Universitat de Lleida

TESI DOCTORAL

TITLE

REDESIGNING RICE FOR ENVIRONMENTAL SUSTAINABILITY, ENHANCED PRODUCTIVITY AND SPECIALTY USES

Can Baysal

Memòria presentada per optar al grau de Doctor per la Universitat de Lleida
Programa de Doctorat en Ciència i Tecnologia Agrària i Alimentària

Director/s
Paul Christou
Changfu Zhu

Front Cover: Rice Terraces in Yunan, China: an interpretation (V. Medina)

Supervisors: **Paul Christou**

Changfu Zhu

Department of Plant Production and Forest Science

School of Agricultural and Forestry Engineering

University of Lleida

Signatures P. Christou C. Zhu

DEDICATION

To my beloved family

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisors Dr. Paul Christou and Dr. Changfu Zhu for their continuous support, great enthusiasm, patience and immense knowledge. It has been an absolute pleasure to have their guidance during my research. Very special thanks to Dr. Teresa Capell for making me a member of this great family for her constant feedback, her limitless motivation and energy. Without all their encouragement this PhD thesis would not has been achievable.

Besides my advisors, my sincere thanks also go to Dr. Luis Rubio ("Universidad Politécnica de Madrid", "Centro de Biotecnología y Genómica de Plantas (CBGP)" – "Instituto Nacional de Investigaciones Agrarias", INIA) who provided me an opportunity to work with his team, and who gave me access to laboratory and research facilities. It was a great pleasure to learn from him and all his lab members especially Dr. Stefan Burén. Thanks for the brilliant work, dedication and patience. I would also like to thank Dr. Paul Fraser (The Royal Holloway, University of London), for accepting me to his laboratory and making available facilities for metabolomic analysis and Dr. Margit Drapal for her kindness and her contribution to my research.

I am grateful to "Agència de Gestió d`Ajuts Universitaris i de Recerca" (AGAUR, Spain) and "Centre de Recerca en Agrotecnologia" (AGROTECNIO) for financial support. Also, thanks to The Bill and Melinda Gates Foundation, Ministry of Economy and Competitiveness of Spain and Generalitat de Catalunya for their financial support. Thanks to the University of Lleida, especially to the "Departament de Producció Vegetal i Ciència Forestal", for the use of labs and infrastructure.

I would like to thank Dr. Ludovic Bassie for his assistance, technical advice and the stimulating discussions. Special thanks to Nuria Gabernet for organizing the official paperwork and Jaume Capell for all his effort to care of my plants.

I would like to thank all my present and ex-colleagues; Ravi, Gemma F, Gemma M, Daniela, Marlies, Giobbe, Sarah, Lucia, Erika, Xin J, Andrew, Marco, Amaya, Vicky, Sarai, Maria, Andrea, Hugo, Jaume, Jose, Derry, Pedro, Guillermo, Ashwin, Wenshu and Xin H. for their friendship, assistance and encouragement throughout my degree. The social events we attended together including numerous celebrating dinners and nights out are memorable. Specially Sun Yi, words are powerless to express my feelings but I would like to say thank you so much for all your effort and unforgettable friendship "rest in peace my dear friend". All your presence has had a great impact on my success for which I will always be grateful.

Last but not the least, I would like to thank my mother, father, brother, my cute nephew and to you Vicente for being my family here, sharing your home, for all your support and encouragement. All my love and thanks for everything.

SUMMARY

My thesis focuses on the optimization of multi-gene transformation and genome-editing techniques to modify and analyze important agronomical traits and complex metabolic pathways in rice.

I targeted nuclear-encoded green fluorescent protein (eGFP) to rice mitochondria using six mitochondrial pre-sequences with diverse phylogenetic origin. I investigated their effectiveness by immunoblot analysis, confocal and electron microscopy in callus and regenerated plants. I confirmed that COX4, SU9 targeting peptides effectively targeted eGFP to rice mitochondria regardless of their origin. I used this methodology to dissect and express extremely oxygen sensitive nitrogenase Fe Protein NifH and its maturase NifM in rice mitochondria to overcome the most challenging bottleneck of engineering biological nitrogen fixation in cereals. Therefore, I was able to express and purify substantial amounts of NifH from rice callus, and I confirmed that the isolated NifH carried out the fundamental role of NifH enzymatic activity required to engineer nitrogen fixation, including electron transfer, and FeMo-co biosynthesis. I was thus able to demonstrate for the first-time stable expression and activity of a nitrogenase component in any cereal.

I have further used CRISPR/Cas9 to mutate the *OsSBEIIb* (encoding starch branching enzyme IIb), which is required for the synthesis of densely branched amylopectin in the rice endosperm. I investigated the effects of the inactivation of individual starch biosynthesis enzyme in rice starch and overall metabolism. Homozygous mutant plants, in which OsSBEIIb was completely inactivated, produced opaque seeds with depleted starch reserves whose amylose content increased from 19.6 to 27.4% and resistant starch (RS) content increased from 0.2 to 17.2%. Because of the mutation there was a general increase in the accumulation of sugars, fatty acids, amino acids, and phytosterols in the mutant endosperm, suggesting that intermediates in the starch biosynthesis pathway increased flux through spillover pathways causing a profound impact on the accumulation of multiple primary and secondary metabolites.

RESUMEN

Mi tesis se centra en la optimización de la transformación múltiple de genes y las técnicas de edición del genoma para modificar y analizar rasgos agronómicos importantes y vías metabólicas complejas en el arroz.

Conseguí dirigir la proteína verde fluorescente codificada en el núcleo (eGFP) a las mitocondrias de arroz utilizando seis pre-secuencias mitocondriales con diverso orígen filogenético. Investigué su efectividad mediante análisis de inmunotransferencia, microscopía confocal y electrónica tanto en callos como en plantas regeneradas. Confirmé que los péptidos COX4, SU9 dirigían eficazmente eGFP a las mitocondrias de arroz independientemente de su origen. Utilicé esta metodología para diseccionar y expresar la proteína NifH nitrogenasa Fe, extremadamente sensible al oxígeno, y su maturasa NifM en las mitocondrias de arroz, superando así el cuello de botella más desafiante de la ingeniería de la fijación biológica de nitrógeno en los cereales. Por lo tanto, pude expresar y purificar cantidades sustanciales de NifH a partir de callos de arroz y confirmé que la NifH aislada desempeñaba el papel fundamental de la actividad emzimática de NifH necesaria para diseñar la fijación de nitrógeno, incluida la transferencia de electrones y la biosíntesis de FeMo-co. De esta manera pude demostrar, por primera vez, la expresión y actividad estables de un componente nitrogenasa en cualquier cereal.

También usé CRISPR/Cas9 para mutar el *OsSBEIIb* (que codifica la enzima ramificadora de almidón IIb), que se requiere para la síntesis de amilopectina densamente ramificada en el endospermo del arroz. Investigué los efectos de la inactivación de la enzima de biosíntesis de almidón individual en el almidón de arroz y en su metabolismo general. Las plantas mutantes homocigotas, en las que OsSBEIIb estaba completamente inactivado, produjeron semillas opacas con reservas de almidón agotadas cuyo contenido de amilosa aumentó de 19.6 a 27.4% y el contenido de almidón resistente (RS) aumentó de 0.2 a 17.2%. Debido a la mutación, hubo un aumento general en la acumulación de azúcares, ácidos grasos, aminoácidos y fitoesteroles en el endospermo mutante, lo que sugiere que los intermediarios en la vía de biosíntesis del almidón aumentaron el flujo a través de vías indirectas causando un profundo impacto en la acumulación de múltiples metabolitos primarios y secundarios.

RESUM

La meva tesi se centra en l'optimització de la transformació múltiple de gens i les tècniques d'edició del'genoma per modificar i analitzar trets agronòmics importants i vies metabòliques complexes en l'arròs.

Vaig dirigir la proteïna verda fluorescent codificada en el nucli (eGFP) a les mitocòndries d'arròs utilitzant sis pre-seqüències mitocondrials amb divers origen filogenètic. Vaig investigar la seva efectivitat mitjançant anàlisi de immunotransferència, microscòpia confocal i electrònica tant en calls com en plantes regenerades. Vaig confirmar que els pèptids COX4, SEU9 dirigien eficaçment eGFP a les mitocòndries d'arròs independentment del seu origen. Vaig utilitzar aquesta metodologia per disseccionar i expressar la proteïna NifH nitrogenasa Fe, extremadament sensible a l'oxigen, i la seva maturasa NifM en les mitocòndries d'arròs, superant així el coll d'ampolla més desafiant de l'enginyeria de la fixació biològica de nitrogen en els cereals. Per tant, vaig poder expressar i purificar quantitats substancials de NifH a partir de calls d'arròs i vaig confirmar que la NifH aïllada exercia el paper fonamental de l'activitat emzimática de NifH necessària per dissenyar la fixació de nitrogen, inclosa la transferència d'electrons i la biosíntesi de FeMo-co. D'aquesta manera demostre, per primera vegada, l'expressió i activitat estables d'un component nitrogenasa en qualsevol cereal.

També vaig fer servir CRISPR/CAS9 per mutar el *OsSBEIIb* (que codifica l'enzim ramificadora de midó IIb), que es requereix per a la síntesi d'amilopectina densament ramificada en l'endosperma de l'arròs. Vaig investigar els efectes de la inactivació de l'enzim de biosíntesi de midó individual en el midó d'arròs i en el seu metabolisme general. Les plantes mutants homocigotas, en què OsSBEIIb estava completament inactivat, van produir llavors opaques amb reserves de midó esgotades el contingut d'amilosa va augmentar de 19.6 a 27.4% i el contingut de midó resistent (RS) va augmentar de 0.2 a 17.2%. A causa de la mutació, hi va haver un augment general en l'acumulació de sucres, àcids grassos, aminoàcids i fitosterols en l'endosperma mutant, el que suggereix que els intermediaris en la via de biosíntesi del midó van augmentar el flux a través de vies indirectes causant un profund impacte en l'acumulació de múltiples metabòlits primaris i secundaris.

Table of contents

INDEX OF FIGURES

Chapter 1

Chapter 2

Chapter 3

Chapter 4

INDEX OF TABLES

Chapter 2

C**hapter 3**

Chapter 4

ABBREVIATIONS & ACRONYMS

ACC65I	Restriction endonuclease from *Acinetobacter calcoaceticus* 65I
AGAUR	Agència de Gestió d'Ajuts Universitaris i de Recerca
AGPase	ADP-glucose pyrophosphorylase
ARA	Acetylene reduction assay
Ar	Argon
ATP	Adenosine triphosphate
ATPA	ATP synthase protein
*Av*NifM	*Azotobacter vinelandii* NifM
BamHI	Type II restriction endonuclease from *Bacillus amyloliquefaciens*
BCIP/NBT	5-bromo-4-chloro-3-indolyl-phosphate/ nitro blue tetrazolium
BNF	Biological N_2 fixation
BSA	Bovine serum albumin
CBGP	Centro de Biotecnología y Genómica de Plantas
CFE	Cell-free extract
COX4	Cytochrome c oxidase subunit 4
CRISPR/Cas9	Clustered Regularly Interspaced Short Palindromic Repeats/Cas9 enzyme
DBE	Debranching enzymes
DNA	Deoxyribonucleic acid
DTH	Sodium dithionite
DTT	Dithiothreitol
EDTA	Ethylene diamine tetraacetic acid
eGFP	Enhanced green fluorescent protein
ER	Endoplasmic reticulum
ETSEA	Escola Tècnica Superior d'Enginyeria Agrària
Fe	Iron
FeMo-co	Iron molybdenum cofactor
GBSS	Granule-bound starch synthase
GC-MS	Gas chromatography–mass spectrometry
GFP	Green fluorescent protein
GST	Glutathione S-transferase
HCl	Hydrochloric acid
hpt	Hygromycin phosphotransferase
*Ht*NifH	*Hydrogenobacter thermophilus* NifH
IgG	Immunoglobulin
INIA	Instituto Nacional de Investigaciones Agrarias
IPTG	Isopropyl β- d-1-thiogalactopyranoside
LB	Lysogeny broth
LSD	Least significant difference
MEP	2-C-methyl-Derythritol 4-phosphate
Mo	Molybdenum
MoFe	Molybdenum Iron
mRNA	Messenger ribonucleic acid
MS	Murashige and Skoog
MTS2	Mitochondrial targeting sequence
N	Nitrogen

NaCl	Sodium chloride
NADPH	Nicotinamide adenine dinucleotide phosphate
*Nb*NifH^{Kp}	*Klebsiella oxytaca NifH* expressed in *Nicotiana benthamiana*
NCBI	National Center for Biotechnology Information
NH_3	Ammonia
Nif / *nif*	Nitrogen fixation proteins and genes
NifB	Nitrogenase cofactor biosynthesis protein
NifDK	MoFe protein protein (encoded by *nifD* and *nifK*) dinitrogenase
NifEN	Scaffold Nif protein
NifH	Fe protein (*nifH*-encoded) dinitrogenase reductase
NifHDK	Nitrogenase protein complex
NifM	Nif maturase (putative peptidyl-prolyl isomerase)
NifS	Scaffold Nif protein (cysteine desulfurase)
NifU	[Fe-S] cluster scaffold protein
NO_3-	Nitrate
*Os*NifH^{Ht}	*Hydrogenobacter thermophilus* NifH expressed in *Oryza sativa*
*Os*NifM^{Av}	*Azotobacter vinelandii* NifM expressed in *Oryza sativa*
*Os*SCSb	*Oryza sativa* ssp. *japonica* succinyl-CoA synthetase protein β-chain
PCA	Principal Component Analysis
PC1	First principal component
PC2	Second principal component
PEG	Polyethylene glycol
PCR	Polymerase chain reaction
Pho	Phosphorylase
PIC	Protease inhibitor cocktail
Pfa	*Arabidopsis thaliana* c subunit of the mitochondrial ATP synthase
PMSF	Phenyl methane sulfonyl fluoride
PstI	Restriction endonuclease from *Providencia stuartii*
PVDF	Poly vinylidene difluoride
PUL	Pullulanase
pUC57	GenScript standard vector for synthetic genes
qRT-PCR	Quantitative real-time PCR
RBS1	Rubisco small chain 1
RCA	Rubisco activase
RNA	Ribonucleic acid
RS	Resistant starch
RT-PCR	Reverse transcription polymerase chain reaction
SAM	S-adenosylmethionine
SBE	Starch branching enzyme
*Sc*NifH^{Ht}	*Hydrogenobacter thermophilus* NifH expressed in *Saccharomyces cerevisiae*
SD	Standard deviation
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEM	Standard error mean
sGFP	Synthetic green fluorescent protein
SRP	Signal recognition particle
SS	Soluble starch synthases

STAC	Strep-tag affinity chromatography
SU9	*Neurospora crassa* subunit 9 mitochondrial ATPase signal
TBS-T	Tris-buffered saline with Tween-20
TOM20	Mitochondrial import receptor subunit TOM20.
TP_S	Transit peptides
UdL	Universitat de Lleida
UPLC	Ultra Performance Liquid Chromatography
V	Vanadium
ZmUbi1	*Zea mays* ubiquitin-1

Chapter 1

General Introduction

1.1. GENERAL INTRODUCTION

1.1.1. Global importance of rice as a food security crop

Rice is one of the most important food crops and the staple food of more than half the global population. Worldwide, over 3.5 billion people depend on rice for approximately 20% of their daily calorie intake (Gnanamanickam, 2009). Globally, rice provides the major source of nutritional calories (23% per capita energy and 16% per capita protein), particularly in East and Southeast Asia (Zeigler and Barclay, 2008). Approx. 95% of rice grown is consumed as an unprocessed food; only 5% is used in processed food, industrial products and alcoholic beverages (Alcázar-Alay and Meireles, 2011). To emphasize the importance of rice in human life and food security, the United Nations declared the year 2004 as "The International Year of Rice" and underlined the role of rice in alleviating poverty and malnutrition. During the global food chrisis of 2007-2008, rice prices increased in Asia and millions of Asians struggled to afford their staple food (Gnanamanickam, 2009).

