *Plant J* 74, 935-945.
- Choury D., Farré J.C., Jordana X., Araya A., 2004. Different patterns in the recognition of editing sites in plant mitochondria. *Nucleic Acids Res* 32, 6397-6406.
- Clough S., Bent A., 1998. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J* 16, 735-743.
- Covello P.S., Gray M.W., 1989. RNA editing in plant mitochondria. *Nature* 341, 662-666.

- Ding Y.H., Liu N.Y., Tang Z.S., Liu J., Yang W.C., 2006. Arabidopsis GLUTAMINE-RICH PROTEIN23 is essential for early embryogenesis and encodes a novel nuclear PPR motif protein that interacts with RNA polymerase II subunit III. *Plant Cell* 18, 815-830.
- Farré J.C., Araya A., 2001. Gene expression in isolated plant mitochondria: High fidelity of transcription, splicing and editing of a transgene product in electroporated organelles. *Nucleic Acids Res* 29, 2484-2491.
- Fujii S., Small I., 2011. The evolution of RNA editing and pentatricopeptide repeat genes. *New Phytologist* 191, 37-47.
- Gagliardi D & Binder S (2007) Expression of the Plant Mitochondrial Genome. D. Logan, Ed. In *Annual Plant Reviews Volume 31: Plant Mitochondria*, pp. 50-96. Blackwell Publishing. Ud, UK.
- García-Andrade J, Ramirez V, Lopez A & Vera P (2013) Mediated plastid RNA editing in plant immunity. *PLoS Pathog* 9, e1003713, doi: 10.1371/journal.ppat.1003713
- Giegé P., Brennicke A., 1999. RNA editing in Arabidopsis mitochondria effects 441 C to U changes in ORFs. *Proc. Natl. Acad. Sci. U S A* 96, 15324-15329.
- Hammani K., Okuda K., Tanz S.K., Chateigner-Boutin A.L., Shikanai T., Small I., 2009. A study of new Arabidopsis chloroplast RNA editing mutants reveals general features of editing factors and their target sites. *Plant Cell* 21, 3686-3699.
- Hammani K., Gobert A., Hleibieh K., Choulier L., Small I., Giegé P., 2011. An Arabidopsis dual-localized pentatricopeptide repeat protein interacts with nuclear proteins involved in gene expression regulation. *Plant Cell* 23, 730-740.
- Hammani K., des Francs-Small CC., Takenaka M., Tanz S., Okuda K., Shikanai T., Brennicke A., Small I., 2011. The pentatricopeptide repeat protein OTP87 is essential for RNA editing of nad7 and atp1 transcripts in Arabidopsis mitochondria. *J Biol Chem* 286, 21361-21371.
- Hammani K., Giegé P., 2014. RNA metabolism in plant mitochondria. *Trends Plant Sci* 19, 380-389.
- Härtel B., Zehrmann A., Verbitskiy D., Takenaka M., 2013. The longest mitochondrial RNA editing PPR protein MEF12 in Arabidopsis thaliana requires the full-length E domain. *RNA Biol* 10, 1543-1548.
- Hiesel R., Wissinger B., Schuster W., Brennicke A., 1989. RNA editing in plant mitochondria. *Science* 246, 1632-1634.
- Jaillon O., Aury J.M., Noel B., Policriti A., Scarpelli C., Adam-Blondon A.F., Weissenbach J., Quetier F., Wincker P., 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449, 463-467.
- Kelley L., Mezulis S., Yates C., Wass M., Sternberg M., 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protocols* 10, 845-858.