1.1.1.1. *Origin, evolution and distribution*

About 140,000 varieties of cultivated rice are thought to exist and more than 90,000 samples of cultivated rice and wild species are stored in the International Rice Gene bank at the International Rice Research Institute (IRRI, The Philipines) for use by researchers around the world. The genus *Oryza sativa* L. which is probably originated at least 130 million years ago and spread as a wild grass in Gondwanaland, the super continent that eventually broke up and drifted apart to become Asia, Africa, the Americas and Australia. Today's species of the genus *Oryza sativa* L. are distributed in all of these continents except Antarctica (Khush, 1997).

Cultivated rice *Oryza sativa* L. has three sub-species, which are recognized as distinct ecotypes: indica, japonica and javanica (Garris *et al*. 2005). Indica rice, accounting for approximately 80% of cultivated rice; it is known as tropical long grain rice and it has a higher amylose content than japonica and javanica sub-species (Ayres and Park, 1994). The bulk of indica rice is grown in the tropics and subtropics including the Indian sub-continent, Southern China, and South America. In contrast, japonica rice has short or round grains, lower amylose content and is more adapted to temperate climates such as

Japan, Southern Europe, Central and Northern China. Javanicas are primarily grown in Java and Bali, in the rice terraces of the Philippines and the mountainous region of Madagascar (Ikehashi, 2014).

1.1.1.2. *Economic impact of rice cultivation*

More than 90% of all rice is grown and consumed in Asia where 60% of the earth's people live. Rice is the grain that has shaped the cultures, diets, and economies of hundreds of millions of Asians (FAO, 2019). Today, 1.6 billion ha is currently in use for cultivation of agricultural crops and rice is produced on about 10% of all cropland (160 million hectares) in over 117 countries, and ranks second only after maize in hectares harvested (Statista, 2020). The major rice-producing countries are illustrated in **Figure 1.1.**

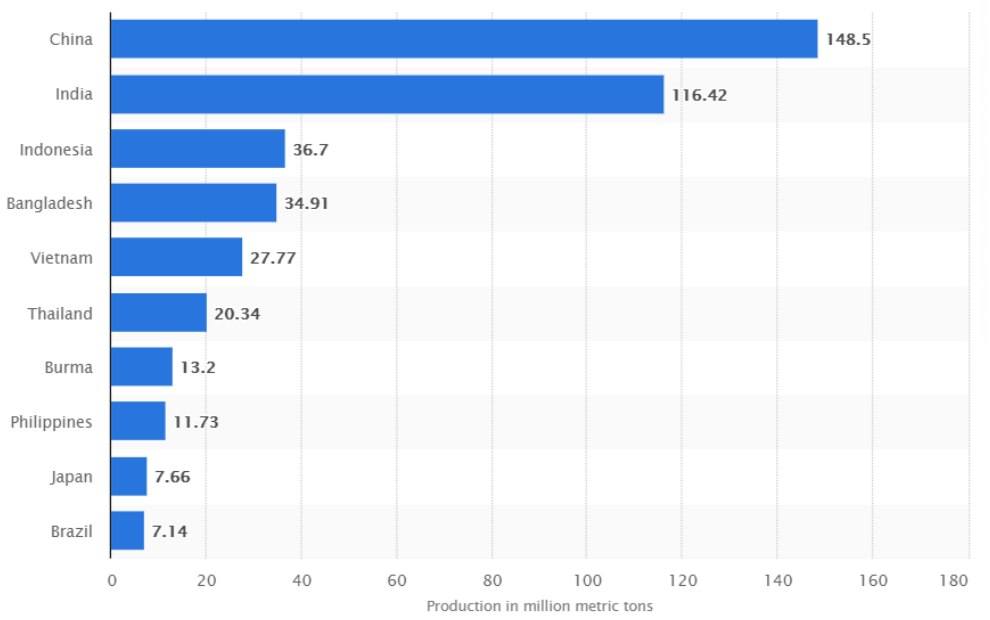


Figure 1.1. Top 10 rice producing countries. Source: Statista 2020.

For the majority of Asians, the total calorific intake is ca: 2,500 calories per person per day, with 35% coming from rice (Zeigler and Barclay, 2008). Rice contains approximately 10% protein, 75% available carbohydrates, 2% fat. In contrast, rice grain is relatively low in essential micronutrients such as iron (Fe), zinc (Zn) and calcium (Ca) compared to wheat or maize (Hoogenkamp *et al*., 2015). Rice bran is also a good source for various vitamins (e.g., thiamine, niacin, and riboflavin), antioxidants, and phytosterols (Prasad *et al*., 2011; Hoogenkamp *et al*., 2015).

Therefore, rice has become one of the most important sources of human food, as indicated by the increasing portion of the total arable land occupied by the crop. However, with accelerating loss of productive rice land to rising sea levels and salinity, rice supply cannot meet with the demand. If we are to provide food for the predicted world population of 10 million people by 2050, we need to achieve substantial increases in the yield of rice and other staple food security crops (Albajes *et al*., 2013; Berman *et al.*, 2013). To achieve this goal conventional breeding has to be aided by additional techologies, including molecular genetics and biotechnology.

1.1.1.3. *Biotechnological strategies for the improvement of agronomic traits in rice*

There is an urgent need of new varieties with improved agronomic characteristics, such as tolerance to different biotic and abiotic stresses, higher yield and better nutritional value (Berman *et al*., 2013). Discovery of genomic regions of key traits, over-expression of heterologous genes or down-regulation of endogenous genes are major components of crop improvement. Cloning and isolation of heterologous genes facilitated crop improvement programs through genetic transformation of elite varieties (Farré *et al*., 2015). One of the most succesful examples is the expression of *Bacillus thuringiensis* "*Cry*" gene in maize, potato, eggplant, soybean and cotton (Van Frankenhuyzen *et al*., 1997; Letourneau *et al*., 2002; Tabashnik *et al*., 2002; Koch *et al*., 2015).

Several strategies involving engineered nucleases are available to induce targeted mutations, namely zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and recently the clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein (Cas) system. All above engineered nucleases are typically designed to introduce a double-strand break (DSB) in the coding sequence of a target gene so that repair by the endogenous non-homologous end joining (NHEJ) pathway results in the introduction of small insertions or deletions (indels). However, the CRISPR/Cas system is more versatile because it offers major improvements in efficiency and accuracy compared with earlier methods based on site-specific recombinases or nuclease-free homologous recombination (Baysal *et al*., 2016; Zhu *et al.*, 2017). The Cas endonuclease is guided to its target by DNA–RNA pairing, which increases the diversity of available target sites and simplifies the targeting strategy (Jinek *et al*. 2012; Bortesi and Fischer, 2015). In the most widely used CRISPR/Cas system, the

Cas9 endonuclease from *Streptococcus pyogenes* is combined with a single or multiple synthetic guide RNA (sgRNA) that provides targeting specificity (Zhu *et al*., 2017). Promoter regions and exons mostly the beginning of the gene are preferred target regions because the indels cause frameshift mutations that result in early translational termination, achieving a complete loss of gene function (Baltes and Voytas 2015, Baysal *et al.,* 2016). Although other outcomes, such as allele replacement or the insertion of DNA sequences at the target site can be achieved by providing donor DNA and selecting for rare homologous recombination events, the NHEJ repair pathway is more efficient in plants and gene knockout is by far the most common (Bortesi *et al*., 2016; Zhu *et al*., 2017)

The first report of genome editing in rice using CRISPR/Cas9 described knock-out experiments of the phytoene desaturase gene "*OsPDS"*. Biallelic mutations had an albino and dwarf phenotype (Shan *et al*., 2013). Since then, there have been numerous reports of genome editing of important agronomical traits in rice such as increased resistance to *Xanthomonas oryzae* by mutation of the *OsSWEET* gene (Zeng *et al*., 2020) and increased resistance to *Magnaporthe oryzae* by mutation of the ethylene responsive factor *ERF922* (Müller and Munné-Bosch, 2015). Another important desirable trait is herbicide resistance. Thus, CRISPR/Cas9 modification of acetolactate synthase (*ALS*) resulted in herbicide resistance in rice (Sun *et al*., 2016). Grain yield is a complex trait governed by different factors, many genes and quantitative trait loci (QTLs). Multiplex genome editing using CRISPR/Cas9 now permits manipulation of multiple alleles or muptiple paralogues simultaneously in the same plant. Xu *et al*. (2016) edited three QTLs related to grain weight, and the resulting mutation led to significant increases in grain weight. More recently, Li *et al*. (2018) demonstrated that the replacement of nitrate transporter *NRT1.1B* increased nitrogen use efficiency in rice.

1.1.2. Importance of nitrogen in plant growth and development

Nitrogen (N) is one of the essential elements and primary nutrients that limits plant growth, development and reproduction (Ohyama, 2010). Plants utilize nitrogen (N) in two different forms, namely nitrate (NO_3-) and ammonium (NH_4+). N is the fundamental element which shows its significant effects directly as increase in the yield and enhances plant quality by playing a vital role in many biochemical and physiological functions (Lawlor *et al*., 2001).

Nitrogen is one of the major components of chlorophyll, energy-transfer compounds such as ATP, nucleic acids and proteins. Therefore, nitrogen availability in plants limits photosynthesis, cell growth, metabolism, and protein synthesis (Ohyama, 2010). Thus, crop production totally depends on an adequate N supply especially in cereals such as rice, maize and wheat (Chapin *et al*., 2011).

1.1.2.1. *Nitrogen assimilation and environmental problems caused by excess of nitrogen*

Crop production requires high amounts of nitrogen and world demand for nitrogen fertilizer use increasing dramatically (**Figure 1.2.**). However, plants do not assimilate more than half of the N applied and it is a major source of pollution in agricultural systems (Cui *et al*., 2013; Kronzucker and Coskun, 2015).

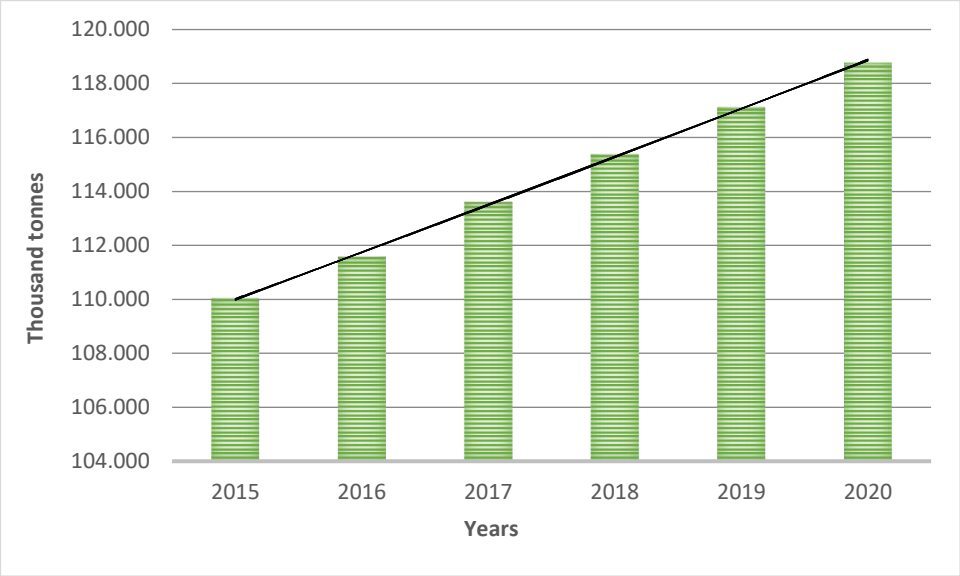


Figure 1.2. World demand for nitrogen fertilizer use, 2015-2020 (FAO).

In addition to its substantial economic cost, in high intensity agricultural production systems, the application of nitrogen fertilizers in excessive amounts causes environmental problems, such as water pollution, promotion of weeds, generation of greenhouse gasses, etc (Sutton *et al*., 2011, Stokstad, 2016).

Furthermore, although over fertilization is a problem in many industrialized countries, in developing countries lack of fertilizer availability is the major constraint in crop yields (Mueller *et al*., 2014) (**Figure 1.3.**).

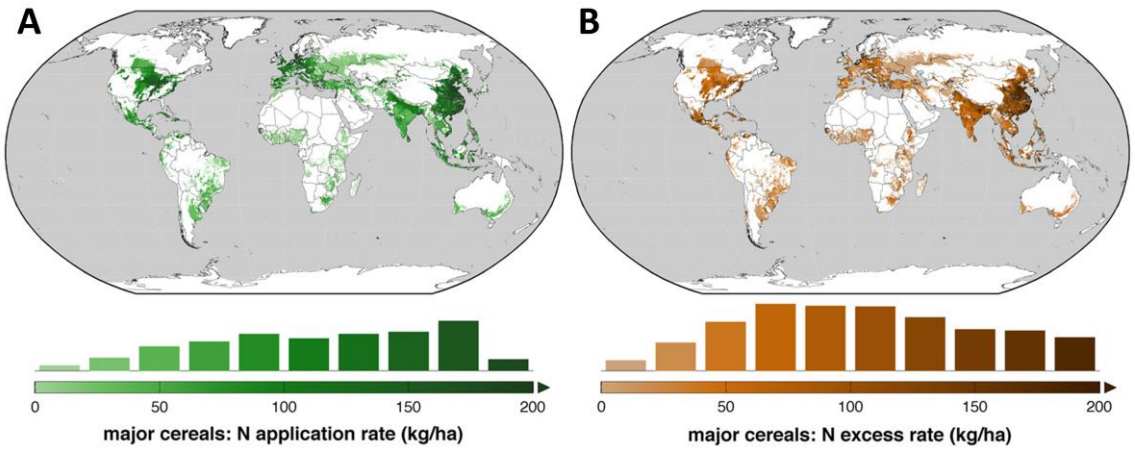


Figure 1.3. Nitrogen application rates for maize, wheat and rice. A) Avarage nitrogen application. B) Excess of applied nitrogen (Mueller *et al*., 2014).

1.1.2.2. *Alternative solutions to reduce N_2 usage*

Strategies to reduce the global dependence on nitrogen fertilizers are paramound. Improving the ability of plants and plant-associated organisms to fix and assimilate atmospheric nitrogen has been a major goal of plant biotechnologists for decades. Early strategies focused on increasing bacterial colonization in close proximity to the cereals root. Such bacteria are able to fix atmospheric nitrogen and improve release and assimilation of NH_3 produced in plant cells (Stoltzfus *et al*., 1997; Mus *et al*., 2016).

More recent strategies aim to introduce the N_2 fixation machinery into the plant itself, either indirectly using endosymbiotic bacteria or directly by transfer of prokaryotic N_2 fixation genes (López-Torrejón *et al*., 2016; Ivleva *et al.*, 2016; Allen *et al*., 2017; Burén *et al*., 2018; Eseverri *et al*., 2020; Jiang *et al.*, 2021).

The first strategy aimed to develop novel symbiosis mechanisms in non-legume plants, i.e., engineering cereals to associate with N_2-fixing bacteria and to form nodules similarly to legumes (Charpentier and Oldroyd 2010; Rogers and Oldroyd 2014; Mus *et al*. 2016). For this to succeed, two main bottlenecks need to be overcome. First, the plant has to be able to interact with bacteria, through crosstalk in a way that bacteria will recognize the plant as a suitable host. The other bottleneck involves formation of nodules or nodule-like structures that provide a low-O_2 environment to enable interchange of nutrients, mainly C, N and metals (Burén *et al*., 2018). It is recognized that Myc factors, involved in signaling between soil fungi and most plants (including cereals) when forming

symbiotic arbuscular mycorrhiza (Maillet *et al*. 2011) are similar to Nod factors secreted by symbiotic N-fixing bacteria. As Myc-factors are already recognized by most plants, engineering cereals capable of perceiving Nod-factors may be one approach to engineering N-fixation (Rogers and Oldroyd, 2014).

A more direct strategy involves transfer of prokaryotic *nif* genes into the plant genome. The main goal of this strategy is to engineer N_2-fixing machinery in the plant itself and to make plant fixing its own nitrogen without the need for bacterial interactions (Curatti and Rubio, 2014). However, the genetic complexity and fragility of the *nif* regulon (set of a number of operons used to regulate BNF), the extreme O_2 sensitivity of nitrogenase, necessity of many accessory proteins and metal clusters for maturation of the nitrogenase components are major barriers that make this strategy unachievable to date (Shah *et al.*, 1994; Dixon and Kahn 2004; Temme, Zhao and Voigt 2012; Poza-Carrión *et al*. 2014).

1.1.3. Biological Nitrogen fixation

Biological N_2 fixation (BNF), is the reduction of inert N_2 gas to ammonia (NH_3) through a reaction catalysed by nitrogenase. Nitrogenase is a complex and extremely oxygen-sensitive enzyme that requires multiple sub-units encoded by many different genes for its assembly (Rubio and Ludden 2008; Jasniewski *et al.*, 2018; Burén *et al*., 2018).

BNF is performed by a group of prokaryotes belonging to the bacteria or archaea domains. Eukaryotes are not capable of converting N_2 into a biologically useful form. Depending on the habitat in which these diazotrophic organisms grow, they can be divided into three groups: (i) free living (e.g., *Azotobacter vinelandii*), (ii) symbiotic, mainly bacteria living within plant root nodules, such as *Rhizobium* in legumes, and (iii) in animals such as the digestive tract of termites and coral reef sponges (Burén *et al*., 2018). In all cases, N_2 fixation is performed by the nitrogenase protein complex.

1.1.3.1. *Types of nitrogenases*

There are three structurally and functionally similar but genetically distinct nitrogenases distinguished, in part, by the metal composition of their corresponding active-site cofactors (Burén *et al*., 2018). The most abundant and ecologically relevant molybdenum (Mo) nitrogenase, and the alternative vanadium (V) and iron-only (Fe) nitrogenases

(Rubio and Ludden, 2008, Eseverri *et al*., 2020). "Mo-dependent" nitrogenase is known as the most-studied and well-characterized nitrogenase.

1.1.3.2. *Molybdenum-nitrogenase and its key subunits*

Mo-nitrogenase is an oxygen sensitive metalloenzyme which comprises two interacting components (subunits), dinitrogenase reductase and dinitrogenase protein. The Mo dinitrogenase reductase component is also refer to iron (Fe) protein because Fe is the only metal present in the cluster. It is a homodimer of the *nifH* product (López-Torrejón *et al*., 2016). Fe protein carries a single [4Fe-4S] cluster linking the two subunits and serves to mature the metalloclusters of the MoFe protein. Its main function is to provide electrons to the MoFe protein during N_2 reduction (Burén *et al*., 2019) **(Figure 1.4.)**. Therefore, Fe protein maturation is the key step of N_2 reduction and it involves the correct folding (formation) of the NifH homodimer, followed by the acquisition of the [4Fe-4S] cluster to generate active Fe protein (Burén *et al*., 2018).

The dinitrogenase component is also known as molybdenum iron (MoFe) protein because of the presence of Mo and Fe in its active site. The MoFe-protein is a heterodimer of the NifD (α-subunit) and NifK (β-subunit) gene pruducts. NifD forms an $\alpha_2\beta_2$ tetramer with NifK (Rubio and Ludden, 2008; Esteves-Ferreira *et al*., 2017) (**Figure 1.4**.).

The Fe and MoFe-protein together catalyze the reduction of nitrogen to ammonia. This enzymatic complex constitutes approximately 10% of the total cellular protein of many diazotrophs. It uses 8 electrons and 16 ATPs per molecule of fixed N_2 and catalyzes the synthesis of approximately half of all the fixed N_2 on earth (Berman-Frank *et al*., 2003, Esteves-Ferreira *et al*., 2017).

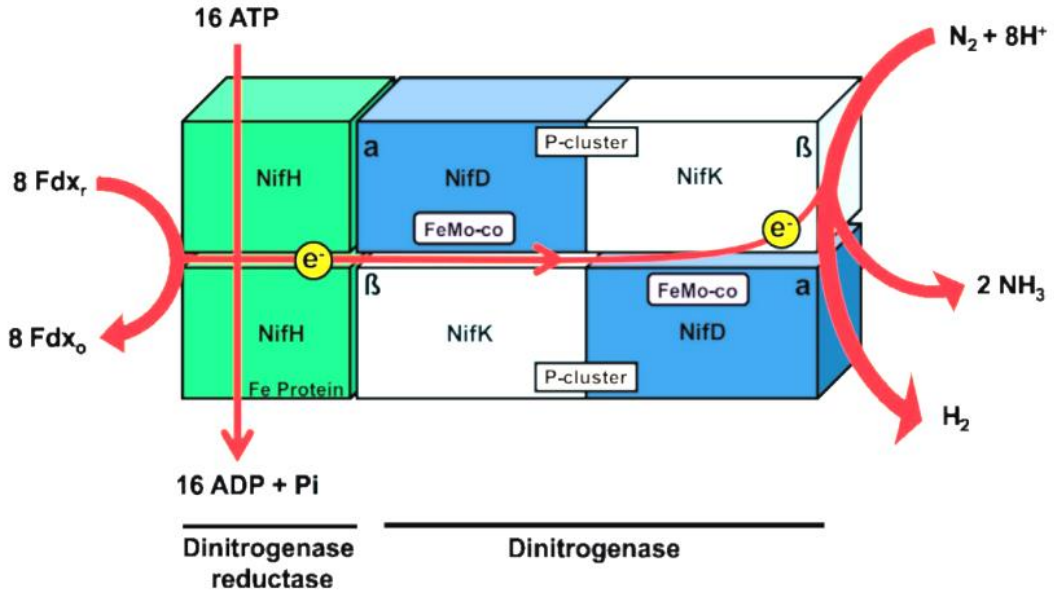


Figure 1.4. The structure of nitrogenase metalloenzyme. Dinitrogenase reductase is a homodimer with one [4Fe-4S]-cluster (Fe protein) encoded by *nif*H, whereas dinitrogenase is a MoFe-protein encoded by the genes *nifD* (α-subunit) and *nifK* (β-subunit) and organized in a α2β2 tetramer associated with two FeMo-cofactors and two P-clusters (Esteves-Ferreira *et al*., 2017).

1.1.4. Eukaryotic N_2 fixation

1.1.4.1. *Towards engineering nitrogen fixation in eukaryotes*

One of the most successful examples of engineering active nitrogenase components in eukaryotes to date is the production of active NifU and NifH within the mitochondria of yeast growing aerobically, overcoming the O_2 sensitivity of Nif proteins (López-Torrejón *et al*, 2016). Following this study, Ivleva *et al*. (2016) demonstrated the activity of stably expressed nitrogenease Fe protein in tobacco. However, it was active only when plants were grown under a low O_2 atmosphere. Eseverri *et al*. (2020) reported transient expression and accumulation of soluble NifH, M, U and S proteins within the stroma of mesophyll cells and the activity of NifH and NifU proteins in *N. benthamiana* chloroplasts. More recently, NifH expressed transiently in *N. benthamiana* leaves exhibited lower nitrogenase activity than that in yeast. However, transfer of [4Fe4S] clusters from NifU to NifH *in vitro* increased the activity of the *N. benthamiana*-isolated NifH by 10-fold, demonstrating that *N. benthamiana* mitochondria [Fe-S] cluster

availability still remains the main bottleneck to engineer plant nitrogenases (Jiang *et al* 2021).

1.1.4.2. *Transfer of bacterial genes to eukaryotes and minimum gene requirement for biological activity*

The total number of genes required for a functional nitrogenase enzyme differs among different organisms, but is usually estimated to ca. 10–20 genes (Zhao and Voigt 2012; Echavarri-Erasun and Rubio 2015, Burén *et al*., 2020). However, the entire prokaryotic *nif* gene cluster cannot be transferred directly to higher eukaryotes. Molecular strategies to achieve the correct stoichiometric expression of each individual gene to generate active nitrogenase are currenty out of reach technically. Therefore, an interim achieveable target, at least in principle is to transfer the genes encoding for the two main nitrogenase components and the genes necessary to generate active Fe and Mo-Fe proteins.

At least three accessory genes (*nifM, nifS and nifU*) genes are required to render active Fe protein in *Azotobacter vinelandii* (Burén *et al*., 2018). NifM, known as a putative peptidyl-prolyl cis-trans isomerase, assists in the proper folding of NifH (Gavini *et al*., 2006). NifS is a cysteine desulfurase responsible for the mobilization and transfer of sulphur to NifU (Burén *et al*., 2019), where [Fe-S] clusters are synthesized and later donated to the Fe protein for activation (Dos Santos *et al*., 2004). This is an oxygen sensitive process in which intermediates are generated and carried directly from one protein to another to protect them from reactive oxygen species (Zheng and Dos Santos, 2018). FeMo-co is produced in a pathway independent of NifDK and it is inserted into apo-NifDK to generate active NifDK (Burén *et al*., 2020). Therefore, the first crucial steps to generate active nitrogenase in higher eukaryotes is to transfer bacterial nitrogenase Fe and MoFe component genes *NifH* and *NifM, NifD and NifK* (with or without *NifS* and *NifU*), obtain correctly folded and matured NifH which is capable to provide electrons to NifDK.

1.1.4.3. *Optimum target locations to express Nif proteins*

As stated earlier, nitrogenase is a complex and an extremely oxygen-sensitive enzyme that requires specific cellular compartments for its assembly (Curatti and Rubio, 2014). Mitochondria and chloroplasts have been considered as potential locations to engineer

nitrogen fixation in plants. Because of the low O_2 environment and presence of bacterial type iron sulfur cluster biosynthesis machinery, mitocondria have been proposed as an appropriate location for the assembly of functional nitrogenase components (López-Torrejón *et al*., 2016, Jiang *et al.,* 2021).

Chloroplasts are also good candidates as a traget organelle because of the availability of the immediate products of photosynthesis, NADPH and ATP, that could satisfy nitrogen fixation demands of energy and reducing power (Merrick and Dixon 1984, Ivleva *et al.*, 2016; Eseverri *et al*., 2020). Moreover, the ammonia produced as a consequence of nitrogen fixation could be easily assimilated into amino acids by the glutamine synthetase/glutamate synthase pathway, found in plastids (Eseverri *et al*., 2020).

1.1.5. Metabolic engineering in rice for the production of valuable metabolites

1.1.5.1. *Starch metabolism*

Starch is made up of two types of glucose polymers namely amylose and amylopectin. Amylose is a linear molecule which composed of α-1,4 glucan chains and constitutes approximately 20% of total rice starch composition. Amylopectin is the major component of rice starch, up to 80%. It is a highly branched polysaccharide because of its α-1,6 glucan chains which connect linear chains (Nishi *et al*., 2001; Crofts *et al*., 2015).

Amylose is synthesized by ADP-Glc pyrophosphorylase and granule-bound starch synthase (GBSS), whereas amylopectin is a product of a complex set of reactions catalyzed by starch synthase (SS), starch-branching enzymes (SBE), starch-debranching enzymes (SDBE) pullulanase (PUL) and starch phosphorylase (Pho) (Satoh *et al*., 2008; Abe *et al*., 2014).

The rice genome encodes eleven isozymes of starch synthase (SSs) to elongate α-1,4-linked linear glucan polymers using ADP-glucose as a substrate, three isozymes of SBEs to produce α-1,6-linked branches for amylopectin formation, four isozymes of DBEs and two isozymes of Pho to hydrolyze and remove the α-1,6-linked branches of the amylopectin structure (Ohdan *et al*., 2005). Starch branching enzymes (SBEs) are the only enzyme that can introduce α-1,6-glucosidic linkages into α-polyglucans, and they are therefore required to introduce the branch points during amylopectin biosynthesis (Smith *et al*., 1999; Nakamura, 2002).

Three isoforms of branching enzyme SBEI, SBEIIa and SBEIIb are present in rice seeds. SBEIIb is the major isozyme in rice starch biosynthesis because it is specifically expressed in the endosperm and plays a unique role in the transfer of short chains in the crystalline lamellae of amylopectin, whereas *SBEIIa* is primarily expressed in leaves and lacks this function (Nishi *et al*., 2001; Nakamura, 2002).

Inactivation of *SBEIIb* generates the rice amylose extender mutant (*ae*) (Mizuno *et al*., 1993). The SBEIIb mutant exhibits an opaque seed phenotype, has smaller grains, less glucan chains and and an increase in amylose content or long chains of amylopectin. In contrast, loss of SBEI and SBEIIa does not alter seed morphology in rice (Satoh *et al.*, 2003).

1.1.5.2. *Modified rice starches in industrial applications*

Starch is one the most important raw material for the food industry not only for its nutritional value, but also for its broad functionality (Fuentes-Zaragoza *et al*., 2011). Therefore, the content of amylose and amylopectin together with lipids, proteins and phosphorus in starch granules have significant effects on the physicochemical properties and functionality of starch (Baysal *et al.*, 2016; Baysal *et al.*, 2020). Starch in general it is rarely consumed in its native form because of its low water solubility, which limits industrial applications (Fuentes-Zaragoza *et al*., 2011). Physical and chemical modifications can greatly improve the characteristics of native starch by altering its properties. Physical modification of starch can increase its solubility in water and reduce the size of the starch granules for specialized uses. Chemical modifications are made by introducing functional groups to the starch molecule without affecting the morphology or granule size. These modifications alter the dielectric properties of granules, reduce the paste temperature and increase its viscosity (Wang *et al.*, 2020). When starch is physically or chemically modified, it can be adapted for different purposes in food and non-food applications. In the food industry modified starch mainly is used as a modifier of texture, viscosity, adhesion, moisture retention, gel formation and films and in the non-food industry as raw material for paper manufucturing, cosmetics, adhesives and textiles (Waterschoot *et al*., 2015).

1.1.5.3. *Resistant starch (RS) and its effect on human health*

Starch is the one of the major energy sources in the human diet. Resistant starch refers to the portion of starch or starch products that resist digestion, passes through the gastrointestinal channels to reach the end of the digestive system and to be used as substrate of fecal microflora (Birt *et al*., 2013). Factors that affect starch digestibility include the structural characteristics of the starch, primarly due to amylose/amylopectin ratio and presence of amylose-lipid complexes on the starch molecule (Cornejo-Ramirez *et al*., 2018). Lipids interact with amylose and amylose-lipid forms decrease amylase activity and consequently reduce starch digestibility (Taylor *et al*., 2015).

Highly branched amylopectin molecules are easily digested by α-amylase in the oral cavity and this rapid digestion resuls in an immediate increase in blood glucose levels (Baysal *et al*., 2020). In contrast, the linear structure of amylose limits the accessibility of digestive enzymes and reduces starch degradation products that are not absorbed by the small intestine. Fermentation products of RS are short chain fatty acids with different physiological and probiotic effects (Conde-Petit *et al*., 2001).

There is a general interest in increasing the level of RS because of its capacity to produce high levels of butyrate throughout the colon (Fuentes-Zaragoza *et al*., 2011). Butyrate is the most important energy source for colonocytes and it has demonstrated many beneficial effects, such as exhibition of anti-inflammatory and anti-cancer functions in the human gut and inhibition of a variety of factors that propagate the initiation and growth of tumors in the mammalian colon (Champ *et al*., 2003; Fuentes-Zaragoza *et al*., 2011; Donohoe *et al*., 2011). Therefore, RS inhibits insulin release, lowers the risk of diabetes, obesity, and heart disease, and also acts as a prebiotic to support a healthy colon microbiome. Usually, rice mostly has 18-20 % amylose content and 0.2-2% resistant starch (Tetlow, 2010). Over-expression of granule-bound starch synthase I (GBSSI), encoded by the waxy (Wx) or down regulation of SBEIIb are two distinct ways to increase amylose and resistant starch content in rice (Baysal *et al*., 2016; Pérez *et al*., 2019).