- Kim S.R., Yang J.I., Moon S., Ryu C.H., An K., Kim K.M., Yim J., An G., 2009. Rice OGR1 encodes a pentatricopeptide repeat-DYW protein and is essential for RNA editing in mitochondria. *Plant J* 59, 738-49.
- Kindgren P., Yap A., Bond C.S., Small I., 2015. Predictable alteration of sequence recognition by RNA editing factors from Arabidopsis. *Plant Cell* 27, 403-416.
- Klodmann J., Braun H.P., 2011, Proteomic approach to characterize mitochondrial complex I from plants. *Phytochemistry* 72, 1071-1080.
- Kotera E., Tasaka M., Shikanai T., 2005. A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. *Nature* 433, 326-330.
- Lurin C., Andres C., Aubourg S., Bellaoui M., Bitton F., Bruyere C., Caboche M., Debast C., Gualberto J., Hoffmann B., Lecharny A., Le Ret M., Martin-Magniette M.L., Mireau H., Peeters N., Renou J.P., Szurek B., Taconnat L., Small I., 2004. Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell* 16, 2089-2103.
- Malek O., Lattig K., Hiesel R., Brennicke A., Knoop V., 1996. RNA editing in bryophytes and a molecular phylogeny of land plants. *EMBO J* 15, 1403-1411.
- Marienfeld J., Unseld M., Brennicke A. 1999. The mitochondrial genome of Arabidopsis is composed of both native and immigrant information. *Tr Plant Sci* 4, 495-502.
- Millar H., Heazlewood J.L., Kristensen B.K., Braun H.P., Møller I.M., 2003. The plant mitochondrial proteome. *Tr Plant Sci* 10, 1360-1385.
- Murayama M., Hayashi S., Nishimura N., Ishide M., Kobayashi K., Yagi Y., Asami T., Nakamura T., Shinozaki K., Hirayama T., 2012. Isolation of Arabidopsis ahg11, a weak ABA hypersensitive mutant defective in nad4 RNA editing. *J Exp Bot* 63, 5301-5310.
- Nakai K., 2000. Protein sorting signals and prediction of subcellular localization. *Adv. Protein Chem.*, 54, 277-344.
- Ohtani S., Ichinose M., Tasaki E., Aoki Y., Komura Y., Sugita M., 2010. Targeted gene disruption identifies three PPR-DYW proteins involved in RNA editing for five editing sites of the moss mitochondrial transcripts. *Plant Cell Physiol* 51, 1942-1949.
- Okuda K., Nakamura T., Sugita M., Shimizu T., Shikanai T., 2006. A pentatricopeptide repeat protein is a site recognition factor in chloroplast RNA editing. *J Biol Chem* 281, 37661–37667.
- Okuda K., Myouga F., Motohashi R., Shinozaki K. Shikanai T., 2007. Conserved domain structure of pentatricopeptide repeat proteins involved in chloroplast RNA editing. *Proc Natl Acad Sci U S A* 104, 8178-8183.
- Okuda K., Habata Y., Kobayashi Y., Shikanai T., 2008. Amino acid sequence variations in nicotiana CRR4 orthologs determine the species-specific efficiency of RNA editing in plastids. *Nucleic Acids Res* 36, 6155-6164.

- Okuda K., Chateigner-Boutin A.L., Nakamura T., Delannoy E., Sugita M., Myouga F., Motohashi R., Shinozaki K., Small I., 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* 21, 146-156.
- Okuda K., Shikanai T. 2012. A pentatricopeptide repeat protein acts as a site- specificity factor at multiple RNA editing sites with unrelated cis-acting elements in plastids. *Nucleic Acids Res* 40, 5052–5064.
- Okuda K., Shoki H., Arai M., Shikanai T., Small I., Nakamura T., 2014. Quantitative analysis of motifs contributing to the interaction between PLS-subfamily members and their target RNA sequences in plastid RNA editing. *Plant J* 80, 870-882.
- O'Toole N., Hattori M., Andres C., Iida K., Lurin C., Schmitz-Linneweber C., Sugita M., Small I., 2008. On the expansion of the pentatricopeptide repeat gene family in plants. *Mol. Biol. Evol.* 25, 1120–1128.
- Rivals E., Bruyere C., Toffano-Nioche C., Lecharny A., 2006. Formation of the *Arabidopsis* pentatricopeptide repeat family. *Plant Physiol* 141, 825–839.
- Robbins J., Heller, W., Hanson, M., 2009. A comparative genomics approach identifies a PPR-DYW protein that is essential for C-to-U editing of the *Arabidopsis* chloroplast *accD* transcript. *RNA* 15, 1142-1153.
- Rudinger M., Polsakiewicz M., Knoop V., 2008. Organellar RNA editing and plant-specific extensions of pentatricopeptide repeat proteins in jungermanniid but not in marchantiid liverworts. *Mol Biol Evol* 25, 1405-1414.
- Rudinger M., Fritz-Laylin L., Polsakiewicz M., Knoop V., 2011. Plant-type mitochondrial RNA editing in the protist *Naegleria gruberi*. *RNA* 17, 2058-2062.
- Ruwe H., Schmitz-Linneweber C., 2012. Short non-coding RNA fragments accumulating in chloroplasts: footprints of RNA binding proteins? *Nucleic Acids Res.* 40, 3106–3116.
- Salone V., Rudinger M., Polsakiewicz M., Hoffmann B., Groth-Malonek M., Szurek B., Small I., Knoop V., Lurin C., 2007. A hypothesis on the identification of the editing enzyme in plant organelles. *FEBS Lett* 581, 4132-4138.
- Sew YS., Stroher E., Holzmann C., Huang S., Taylor NL., Jordana X., Millar AH., 2013. Multiplex micro-respiratory measurements of *Arabidopsis* tissues. *New Phytol* 200, 922-932.
- Schmitz-Linneweber C., Small I., 2008. Pentatricopeptide repeat proteins: a socket set for organelle gene expression. *Tr Plant Sci* 13, 663-670.
- Small I.D., Peeters N., 2000. The PPR motif A TPR-related motif prevalent in plant organellar proteins. *Tr Biochem Sci* 25, 46–47.
- Sung T., Tseng C., Hsieh M., 2010. The SLO1 PPR protein is required for RNA editing at multiple sites with similar upstream sequences in *Arabidopsis* mitochondria. *Plant J* 63, 499-511.

Takenaka M., Brennicke A., 2009. Multiplex single-base extension typing to identify nuclear genes required for RNA editing in plant organelles. *Nucleic Acids Res* 37, e13.