1.1.5.4. *Generation of rice starch mutants through genome editing*

Genome editing studies mainly focused on the alteration of total starch content or modification of two types of starches which are known as “waxy” and “high amylose” starch. Waxy starches due to the waxy appearance of the endosperm tissue contains a minimal amount of amylose in their granule composition (<15%). In contrast high amylose starches have a high content of amylose (>15%); these starches exhibit a slight deformation in granule appearance (Hung *et al.*, 2007; Pérez *et al*., 2019).

CRISPR/Cas9 inactivation of endosperm-specific Waxy/GBSSI gene reduced the amylose content of the endosperm to 5% in homozygous seeds and 8–12% in heterozygous seeds resulting in a fully translucent phenotype (Pérez *et al*., 2019). CRISPR/Cas9 mutation of cytosolic AGPase large subunit gene “APL2” induced the ectopic expression of APL2 and the corresponding small subunit gene APS2b in leaves decreased total starch content to 85% and increased soluble sugar content in leaves by 40% (Pérez *et al*., 2019).

A CRISPR/Cas9 induced mutation of *SBEIIb* resulted in higher proportion of long chains in the form of debranched amylopectin, elevated amylose content to 25% and resistant starch content to 9.8% (Sun *et al*., 2017).

In addition to genes directly involved in starch biosynthesis, CRISPR/Cas9 mutations of related metabolic genes such as sugar metabolism also influenced starch biosynthesis. The double knockout of *OsSPS1* and *OsSPS11* caused the accumulation of leaf starch, although there was no significant impact on plant growth (Hashida *et al*., 2016). Ma *et al*. (2017) knocked out *SWEET11*, reduced the sucrose concentration by 40% in the mutant embryo sacs, resulted in defective grain filling and a 5% drop in the starch content of the mature caryopses.

Deng *et al*. (2020) generated single and double knockouts of the sugar metabolism vacuolar invertase genes *INV2* and *INV3*. The grain size of the *inv2* mutant was normal, but that of the *inv3* mutant and *inv2-inv3* double-knockout mutant was smaller, reducing the grain weight by 33.5%. In all knockout mutants, the sucrose level was higher but the total hexose content was lower. Furthermore, the total starch content was similar in the wild-type plants and knockout mutants but the amylose content of the mutants was 3–6%

lower (Deng *et al*., 2020). Similarly, inactivating *PGM* (encoding plastidial phosphoglucomutase) and *APL4* (encoding the plastidial AGPase large subunit) inhibited starch biosynthesis in the seeds (Lee *et al*., 2016).

1.1.6. References

Abe, N., Asai, H., Yago, H., Oitome, N. F., Itoh, R., Crofts, N., Nakamura, Y., & Fujita, N. (2014). Relationships between starch synthase I and branching enzyme isozymes determined using double mutant rice lines. *BMC Plant Biology*, 14, 80.

Alcázar-Alay, S.C., & Meireles, M.A.A. (2015). Physicochemical properties, modifications and applications of starches from different botanical sources. *Food Science and Technology*, 35, 215-236.

Albajes, R., Cantero-Martínez, C., Capell, T., Christou, P., Farre, A., Galceran, J., López-Gatius, F., Marin, S., Martín-Belloso, O., Motilva, Ma-J., Nogareda, C., Peman, J., Puy, J., Recasens, J., Romagosa, I., Romero, Ma-P., Sanchis, V., Savin, R., Slafer, G.A. Soliva-Fortuny, R., Viñas, I. & Voltas, J. (2013). Building bridges: an integrated strategy for sustainable food production throughout the value chain. *Molecular Breeding*, 32, 743–770.

Allen, R. S., Tilbrook, K., Warden, A. C., Campbell, P. C., Rolland, V., Singh, S. P., & Wood, C. C. (2017). Expression of 16 Nitrogenase Proteins within the Plant Mitochondrial Matrix. *Frontiers in Plant Science*, 8, 287.

Ayres, N.M., & Park, W.D. (1994). Genetic Transformation of Rice. *Critical Reviews in Plant Science*, 13, 219-239

Baltes, N. J., & Voytas, D. F. (2015). Enabling plant synthetic biology through genome engineering. *Trends in Biotechnology*, 33, 120–131.

Baysal, C., Bortesi, L., Zhu, C., Farre, G., Schillberg, S., & Christou, P. (2016). CRISPR/Cas9 activity in the rice *OsBEIIb* gene does not induce off-target effects in the closely related paralog *OsBEIIa*. *Molecular Breeding*, 36,108.

Baysal, C., He, W., Drapal, M., Villorbina, G., Medina, V., Capell, T., Khush, G. S., Zhu, C., Fraser, P. D., & Christou, P. (2020). Inactivation of rice starch branching enzyme IIb triggers broad and unexpected changes in metabolism by transcriptional reprogramming. *Proceedings of the National Academy of Sciences of the United States of America*, 117, 26503–26512.

Berman, J., Zhu, C., Pérez-Massot, E., Arjó, G., Zorrilla-López, U., Masip, G., Banakar, R., Sanahuja, G., Farré, G., Miralpeix, B., Bai, C., Vamvaka, E., Sabalza, M., Twyman, R. M., Bassié, L., Capell, T., & Christou, P. (2013). Can the world afford to ignore biotechnology solutions that address food insecurity? *Plant Molecular Biology*, 83, 5–19.

Berman-Frank, I., Lundgren, P., & Falkowski, P. (2003). Nitrogen fixation and photosynthetic oxygen evolution in cyanobacteria. *Research in Microbiology*, 154, 157–164.

Birt, D. F., Boylston, T., Hendrich, S., Jane, J. L., Hollis, J., Li, L., McClelland, J., Moore, S., Phillips, G. J., Rowling, M., Schalinske, K., Scott, M. P., & Whitley, E. M. (2013). Resistant starch: promise for improving human health. *Advances in Nutrition*, 4, 587–601.

Bortesi, L., & Fischer, R. (2015). The CRISPR/Cas9 system for plant genome editing and beyond. *Biotechnology Advances*, 33, 41–52.

Burén, S., Young, E. M., Sweeny, E. A., Lopez-Torrejón, G., Veldhuizen, M., Voigt, C. A., & Rubio, L. M. (2017). Formation of Nitrogenase NifDK Tetramers in the Mitochondria of *Saccharomyces cerevisiae*. *ACS Synthetic Biology*, 6, 1043–1055.

Burén, S., & Rubio, L. M. (2018). State of the art in eukaryotic nitrogenase engineering. *FEMS Microbiology Letters*, 365, fnx274.

Burén, S., Pratt, K., Jiang, X., Guo, Y., Jimenez-Vicente, E., Echavarri-Erasun, C., Dean, D. R., Saaem, I., Gordon, D. B., Voigt, C. A., & Rubio, L. M. (2019). Biosynthesis of the nitrogenase active-site cofactor precursor NifB-co in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America*, 116, 25078–25086.

Burén, S., Jiménez-Vicente, E., Echavarri-Erasun, C., & Rubio, L. M. (2020). Biosynthesis of Nitrogenase Cofactors. *Chemical Reviews*, 120, 4921–4968.

Cornejo-Ramírez, Y. I., Martínez-Cruz, O., Toro-Sánchez, C. L. D., Wong-Corral, F. J., Borboa-Flores, J., & Cinco-Moroyoqui, F. J. (2018). The structural characteristics of starches and their functional properties. *CYTA - Journal of Food*, 16, 1003-1017.

Cui, S., Shi, Y., Groffman, P. M., Schlesinger, W. H., & Zhu, Y. G. (2013). Centennial-scale analysis of the creation and fate of reactive nitrogen in China (1910-2010). *Proceedings of the National Academy of Sciences of the United States of America*, 110, 2052–2057.

Champ, M., Langkilde, A. M., Brouns, F., Kettlitz, B., & Bail-Collet, Y. L. (2003). Advances in dietary fibre characterisation. 2. Consumption, chemistry, physiology and measurement of resistant starch; implications for health and food labelling. *Nutrition Research Reviews*, 16, 143–161.

Chapin, F.S.III., Matson, P.A., & Vitousek, P. (2011). *Principles of Terrestrial Ecosystem Ecology*. New York, NY: Springer Science & Business Media.

Charpentier, M., & Oldroyd, G. (2010). How close are we to nitrogen-fixing cereals? *Current Opinion in Plant Biology*, 13, 556–564.

Conde-Petit, B., Nuessli, J., Arrigoni, E., Escher, F., & Amadò, R. (2001). Perspectives of starch in food science. *Chimia*, 55, 201-205.

Crofts, N., Abe, N., Oitome, N. F., Matsushima, R., Hayashi, M., Tetlow, I. J., Emes, M. J., Nakamura, Y., & Fujita, N. (2015). Amylopectin biosynthetic enzymes from developing rice seed form enzymatically active protein complexes. *Journal of Experimental Botany*, 66, 4469–4482.

Curatti, L., & Rubio, L. M. (2014). Challenges to develop nitrogen-fixing cereals by direct nif-gene transfer. *Plant Science*, 225, 130–137.

Deng, X., Han, X., Yu, S., Liu, Z., Guo, D., He, Y., Li, W., Tao, Y., Sun, C., Xu, P., Liao, Y., Chen, X., Zhang, H., & Wu, X. (2020). *OsINV3* and Its Homolog, *OsINV2*, Control Grain Size in Rice. *International Journal of Molecular Sciences*, 21, 2199.

Dixon, R., & Kahn, D. (2004). Genetic regulation of biological nitrogen fixation. *Nature Reviews. Microbiology*, 2, 621–631.

Donohoe, D. R., Garge, N., Zhang, X., Sun, W., O'Connell, T. M., Bunger, M. K., & Bultman, S. J. (2011). The microbiome and butyrate regulate energy metabolism and autophagy in the mammalian colon. *Cell Metabolism*, 13, 517–526.

Echavarri-Erasun, C., Gonzalez, E., & y Rubio Herrero, L.M. (2015). *A novel method to isolate native NIFB-cofactor*. In: "*XV Congress of the Spanish Society of Nitrogen Fixation, and the IV Portuguese-Spanish Congress on Nitrogen Fixation*", 16/06/2015-18/06/2015, León. p. 1.

Eseverri, A., López-Torrejón, G., Jiang, X., Burén, S., Rubio, L. M., & Caro, E. (2020). Use of synthetic biology tools to optimize the production of active nitrogenase Fe protein in chloroplasts of tobacco leaf cells. *Plant Biotechnology Journal*, 18, 1882–1896.

Esteves-Ferreira, A. A., Inaba, M., Fort, A., Araújo, W. L., & Sulpice, R. (2018). Nitrogen metabolism in cyanobacteria: metabolic and molecular control, growth consequences and biotechnological applications. *Critical Reviews in Microbiology*, 44, 541–560.

Farré, G., Twyman, R. M., Christou, P., Capell, T., & Zhu, C. (2015). Knowledge-driven approaches for engineering complex metabolic pathways in plants. *Current Opinion in Bbiotechnology*, 32, 54–60.

Food and Agriculture Organization (FAO) (2019). Production quantities of rice, paddy by country. http://www.fao.org/faostat/en/#data/QC/visualize

Food and Agriculture Organization (FAO) (2020). World fertilizer trends and outlook to 2020. http://www.fao.org/3/i6895e/i6895e.pdf

Fuentes-Zaragoza. E., Sánchez-Zapata, E., Sendra, E., Sayas, E., Navarro, C., Fernández-López, J., Pérez-Álvarez, J. (2011). Resistant starch as prebiotic: A review. *Starch*, 63, 406-415.

Gnanamanickam S.S. (2009) Rice and Its Importance to Human Life. In: *Biological Control of Rice Diseases. Progress in Biological Control*, vol 8. Springer, Dordrecht.

Hashida, Y., Hirose, T., Okamura, M., Hibara, K. I., Ohsugi, R., & Aoki, N. (2016). A reduction of sucrose phosphate synthase (SPS) activity affects sucrose/starch ratio in leaves but does not inhibit normal plant growth in rice. *Plant Science*, 253, 40–49.

Hoogenkamp, H.R., Bakker, G-J, Wolf, L., Suurs, P., Dunnewind, B., Barbut, S., Friedl, P., van Kuppevelt, T.H., & Daamen, W.F., & (2015). Directing collagen fibers using counter-rotating cone extrusion. *Acta Biomaterialia*, 12, 113-121.

Hung, P. V., Maeda, T., & Morita, N. (2007). Study on physicochemical characteristics of waxy and high-amylose wheat starches in comparison with normal wheat starch. *Starch*, 59, 125-131.

Ikehashi, H. (2014). Domestication and Long-Distance Dissemination of Rice: A Revised Version. *Journal of Rice Research*, 3:128.

Jasniewski, A. J., Sickerman, N. S., Hu, Y., Ribbe, M. W. (2018). The Fe Protein: An Unsung Hero of Nitrogenase. *Inorganics*. 6, 1-25.

Jiang, X., Payá-Tormo, L., Coroian, D., García-Rubio, I., Castellanos-Rueda, R., Eseverri, Á., López-Torrejón, G., Burén, S., & Rubio, L. M. (2021). Exploiting genetic diversity and gene synthesis to identify superior nitrogenase NifH protein variants to engineer N_2-fixation in plants. *Communications Biology*, 4, 4.

Khush G. S. (1997). Origin, dispersal, cultivation and variation of rice. *Plant Molecular Biology*, 35, 25–34.

Koch, M. S., Ward, J. M., Levine, S. L., Baum, J. A., Vicini, J. L., & Hammond, B. G. (2015). The food and environmental safety of Bt crops. *Frontiers in Plant Science*, 6, 283.

Kronzucker, H. J., & Coskun, D. (2015). "Bioengineering nitrogen acquisition in rice: promises for global food security," In: F. J. de Bruijn (ed.), *Biological Nitrogen Fixation* (Hoboken, NJ: JohnWiley & Sons, Inc.), 47–56.

Lawlor D.W., Lemaire G., Gastal F. (2001) Nitrogen, Plant Growth and Crop Yield. In: Lea P.J., Morot-Gaudry JF. (eds) *Plant Nitrogen*. Springer, Berlin, Heidelberg.

Lee, S. K., Eom, J. S., Hwang, S. K., Shin, D., An, G., Okita, T. W., & Jeon, J. S. (2016). Plastidic phosphoglucomutase and ADP-glucose pyrophosphorylase mutants impair starch synthesis in rice pollen grains and cause male sterility. *Journal of Experimental Botany*, 67, 5557–5569.

Letourneau, D. K., Robinson, G. S., & Hagen, J. A. (2003). Bt crops: predicting effects of escaped transgenes on the fitness of wild plants and their herbivores. *Environmental Biosafety Research*, 2, 219–246.

Li, J., Zhang, X., Sun, Y., Zhang, J., Du, W., Guo, X., Li, S., Zhao, Y., & Xia, L. (2018). Efficient allelic replacement in rice by gene editing: A case study of the NRT1.1B gene. *Journal of Integrative Plant Biology*, 60, 536–540.

López-Torrejón, G., Jiménez-Vicente, E., Buesa, J. M., Hernandez, J. A., Verma, H. K., & Rubio, L. M. (2016). Expression of a functional oxygen-labile nitrogenase component in the mitochondrial matrix of aerobically grown yeast. *Nature Communications*, 7, 11426.

Nagendra Prasad, M.N., Sanjay, K.R., Shravya Khatokar, M., Vismaya, M.N., & Nanjunda Swamy, S. (2011). Health Benefits of Rice Bran - A Review. *Journal of Nutrition and Food Sciences*, 1, 108.

Ma, L., Zhang, D., Miao, Q., Yang, J., Xuan, Y., & Hu, Y. (2017). Essential Role of Sugar Transporter OsSWEET11 During the Early Stage of Rice Grain Filling. *Plant & Cell Physiology*, 58, 863–873.

Maillet, F., Poinsot, V., André, O., Puech-Pagès, V., Haouy, A., Gueunier, M., Cromer, L., Giraudet, D., Formey, D., Niebel, A., Martinez, E. A., Driguez, H., Bécard, G., & Dénarié, J. (2011). Fungal lipochitooligosaccharide symbiotic signals in arbuscular mycorrhiza. *Nature*, 469, 58–63.

Merrick, M., & Dixon, R. (1984). Why don't plants fix nitrogen? *Trends in Biotechnology*. 2, 162

Mizuno, K., Kobayashi, E., Tachibana, M., Kawasaki, T., Fujimura, T., Funane, K., Kobayashi, M., & Baba, T. (2001). Characterization of an isoform of rice starch branching enzyme, RBE4, in developing seeds. *Plant & Cell Physiology*, 42, 349–357.

Mueller, N.D., West, P.C., Gerber, J.S., MacDonald, G.K., Stephen Polasky, S., & Foley, J.A. (2014). A tradeoff frontier for global nitrogen use and cereal production. *Environmental Research Letters*, 9, 5.

Müller, M., & Munné-Bosch, S. (2015). Ethylene Response Factors: A Key Regulatory Hub in Hormone and Stress Signaling. *Plant Physiology*, 169, 32–41.

Mus, F., Crook, M. B., Garcia, K., Garcia Costas, A., Geddes, B. A., Kouri, E. D., Paramasivan, P., Ryu, M. H., Oldroyd, G., Poole, P. S., Udvardi, M. K., Voigt, C. A., Ané, J. M., & Peters, J. W. (2016). Symbiotic Nitrogen Fixation and the Challenges to Its Extension to Nonlegumes. *Applied and Environmental Microbiology*, 82, 3698–3710.

Nakamura Y. (2002). Towards a better understanding of the metabolic system for amylopectin biosynthesis in plants: rice endosperm as a model tissue. *Plant & Cell Physiology*, 43, 718–725

Nishi, A., Nakamura, Y., Tanaka, N., & Satoh, H. (2001). Biochemical and genetic analysis of the effects of amylose-extender mutation in rice endosperm. *Plant Physiology*, 127, 459–472.

Ohyama, T. (2010) Nitrogen as a Major Essential Element of Plants. In: Ohyama, T. and Sueyoshi, K., (eds.), *Nitrogen Assimilation in Plants*, Research Singpot, Kerala, 1-18.

Ohdan, T., Francisco, P. B., Jr, Sawada, T., Hirose, T., Terao, T., Satoh, H., & Nakamura, Y. (2005). Expression profiling of genes involved in starch synthesis in sink and source organs of rice. *Journal of Experimental Botany*, 56, 3229–3244.

Pérez, L., Soto, E., Farre, G., Juanos, J., Villorbina, G., Bassie, L., & Christou, P. (2019). CRISPR/Cas9 mutations in the rice Waxy/GBSSI gene induce allele-specific and zygosity-dependent feedback effects on endosperm starch biosynthesis. *Plant Cell Reports*, 38, 417-433.

Pérez, L., Soto, E., Villorbina, G., Bassie, L., Medina, V., Muñoz, P., Capell, T., Zhu, C., Christou, P., & Farré, G. (2018). CRISPR/Cas9-induced monoallelic mutations in the cytosolic AGPase large subunit gene APL2 induce the ectopic expression of APL2 and the corresponding small subunit gene APS2b in rice leaves. *Transgenic Research*, 27, 423-439.

Poza-Carrión, C., Jiménez-Vicente, E., Navarro-Rodríguez, M., Echavarri-Erasun, C., & Rubio, L. M. (2014). Kinetics of Nif gene expression in a nitrogen-fixing bacterium. *Journal of Bacteriology*, 196, 595–603.

Rogers, C., & Oldroyd, G. E. (2014). Synthetic biology approaches to engineering the nitrogen symbiosis in cereals. *Journal of Experimental Botany*, 65, 1939–1946.

Rubio, L. M., & Ludden, P. W. (2008). Biosynthesis of the iron-molybdenum cofactor of nitrogenase. *Annual Review of Microbiology*, 62, 93–111.

Satoh, H., Shibahara, K., Tokunaga, T., Nishi, A., Tasaki, M., Hwang, S. K., Okita, T. W., Kaneko, N., Fujita, N., Yoshida, M., Hosaka, Y., Sato, A., Utsumi, Y., Ohdan, T., & Nakamura, Y. (2008). Mutation of the plastidial alpha-glucan phosphorylase gene in rice affects the synthesis and structure of starch in the endosperm. *The Plant Cell*, 20, 1833–1849.

Shah, V. K., Allen, J. R., Spangler, N. J., & Ludden, P. W. (1994). In vitro synthesis of the iron-molybdenum cofactor of nitrogenase. Purification and characterization of NifB cofactor, the product of NIFB protein. *The Journal of Biological Chemistry*, 269, 1154–1158.

Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., Zhang, K., Liu, J., Xi, J. J., Qiu, J. L., & Gao, C. (2013). Targeted genome modification of crop plants using a CRISPR-Cas system. *Nature Biotechnology*, 31, 686–688.

Smith, A. M., Denyer, K., & Martin, C. (1997). The synthesis of the starch granule. *Annual Review of Plant Physiology and Plant Molecular Biology*, 48, 67–87.

Statista, 2020 https://www.statista.com/statistics/255945/top-countries-of-destination-for-us-rice-exports-2011/

Stokstad E. (2016). Conservation researchers get a new roost in Cambridge. *Science*, 351, 114.

Stoltzfus, J.R., So, R., Malarvithi, P.P., Ladha, J.K., & de Bruijn, F.J. (1997). Isolation of endophytic bacteria from rice and assessment of their potential for supplying rice with biologically fixed nitrogen. *Plant and Soil*, 194, 25–36).

Sun, Y., Jiao, G., Liu, Z., Zhang, X., Li, J., Guo, X., Du, W., Du, J., Francis, F., Zhao, Y., & Xia, L. (2017). Generation of High-Amylose Rice through CRISPR/Cas9-Mediated Targeted Mutagenesis of Starch Branching Enzymes. *Frontiers in Plant Science*, 8, 298.

Sutton, M.A., Oenema, O., Erisman, J.W., Leip, A., van Grinsven, H., & Winiwarter, W. (2011) Too much of a good thing. *Nature*, 472, 159–161.

Tabashnik, B. E., Dennehy, T. J., Sims, M. A., Larkin, K., Head, G. P., Moar, W. J., & Carrière, Y. (2002). Control of resistant pink bollworm (Pectinophora gossypiella) by transgenic cotton that produces *Bacillus thuringiensis* toxin Cry2Ab. *Applied and Environmental Microbiology*, 68, 3790–3794.

Taylor, J.R., Naushad Emmambux, M., & Kruger, J. (2015). Developments in modulating glycaemic response in starchy cereal foods. *Starch*, 67, 79-89.

Temme, K., Zhao, D., & Voigt, C. A. (2012). Refactoring the nitrogen fixation gene cluster from Klebsiella oxytoca. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 7085-7090.

Van Frankenhuyzen, K., Gringorten, L., & Gauthier, D. (1997). Cry9Ca1 Toxin, a Bacillus thuringiensis Insecticidal Crystal Protein with High Activity against the Spruce Budworm (*Choristoneura fumiferana*). *Applied and Environmental Microbiology*, 63, 4132–4134.

Wang, S., Wang, J., Liu, Y., & Liu, X. (2020). Starch Modification and Application. In: Wang S. (eds) *Starch Structure, Functionality and Application in Foods*. Springer, Singapore.

Waterschoot, J., Gomand, S.V., Fierens, E., & Delcour, J.A. (2015). Starch blends and their physicochemical properties. *Starch*, 67, 1-13.

Xu, Y., Li, P., Yang, Z., & Xu, C. (2016). Genetic mapping of quantitative trait loci in crops. *Crop Journal*, 5,175-184

Temme, K., Zhao, D., & Voigt, C. A. (2012). Refactoring the nitrogen fixation gene cluster from Klebsiella oxytoca. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 7085–7090.

Zeigler, R. S., & Barclay, A. (2008). The relevance of rice. *RICE*, 1, 3-10.

Zhu, C., Bortesi, L., Baysal, C., Twyman, R. M., Fischer, R., Capell, T., Schillberg, S., & Christou, P. (2017). Characteristics of Genome Editing Mutations in Cereal Crops. *Trends in Plant Science*, 22, 38–52.

Zeng, X., Luo, Y., Vu, N., Shen, S., Xia, K., & Zhang, M. (2020). CRISPR/Cas9-mediated mutation of OsSWEET14 in rice cv. Zhonghua11 confers resistance to *Xanthomonas oryzae* pv. *oryzae* without yield penalty. *BMC Plant Biology*, 20, 313.

1.2. AIMS AND OBJECTIVES

The aims of my dissertation were two-fold. Firstly, to use rice as a model to understand and resolve current bottlenecks in engineering biological nitrogen fixation in cereal crops. The second aim was to use genome editing to redesign starch biosynthesis in the endosperm to understand how changes in the expression of starch biosynthetic genes impact broad metabolism.

The following specific **objectives** were addressed:

- Used different mitochondrial targeting sequences for the effective targeting of recombinant proteins to this sub-cellular organelle.
- Transfered the key nitrogenase genes (*nifH, nifM*) and confirmed their expression at mRNA and protein levels in rice dedifferentiated cells and regenerated plants.
- Identified the specific (trans)gene combination(s) required to recover rice callus and plant lines accumulating high levels of NifH (Fe) protein.
- Identified and resolved constraints in obtaining enzymatically active Fe protein.
- Evaluated the impact of mutating endogenous *OsSBEIIb* on starch properties, seed morphology, and the expression of other starch biosynthetic genes in rice endosperm and leaves.
- Determined the effect of the *OsSBEIIb* mutation on overall starch biosynthesis and broader primary and secondary metabolism in the rice endosperm.

Chapter 2

Recognition motifs rather than phylogenetic origin influence the ability of targeting peptides to import nuclear-encoded recombinant proteins into rice mitochondria

2.0. Abstract

Mitochondria fulfil essential functions in respiration and metabolism as well as regulating stress responses and apoptosis. Most native mitochondrial proteins are encoded by nuclear genes and are imported into mitochondria via one of several receptors that recognize N-terminal signal peptides. The targeting of recombinant proteins to mitochondria therefore requires the presence of an appropriate N-terminal peptide, but little is known about mitochondrial import in monocotyledonous plants such as rice (*Oryza sativa*). To gain insight into this phenomenon, nuclear-encoded enhanced green fluorescent protein (eGFP) was targeted rice mitochondria using six mitochondrial pre-sequences with diverse phylogenetic origins, and their effectiveness were investigated by immunoblot analysis as well as confocal and electron microscopy. I found that the ATPA and COX4 (*Saccharomyces cerevisiae*), SU9 (*Neurospora crassa*), pFA (*Arabidopsis thaliana*) and OsSCSb (*Oryza sativa*) peptides successfully directed most of the eGFP to the mitochondria, whereas the MTS2 peptide (*Nicotiana plumbaginifolia*) showed little or no evidence of targeting ability even though it is a native plant sequence. This data therefore indicate that the presence of particular recognition motifs may be required for mitochondrial targeting, whereas the phylogenetic origin of the pre-sequences probably does not play a key role in the success of mitochondrial targeting in dedifferentiated rice callus and plants.

2.1. Introduction

The proteins synthesized by eukaryotic cells are targeted to particular subcellular compartments. In the absence of specific targeting signals, nascent proteins accumulate by default in the cytosol (Kim and Hwang, 2013). Proteins carrying an N-terminal signal peptide recognized by a signal recognition particle (SRP) on the surface of the endoplasmic reticulum (ER) are co-translationally imported into the secretory pathway, from where they can be routed to various other compartments including the nucleus (Luirink and Sinning, 2004; Akopian *et al.,* 2013). Proteins carrying other types of N-terminal or C-terminal peptides are directed post-translationally to peroxisomes, mitochondria or (in plants) to the plastids (Egea *et al.,* 2005). Whereas most organelles are thought to have originated ultimately from the plasma membrane during eukaryote evolution, the mitochondria and plastids are exceptional because they evolved

independently from endosymbionts and therefore carry their own genomes (Dolezal *et al.*, 2006). Accordingly, mitochondrial and plastid proteins can be derived either from the organelle genome or the nuclear genome, hence the need for protein import pathways for nuclear-encoded proteins (Chacinska *et al*., 2009; Endo *et al*., 2011).

Mitochondria carry out multiple essential functions in the eukaryotic cell, including the generation of ATP by oxidative phosphorylation, the biosynthesis of amino acids and lipids, and the regulation of apoptosis (Rasmusson *et al.*, 2004; Sluse *et al.*, 2006). Approximately 98% of the enzymes and other proteins required for mitochondrial functions are encoded by the nuclear genome and the remainder by the mitochondrial genome (Taylor and Pfanner, 2004). Proteins encoded by the nuclear genome must therefore be imported into the mitochondria, and accordingly the pre-proteins carry N-terminal peptides that are recognized by the hydrophobic binding pockets of the receptor Tom20 (Yamamoto *et al*., 2011). The targeting peptides are cleaved off by mitochondrial peptidases during or after import, yielding the mature protein (Brix *et al*., 1997; Obita *et al.*, 2003; Taylor and Pfanner, 2004; Mukhopadhyay *et al*., 2006; Wiedemann and Pfanner, 2017).

The targeting of recombinant proteins to plant mitochondria could be useful when the aim is to modulate a mitochondrial function such as energy generation, iron–sulfur cluster assembly, developmental signals, or responses to biotic and abiotic stress (Pierrel *et al*., 2007; Atkin and Macherel, 2009). Furthermore, the low-oxygen mitochondrial environment is ideal for metabolic engineering with oxygensensitive enzymes (Curatti and Rubio, 2014; López-Torrejón *et al*., 2016) and the control of enzyme metalation, given the abundance of copper, iron, manganese and zinc in the mitochondrial matrix (Pierrel *et al.*, 2007; Pérez-González *et al*., 2017). However, one of the challenges involved in mitochondrial targeting is the complex structure of the organelle, which features a double membrane separated by an inner matrix, and folded internal cristae separated by a further membrane, allowing the localization and separation of proteins that require specific environments for their activity (Lill and Mühlenhoff, 2008). Nuclear-encoded mitochondrial proteins may therefore feature complex targeting peptides up to 90 amino acids in length carrying the information needed for precise localization within mitochondrial compartments or membranes (Huang *et al.*, 2009b). The presequences

comprise different groups of amino acids with distinct physicochemical properties and/or mitochondrial outer membrane recognition motifs (Fukasawa *et al*., 2015). These properties determine the ultimate destination of native and heterologous proteins within the various spaces and membranes of the mitochondrion (Dudek *et al*., 2013).

The direct experimental analysis of the plant mitochondrial proteome has been carried out predominantly in dicot species such as Arabidopsis (Kruft *et al*., 2001; Heazlewood *et al.*, 2004; Lee *et al*., 2012), tobacco (Huang *et al*., 1990; Allen *et al*., 2017) and pea (Bardel *et al*., 2002). In contrast, only a few reports have described the mitochondrial proteome of monocot species such as rice (Heazlewood *et al.*, 2004; Huang *et al*., 2009a) and maize (Hochholdinger *et al*., 2004). In rice, the analysis of 313 nuclear-encoded mitochondrial proteins using four different algorithms revealed that the correct mitochondrial location was predicted in only 60% of cases (Huang *et al*., 2009a). The number of mitochondrial proteins predicted to be involved in the electron transport chain, tricarboxylic acid cycle and stress responses was well conserved between rice and Arabidopsis (Heazlewood *et al*., 2004; Huang *et al*., 2011). The relatively low predictive accuracy of the localization algorithms highlights the importance of experimental validation, particularly when attempting to target recombinant proteins to the mitochondria.