Takenaka M., Neuwirt J., Brennicke A., 2004. Complex cis-elements determine an RNA editing site in pea mitochondria. *Nucleic Acids Res* 32, 4137-4144.

Takenaka M., Verbitskiy D., van der Merwe J.A., Zehrmann A., Brennicke A., 2008. The process of RNA editing in plant mitochondria. *Mitochondrion* 8, 35-46.

Takenaka M., Verbitskiy D., Zehrmann A., Brennicke A., 2010. Reverse genetic screening identifies five E-class PPR proteins involved in RNA editing in mitochondria of *Arabidopsis thaliana*. *J Biol Chem* 285, 27122-27129.

Takenaka M., Zehrmann A., Brennicke A., Graichen K., 2013. Improved computational target site prediction for pentatricopeptide repeat RNA editing factors. *PLoS One* 8, e65343.

Takenaka M., Zehrmann A., Verbitskiy D., Kugelman M., Härtel B., 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.

Tasaki E., Hattori M., Sugita M., 2010. The moss pentatricopeptide repeat protein with a DYW domain is responsible for RNA editing of mitochondrial *ccmFc* transcript. *Plant J* 62, 560-570.

Tuskan G.A., Difazio S., Jansson S., Douglas C., Marra M., Sandberg G., Van de Peer Y., Rokhsar D., 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray) *Science* 313, 1596-1604.

Uchida M., Ohtani S., Ichinose M., Sugita C., Sugita M., 2011. The PPR-DYW proteins are required for RNA editing of *rps14*, *cox1* and *nad5* transcripts in *Physcomitrella patens* mitochondria. *FEBS Lett* 585, 2367-2371.

Unseld M., Marienfeld JR., Brandt P. & Brennicke A. 1997 The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. *Nat Genet* 15, 57-61

van der Merwe, J.A. and Dubery, I.A. 2007. Expression of mitochondrial *tatC* in *Nicotiana tabacum* is responsive to benzothiadiazole and salicylic acid. *J Plant Physiol* 164, 1231-1234.

Verbitskiy D., Zehrmann A., Brennicke A. & Takenaka M. 2010. A truncated MEF11 protein shows site-specific effects on mitochondrial RNA editing. *Plant Signal Behav* 5, 558-560.

Verbitskiy D., Takenaka M., Neuwirt J., van der Merwe JA., Brennicke A., 2006. Partially edited RNAs are intermediates of RNA editing in plant mitochondria. *Plant J* 47, 408-416.

Verbitskiy D., Zehrmann A., Hartel B., Brennicke A., Takenaka M., 2012. Two related RNA-editing proteins target the same sites in mitochondria of *Arabidopsis thaliana*. *J Biol Chem* 287, 38064-38072.

Welchen E., Klodmann J., Braun HP., 2010. Biogenesis and supramolecular organization of the oxidative phosphorylation system in plants. F. Kempken. "Plant mitochondria", *Advances in plant biology* 1 Chapter 13 (pp. 327-329). Kiel: Springer

Wu F., Shen S., Lee L., Lee S., Chan M., Lin C., 2009. Tape-Arabidopsis Sandwich - a simpler Arabidopsis protoplast isolation method. *Plant Methods* 5, 16, doi: 1746-4811-5-16

Yu Q.B., Jiang Y., Chong K., Yang Z.N., 2009. AtECB2, a pentatricopeptide repeat protein, is required for chloroplast transcript accD RNA editing and early chloroplast biogenesis in *Arabidopsis thaliana*. *Plant J.* 59, 1011-1023.

Yu W., and Schuster W., 1995. Evidence for a site-specific cytidine deamination reaction involved in C to U RNA editing of plant mitochondria. *J Biol Chem* 270, 18227-18233.

Zehrmann A., van der Merwe J.A., Verbitskiy D., Brennicke A., Takenaka M., 2008. Seven large variations in the extent of RNA editing in plant mitochondria between three ecotypes of *Arabidopsis thaliana*. *Mitochondrion* 8, 319-327.

Zehrmann A., Verbitskiy D., van der Merwe J.A., Brennicke A., Takenaka M., 2009. A DYW Domain-Containing Pentatricopeptide Repeat Protein Is Required for RNA Editing at Multiple Sites in Mitochondria of *Arabidopsis thaliana*. *Plant Cell* 21, 558-567.

Zhelyazkova P., Hammani K., Rojas M., Voelker R., Vargas-Suarez M., et al. 2012. Protein-mediated protection as the predominant mechanism for defining processed mRNA termini in land plant chloroplasts. *Nucleic Acids Res* 40, 3092-3105.

Zhou W., Cheng Y., Yap A., Chateigner-Boutin A.L., Delannoy E., Hammani K., Small I., Huang J., 2008. The Arabidopsis gene YS1 encoding a DYW protein is required for editing of *rpoB* transcripts and the rapid development of chloroplasts during early growth. *Plant J.* 58, 82-96

Zhu Q., E HM., Van Der Straeten D., 2012. Functional analysis of SLO2 provides new insight into the role of plant PPR proteins. *Plant Signal Behav* 7, doi: 21430