The best way to determine the activity of mitochondrial targeting peptides is to test them using a visible marker, such as enhanced green fluorescent protein (eGFP) adapted for optimal activity in plants (Chiu *et al*., 1996; Snapp, 2005). This will confirm the potential suitability of the import sequence activity, although it does not guarantee the import of other proteins because the structure of the linked protein can also influence the efficiency of targeting peptides (Van Steeg *et al*., 1986). I therefore generated transgenic rice callus and plants in which nuclear-encoded recombinant eGFP was targeted to the mitochondria using six different mitochondrial N-terminal targeting peptides. These sequences were derived from diverse phylogenetic origins (higher and lower eukaryotes) and varied in their targeting probability scores, motifs and physicochemical properties. I analyzed the expression and localization of eGFP by immunoblot as well as confocal and electron microscopy in order to determine factors responsible for the effectiveness of mitochondrial targeting in rice.

2.2. Aims

- Target rice mitochondria using distinct targeting peptides and generate transgenic rice callus and plants.
- Analyze the expression and localization of eGFP by immunoblot as well as confocal and electron microscopy.
- Determine the factors responsible for the effectiveness of mitochondrial targeting in rice.

2.3. Materials and methods

2.3.1. *Expression constructs*

I selected six different mitochondrial pre-sequences representing different phylogenetic origins. Two of the sequences were from the yeast *Saccharomyces cerevisiae*: the alpha subunit of the F1 sector of the mitochondrial F1F0 ATP synthase protein ATPA (NCBI: NP_009453.2) and the cytochrome oxidase subunit IV protein COX4 (NCBI: NP_011328.1). The third sequence was from the ascomycete fungus *Neurospora crassa*: subunit 9 of the mitochondrial ATPase protein, SU9 (NCBI: XM_954801.3). The other three sequences were from plants. Two were derived from dicot species: the *Nicotiana plumbaginifolia* F1-ATPase from the β -subunit protein MTS2 (NCBI: X02868.1) and the *Arabidopsis thaliana* γ-subunit of the mitochondrial ATP synthase pFA (NCBI: NM_128864.4). The final sequence represented the endogenous rice (*Oryza sativa* ssp. *japonica*) succinyl-CoA synthetase protein β-chain OsSCSb (NCBI: Q6K9N6).

The expression plasmids were generated using the MoClo cloning system (Weber *et al.* 2011) and an in-house destination vector comprising the pUC57 vector backbone (GenScript, Piscataway, NJ, USA) joined to the cloning cassette of the Level 1 Position 2 MoClo vector pICH47742. The vector was amplified using primers 591 (5′-AAG CCC ACG AAG TGT GGG GTG CCT AAT GAG TGA GCT AAC TCA CA-3′) and 592 (5′-TTA ACA CAG AGT GGC CAG CCC CGA CAC CCG CCA ACA CCC G-3′) and the 2200-bp AdeI digestion product was ligated to the 650-bp fragment released from pICH47742 by digestion with the same enzyme. The recombinant in-house vector (pUC57-L1P2) is available from Addgene (ID: 109221). All the Level 1 Position 2

plasmids were constructed using the strong constitutive maize ubiquitin-1 promoter including the first intron and the nopaline synthase (*nos*) terminator to control eGFP expression.

After MoClo restriction/ligation, 20µl of the reaction mix was used to transform *Escherichia coli* DH5α competent cells. Positive clones were selected on lysogeny broth (LB) solid medium containing 100µg/ml ampicillin (Sigma-Aldrich, Darmstadt, Germany), 20µg/ml X-gal (Duchefa Biochemie, Haarlem, Netherlands) and 1mM isopropyl β-D-1-thiogalactopyranoside (IPTG; Sigma-Aldrich). Plasmid DNA was extracted using the GenElute Plasmid Miniprep Kit (Sigma-Aldrich). The final constructs were named Ubi1:SU9-eGFP-tNos, Ubi1:COX4-eGFP-tNos, Ubi1:MTS2-eGFP-tNos, Ubi1:pFA-eGFP-tNos, Ubi1:ATPA-eGFP-tNos and Ubi1:OsSCSb-eGFP-tNos. The integrity of all plasmids was verified by sequencing (Macrogen, Madrid, Spain).

2.3.2. *Transformation of rice callus and regeneration of transgenic plants*

The six test vectors containing the individual mitochondrial targeting peptide sequences fused to eGFP were introduced separately into rice embryos, together with the *hpt* gene for selection as described by Christou *et al*. (1991) and Sudhakar *et al*. (1998). We selected five representative independent callus lines and corresponding regenerated plants for each construct for all subsequent analyses from a population of at least 50 independent lines per construct.

2.3.3. *Protein extraction and immunoblot analysis*

Total rice protein extracts were prepared by grinding 0.1–0.2g callus or leaf tissue in liquid nitrogen and thawing the powder in 0.2–0.4mL of extraction buffer: 20mM Tris-HCl pH 7.5, 5mM ethylenediaminetetraacetic acid (EDTA), 0.1% Tween-20, 0.1% sodium dodecylsulfate (SDS), 2mM phenylmethanesulfonylfluoride (PMSF). The mixture was vortexed for 1h at 4 °C. Cell debris was removed by centrifugation at 15,000 x *g* for 20min at 4°C, and the supernatant was collected and stored at –80°C. The protein concentration in the supernatants was determined using the Bradford method (AppliChem, Darmstadt, Germany). I fractionated 80µg of total rice protein by denaturing SDS-PAGE in polyacrylamide gels containing 10% SDS at 200V for 60min, and then electro-transferred the protein to an Immobilon FL polyvinylidene difluoride (PVDF)

membrane (Merck, Darmstadt, Germany) using a semidry transfer apparatus (Bio-Rad, Hercules, CA, USA) at 20V for 45min. The membrane was immersed in 5% non-fat milk in Tris-buffered saline with Tween-20 (TBST) solution (0.2M Tris-HCl pH 7.6, 1.37M NaCl, 0.1% Tween-20) for 1h at room temperature. Membranes were incubated with anti-eGFP polyclonal antibody SAB4301138 (Sigma-Aldrich) diluted 1:2000 in 5% non-fat milk in TBST overnight at 4°C, then rinsed three times for 10min in TBST. The membranes were subsequently incubated with an alkaline phosphatase-conjugated goat anti-rabbit secondary antibody (Sigma-Aldrich) (diluted 1:5000 in 2% non-fat milk in TBS-T) for 1h at room temperature followed by three 10min rinses in TBS-T. Signals were detected using SIGMA*FAST* BCIP/NBT tablets (Sigma-Aldrich).

2.3.4. *Confocal microscopy*

Confocal microscopy was used to confirm the localization of eGFP in small pieces of rice callus ($1mm^3$) or leaf tissue (1 x 10mm) after incubation with the mitochondrial counterstain Mitotracker Red (Molecular Probes/Invitrogen, Paisley, UK) according to the manufacturer's recommendations. The tissues were then fixed with 2% paraformaldehyde in 0.1M sodium phosphate buffer (pH 7.2) and cut into semi-thin sections (30–40µm) using a CM3050S Research Cryostat (Leica Microsystems, Wetzlar, Germany). The sections were collected on standard glass microscope slides pre-coated with poly-L-lysine and images were captured using an FV1000 laser scanning confocal microscope (Olympus, Hamburg, Germany) with illumination at 488nm (excitation wavelength of eGFP, multiline argon laser) and 559nm (excitation wavelength of Mitotracker Red, diode laser). Five different callus lines and leaves (biological replicates) were analyzed per targeting peptide, and the percentage of merged mitochondria (where the green and red color signals co-localized) was counted in a minimum of three images taken from different areas/sections in each sample to determine the mitochondrial targeting efficiency of each peptide.

2.3.5. *Immuno-electron microscopy*

Small callus and leaf samples as described above were fixed in 1% glutaraldehyde and 1% paraformaldehyde in 0.1M sodium phosphate buffer (pH 7.2) for 16–24h at 4°C and washed three times (10min) with the same buffer. After fixation, samples were dehydrated in an ethanol series (30–100%) before embedding in Lowicryl K4M resin

(Polysciences, Hirschberg an der Bergstrasse, Germany) in a cold chamber at –20 to –35°C and inducing polymerization by exposure to ultraviolet light.

Semithin (2µm) and ultrathin (70–90nm) sections were prepared using a Reichert-Jung ultra-cut E cryotome (Leica Microsystems). The sections were stained with Richardson's blue, covered with a drop of DPX slide mounting medium and a coverslip, and observed under a DM4000B microscope (Leica Microsystems). Images were captured using a DFC300 FX 1.4-MP digital color camera equipped with LAS v3.8 (Leica Microsystems). The ultrathin sections were mounted on Formvar carbon-coated gold grids (200mesh) and incubated for 15min in blocking buffer for polyclonal antibodies (200mM Tris-HCl pH 7.4, 1% Tween-20, 0.1% gelatin, 1% BSA) or monoclonal antibodies (10mM Tris-HCl pH 7.4, 0.9% NaCl, 0.05% PEG 20,000, 3% BSA). The grids were then washed in distilled water and incubated overnight at 4°C with primary polyclonal anti-eGFP antibody PA5-22688 (Thermo Fisher Scientific, Waltham, MA, USA) diluted 1:200 in blocking buffer, or primary monoclonal anti-eGFP antibody 11814460001 (Sigma-Aldrich) diluted 1:500 in blocking buffer. I cross-adsorbed the polyclonal antibody following the protocol of Deena and Fletcher (1993). After washing in distilled water, followed by a further 30-min incubation in the appropriate blocking buffer and another wash, the grids were incubated at room temperature for 1h with the 15nm gold-conjugated secondary antibody diluted 1:20 in the appropriate blocking buffer: goat-anti-rabbit IgG for the polyclonal antibody, or EM-grade goat-anti-mouse IgG for the monoclonal antibody (Electron Microscopy Sciences, Hatfield, PA, USA). Finally, the grids were contrasted with 1% uranyl acetate in water (20min) and Reynold's lead citrate (2min) before observation in an EM 910 Transmission Electron Microscope (Zeiss, Oberkochen, Germany). I analyzed at least two grids per treatment and sample (callus or leaf tissue). More than 10 areas containing mitochondria were registered per treatment.

2.3.6. *Bioinformatic analysis*

Targeting peptide cleavage sites, probability scores and physicochemical properties were determined using MitoFates (http://mitf.cbrc.jp/MitoFates/cgibin/top.cgi), MitoProt (https://ihg.gsf.de/ihg/mitoprot.html) and TPpred v3.0 (https://tppred3.biocomp.unibo.it/tppred3) (Claros and Vincens, 1996; Fukasawa *et al.*, 2015; Savojardo *et al.*, 2015).

2.3.7. *Accession Numbers*

Sequence data from this article can be found in National Center for Biotechnology Information (NCBI) under the following accession numbers: ATPA (NCBI: NP_009453.2), COX4 (NCBI: NP_011328.1), SU9 (NCBI: XM_954801.3), MTS2 (NCBI: X02868.1), pFA (NCBI: NM_128864.4), OsSCSb (NCBI: Q6K9N6).

2.4. Results

2.4.1. *Probability scores and recognition motifs of selected mitochondrial targeting peptides*

Six mitochondrial targeting peptides were selected for analysis (ATPA, COX4, SU9, MTS2, pFA and OsSCSb), whose properties and phylogenetic origins are summarized in **Table 2.1**. Bioinformatics analysis revealed that all six targeting peptides are positively charged and achieve high targeting probability scores but feature different numbers of Tom20 recognition motifs and N terminal hexamer motifs, as described by Fukasawa *et al*. (2015). Hexamer motifs are enhancer motifs that interact with hydrophobic binding pockets on the Tom20 receptor and improve mitochondrial targeting effectiveness and accuracy (Obita *et al*., 2003; Fukasawa *et al*., 2015). The structural relationships among the various motifs are shown in **Figure 2.1**.

Table 2.1. Molecular and biochemical characteristics of the six mitochondrial targeting peptides.

Peptide	Source	Targeting probability	Size (aa)	Predicted cleavage position (aa)	No. of Tom20 recognition motifs	No. of N-terminal hexamer motifs
ATPA	*Saccharomyces cerevisiae*	99%	35	24	0	4
COX4	*Saccharomyces cerevisiae*	92%	29	17	0	1
SU9	*Neurospora crassa*	98%	69	35/67	2 (ASRLA, AVRVA)	3
MTS2	*Nicotiana plumbaginifolia*	97%	87	37/42	1 (LNRAV)	0
pFA	*Arabidopsis thaliana*	97%	77	30/42	2 (ITKAM, AMKMV)	1
OsSCSb	*Oryza sativa* ssp. *japonica*	86%	27	13	2 (LGKLA, ASRAL)	2

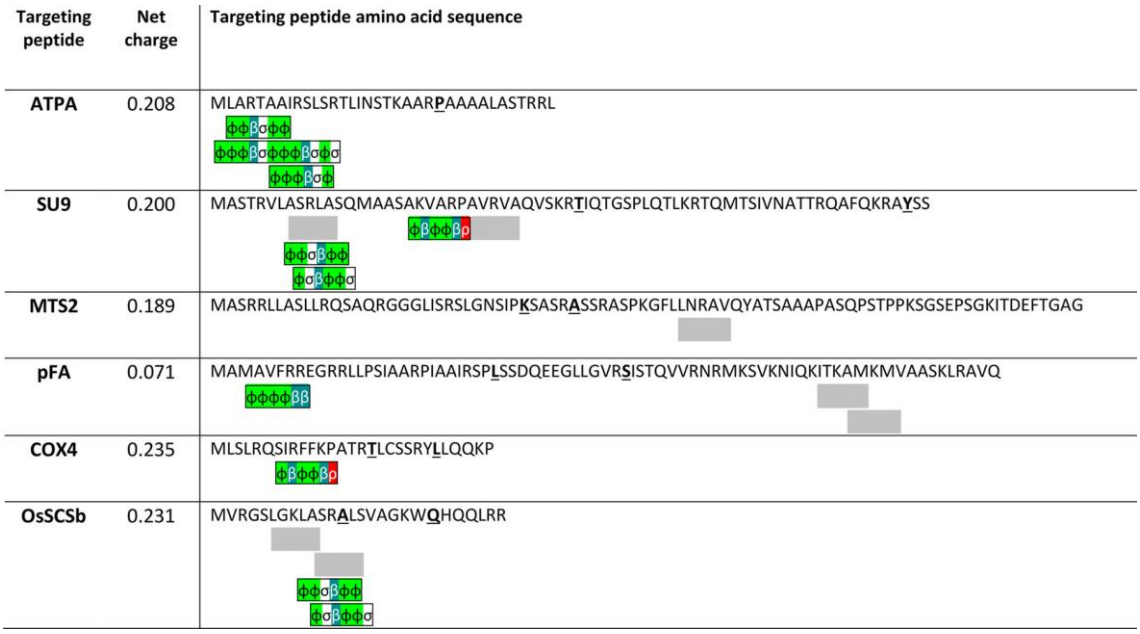


Figure 2.1. Properties of targeting peptides and the structural relationships among the various motifs. N-terminal hexamer motifs are described using the following symbols to denote groups of amino acids: φ = hydrophobic (L, F, I, V, W, Y, M, C or A), β = basic (R, K or H), α = acidic (E, D), σ = polar (S, T, N or Q), and ρ = secondary structure breaker (P or G). Gray boxes indicate Tom20 recognition motifs. Underlined amino acids are predicted cleavage sites.

2.4.2. *Expression of eGFP in rice*

The six mitochondrial targeting peptides we tested included three with single predicted cleavage sites (ATPA, COX4 and OsSCSb) and three with dual predicted cleavage sites (SU9, MTS2 and pFA) based on combined analysis using three bioinformatics programs: MitoFates, MitoProt and TPpred. The six recombinant eGFP constructs were expressed in rice. The callus and leaf extracts were analyzed by immunoblot to determine the size of the recovered eGFP products (**Fig. 2.2**).

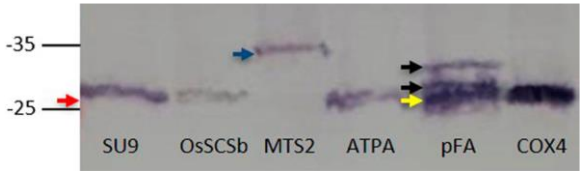


Figure 2.2. Analysis of six constructs targeting the mitochondria in rice (SU9-eGFP, OsSCSb-eGFP, MTS2-eGFP, ATPA-eGFP, pFA-eGFP and COX4-eGFP) by denaturing SDS-PAGE and western blot. The results represent transformed callus lines but identical profiles were seen in the leaves of transgenic plants. The red arrow (~27kDa) indicates the molecular weight of correctly processed eGFP. The blue arrow (~35 kDa) indicates eGFP with a non-cleaved MTS2 peptide. The black arrows (~30 and ~32kDa) indicate the products generated following the correct processing of the pFA peptide and the yellow arrow (~28 kDa) indicates cleavage closer to the native end of the eGFP.

The mature eGFP has a molecular weight of 27kDa, but all six targeting peptides have internal cleavage sites leaving a remnant on the processed protein, so we anticipated the mature processed products would be somewhat higher in molecular weight than native eGFP, depending on the length of the remnant (**Figure 2.1**). The ATPA, COX4 and OsSCSb variants of eGFP yielded bands of ~28kDa, consistent with successful cleavage of the targeting peptide and the presence of a remnant ranging in length from 11 to 14 residues. The SU9 variant of eGFP yielded a ~27kDa band, consistent with cleavage at position 67 of 69 and the presence of a comparatively small dipeptide remnant. For these four constructs, I therefore observed complete and successful cleavage of the targeting peptide in rice. The pFA sequence has predicted cleavage sites at positions 30 and 42 of 77, which should leave a remnant of at least 35 amino acids. Here I observed three distinct bands of ~28, ~30 and ~32kDa. The 30 and 32kDa bands reflected the anticipated sizes of the processed forms of pFA generated by cleavage at the predicted processing sites, whereas the other band indicated that the peptide is also cleaved closer to the native end of the eGFP than expected based on the predictions. Finally, the MTS2 targeting peptide has predicted cleavage sites at positions 37 and 42 of 87, but the ~35 kDa band I detected indicated that neither site was cleaved successfully (**Figure 2.2**).

2.4.3. *Localization of eGFP in rice and the effectiveness of the targeting peptides*

Having established that five of the six peptides were partially or fully cleaved, I analyzed the localization of eGFP in rice callus (**Figure 2.3**) and in the leaves of the corresponding regenerated plants (**Figure 2.4**) by confocal microscopy.

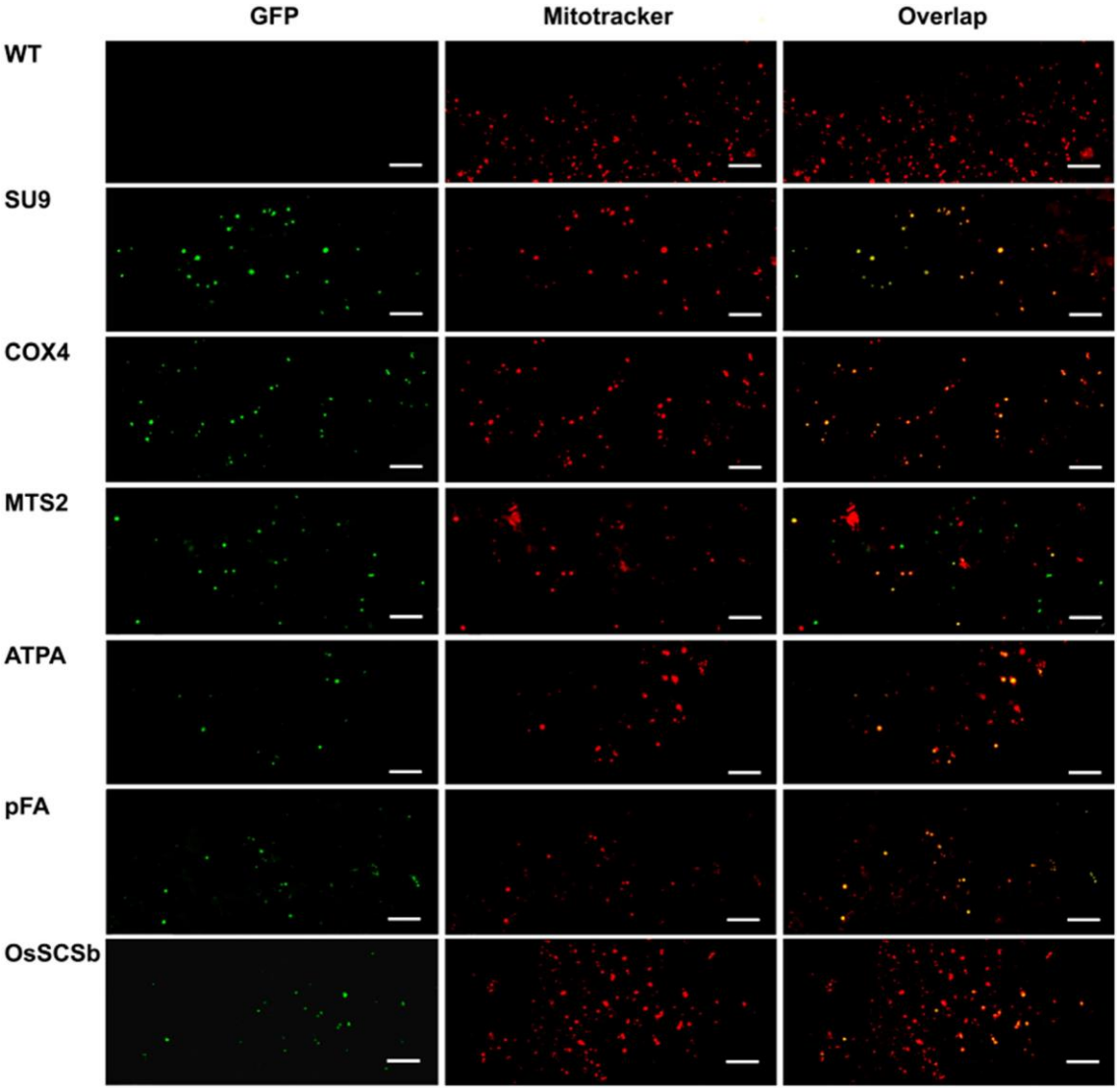


Figure 2.3. Confocal laser scanning microscopy images of wild-type (WT) rice callus and callus lines transformed with eGFP constructs linked to the SU9, COX4, MTS2, ATPA, pFA and OsSCSb mitochondrial targeting peptides. The three columns show the individual signals for eGFP (green fluorescence) and Mitotracker Red (red fluorescence) and the overlap (merged confocal image mixing green and red fluorescence) (bars = 20µm). Five different callus lines and leaves (biological replicates) were analyzed per targeting peptide, and the percentage of merged mitochondria (where the green and red color signals co-localized) was counted in a minimum of three images taken from different areas/sections in each sample to determine the mitochrondrial targeting efficiency of each peptide.

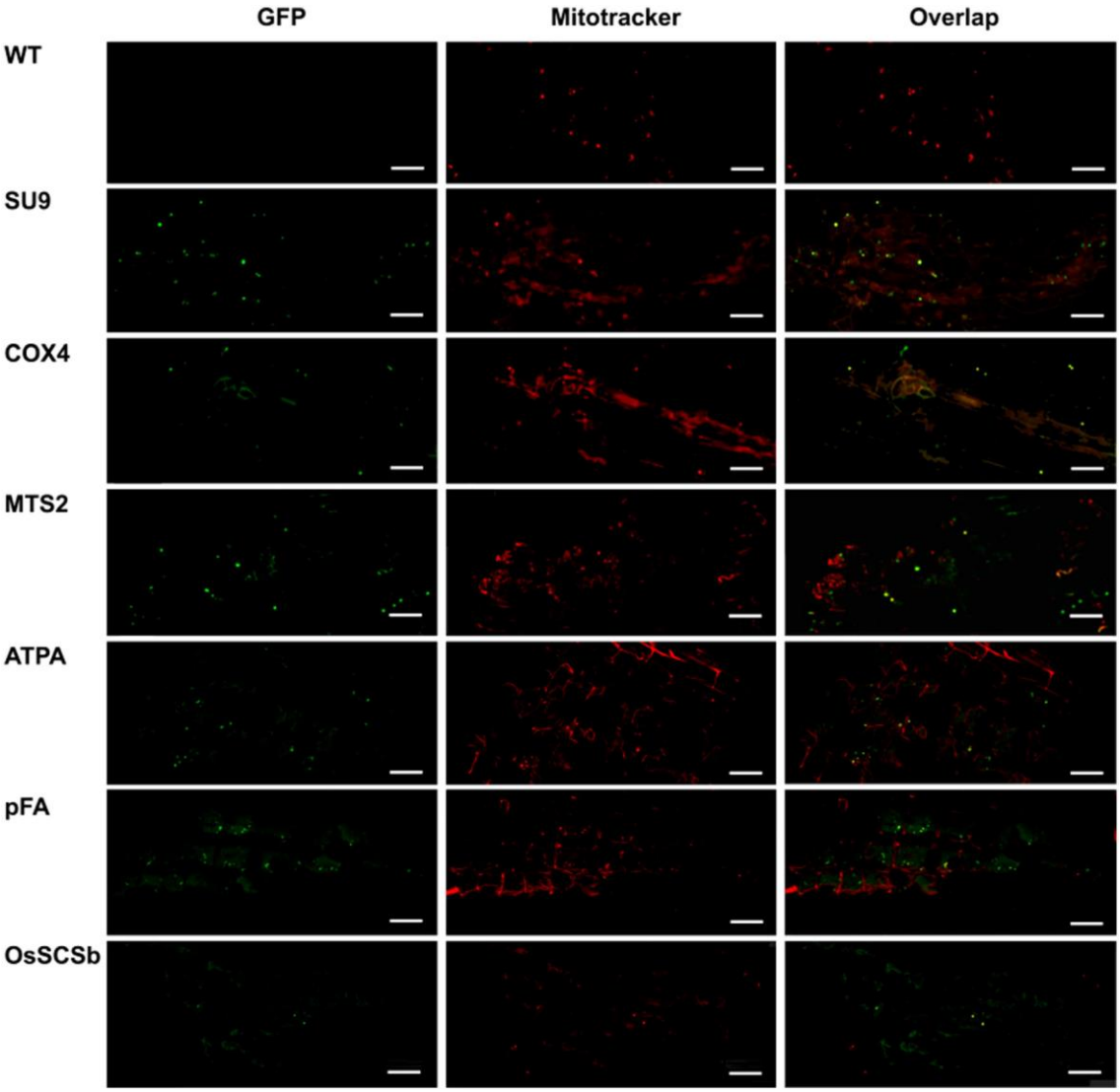


Figure 2.4. Confocal laser scanning microscopy images of wild-type (WT) rice leaves and leaves from transgenic lines transformed with eGFP constructs linked to the SU9, COX4, MTS2, ATPA, pFA and OsSCSb mitochondrial targeting peptides. The three columns show the individual signals for eGFP (green fluorescence) and Mitotracker Red (red fluorescence) and the overlap (merged confocal image mixing green and red fluorescence) (bars = 20µm).

The penetration of the Mitotracker Red marker was more effective in callus than in leaves giving clearer results, but in both cases the green fluorescence of the eGFP and the red fluorescence of the Mitotracker Red marker co-localized in more than 70% of the merged images of tissues transformed with the ATPA, COX4, SU9, pFA and OsSCSb variants of eGFP, whereas the two signals were co-localized in fewer than 15% of the merged images of tissues transformed with the MTS2-eGFP construct. Immunogold labeling was carried out in callus and leaf tissue expressing four of the six constructs (two with the correct, anticipated cleavage patterns as well as pFA with the additional cleavage product and

MTS2 which did not show evidence of cleavage). The 15nm gold particles were strongly associated with the inner mitochondrial membrane and/or mitochondrial matrix in callus (**Figure 2.5**) and leaf (**Figure 2.6**) tissues transformed with the COX4, SU9 and pFA variants, whereas in tissues transformed with the MTS2-eGFP construct the labeling was mostly restricted to the cytosol. For all four targeting peptides, I also observed strongly-labeled protein bodies in the cytosol (**Figure 2.7**), which is consistent with earlier observations of GFP-prolamin fusions in rice (Saito *et al.*, 2009; Shigemitsu *et al*., 2013). These signals most likely represent eGFP aggregates, which I detected in the vacuoles of callus and leaf tissues expressing each of the four constructs (**Figure 2.7**). The ectopic formation of protein bodies in the leaves is likely to reflect the use of a strong, constitutive promoter (Saberianfar *et al*., 2016). The aggregates were larger in callus samples expressing MTS2-eGFP compared to those expressing the COX4, SU9 and pFA variants (**Figures 2.8** and **2.9**). Very occasionally, I observed nonspecific labeling of the nucleus in tissues expressing each of the four constructs (**Figure 2.10**).

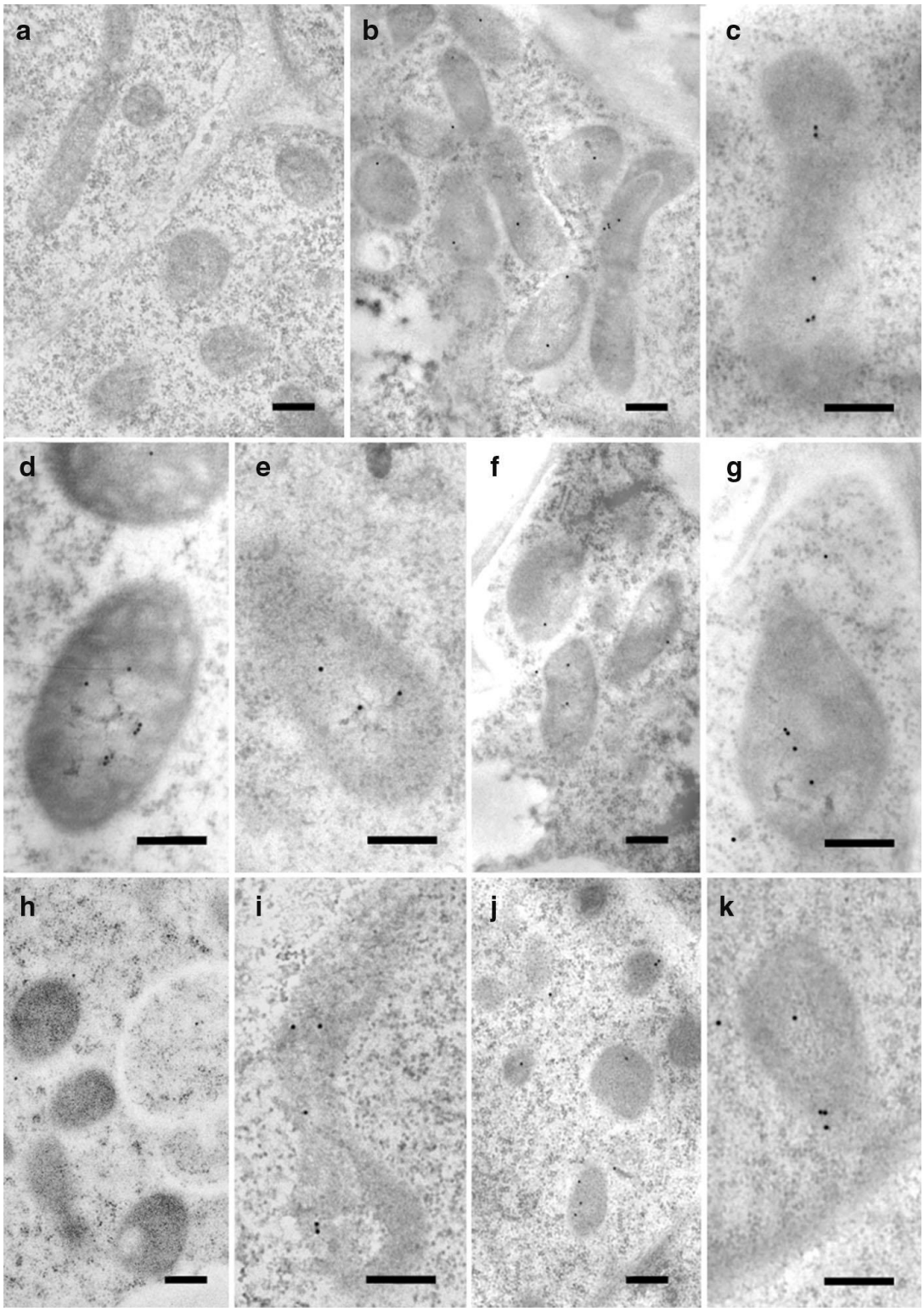


Figure 2.5. Immunogold labeling of eGFP in the mitochondria of rice callus cells using a GFP-specific monoclonal antibody (diluted 1:200) (A–C, F–K) or eGFP polyclonal antibody (diluted 1:500) (D = non-cross adsorbed, E = cross-adsorbed). Monoclonal antibodies were used to confirm the localization profile revealed by the cross-adsorbed polyclonal antibody although the signal produced by the monoclonal antibody was weaker. A. Wild-type cells. B–E. SU9-eGFP (B, C and E = monoclonal antibody, D = polyclonal antibody). F–G. Cox4-eGFP is also labeled outside the mitochondria. H–I. MTS2-eGFP is labeled in only a few mitochondria but also in the cytosol. J–K. pFA-eGFP is labeled inside the mitochondria (bars = 200nm; gold particle size = 15nm).

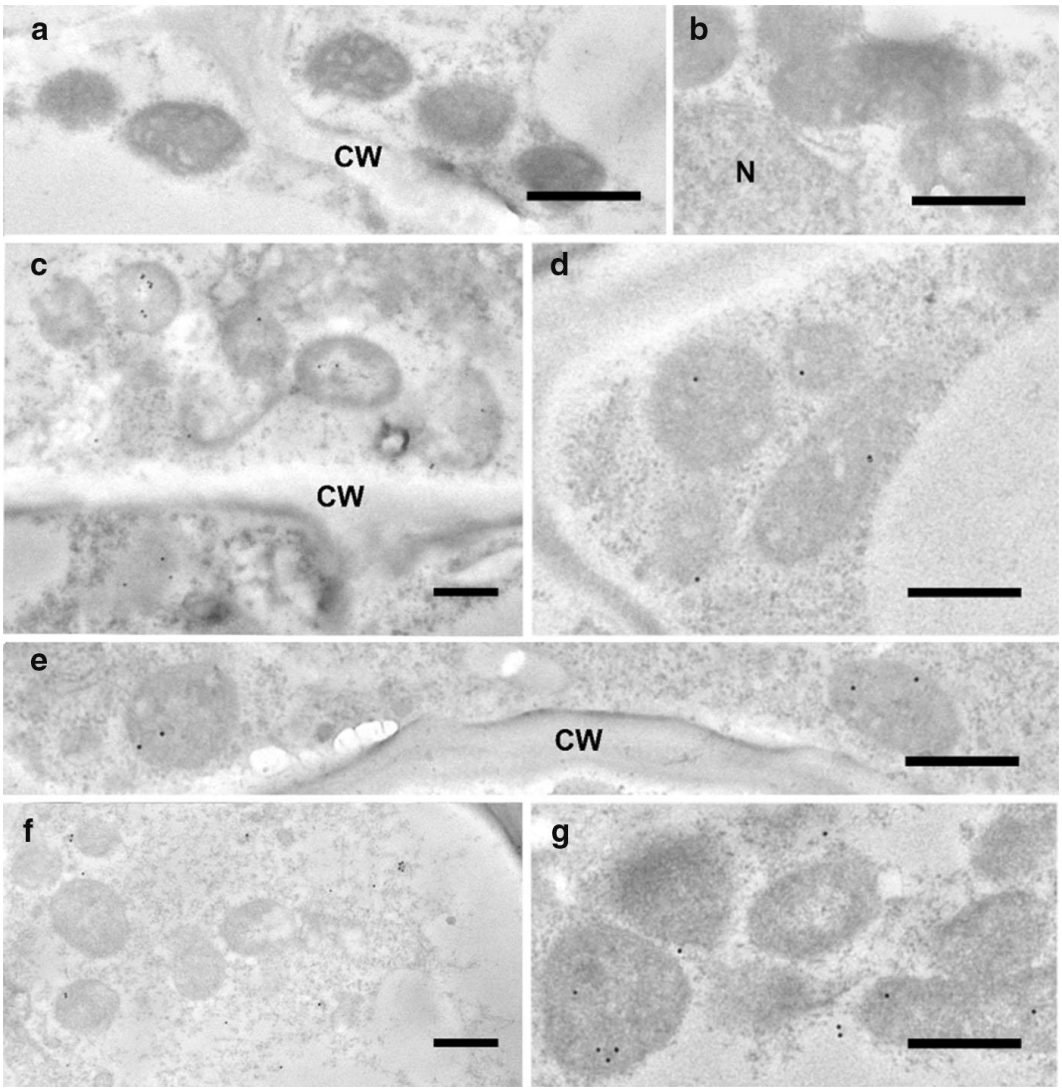


Figure 2.6. Immunogold labeling of eGFP in the mitochondria of rice leaf cells using a GFP-specific monoclonal antibody (diluted 1:200). A–B = wild-type, C–D = SU9-eGFP, E = pFA-eGFP, F = MTS2-eGFP, G = Cox4-eGFP (CW = cell wall, N = nucleus; bars = 500nm; gold particle size = 15nm).

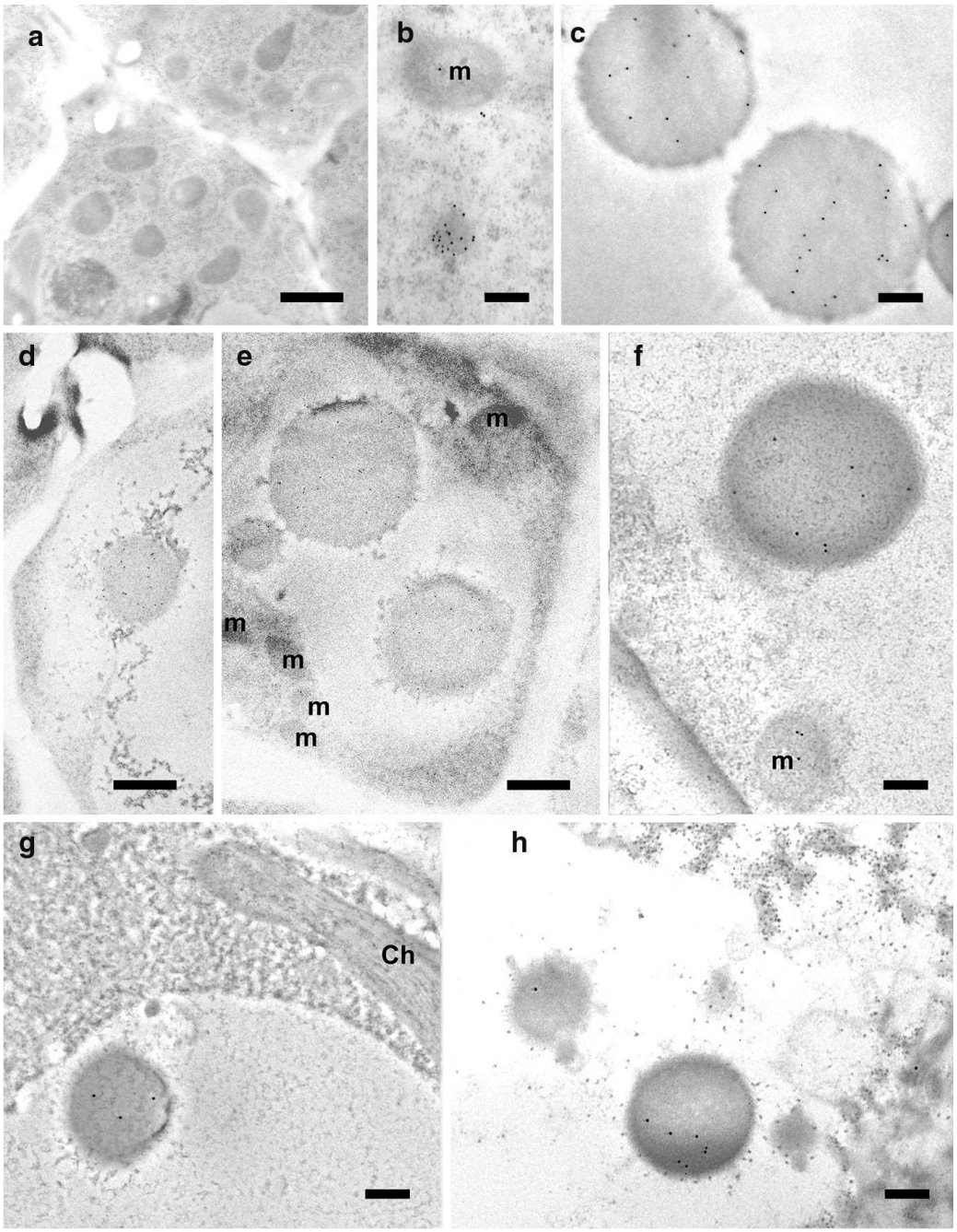


Figure 2.7. Immunogold labeling of protein bodies in rice callus and leaf cells using a GFP-specific monoclonal antibody (diluted 1:200). A. Wild-type negative control, showing no protein bodies. B. Labeled protein body in the cytosol of callus cells expressing Cox4-eGFP. Labeled protein bodies in the vacuole of callus cells (C) and leaf cells (G-H) expressing SU9-eGFP. D–E. Large protein bodies close to (D) and inside (E) the vacuoles of callus cells expressing MTS2-eGFP. F. Labeled protein body in callus cells expressing pFA-eGFP (m = mitochondria; bars A, G and H = 1µm, B–F = 200nm; gold particle size = 15nm).

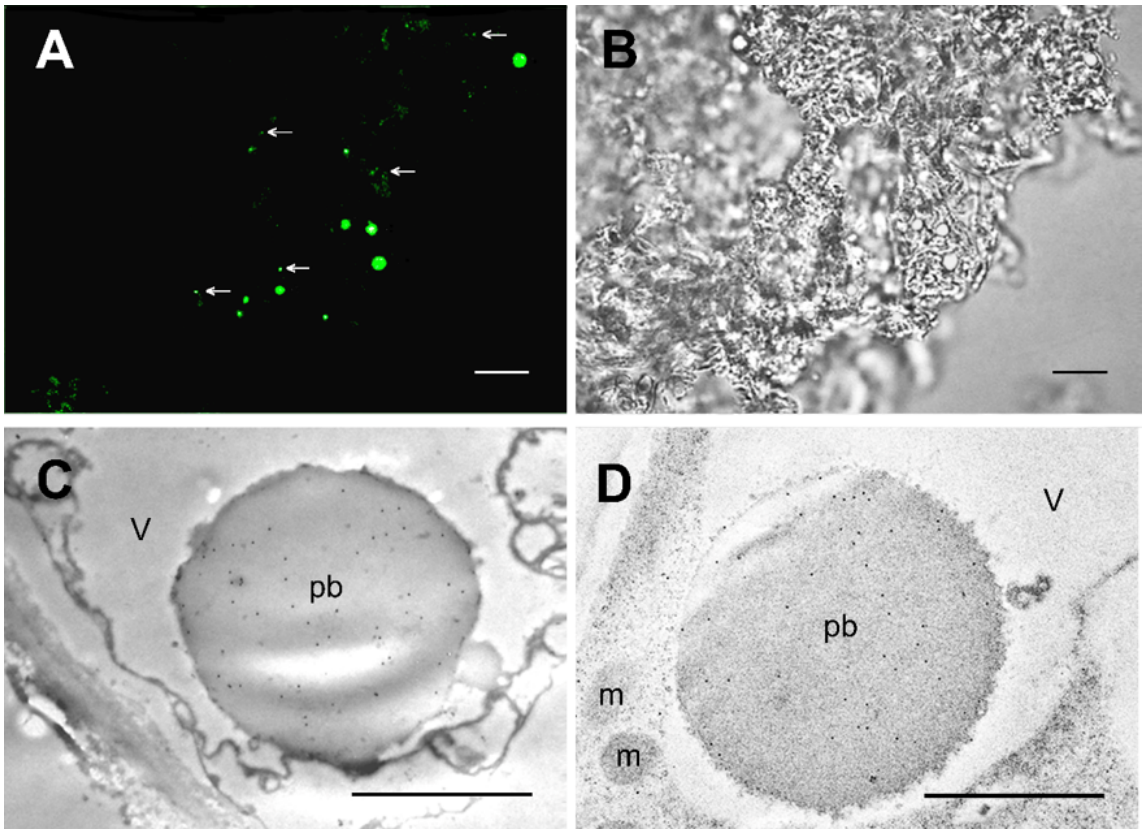


Figure 2.8. Confocal and light microscopy images of callus cells expressing MTS2-eGFP. A. Confocal laser scanning microscopy image, with arrows highlighting the mitochondria. B. Light microscopy image. C-D. Immuno-electron microscopy images showing the detection of eGFP using a polyclonal antibody (diluted 1:500) (m = mitochondria, pb = protein body; bars A-B = 20µm, C-D = 2µm; gold particle size = 15nm).

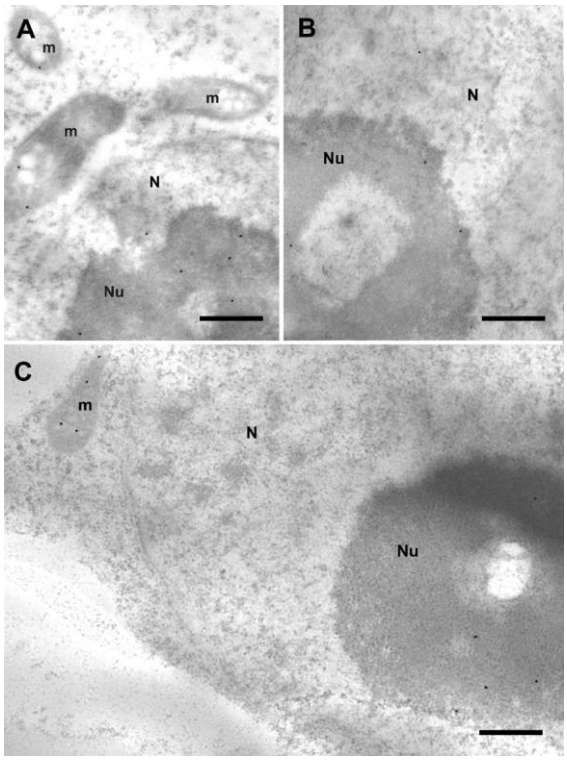


Figure 2.9. Immunogold labeling of the nucleus (N) in rice callus cells using a GFP-specific monoclonal antibody (diluted 1:200). Note the specific labeling of mitochondria (m) and the nucleus/nucleolus in A (SU9-eGFP), B (Cox4-eGFP) and C (pFA-eGFP) (bars = 500nm; gold particle size = 15nm).

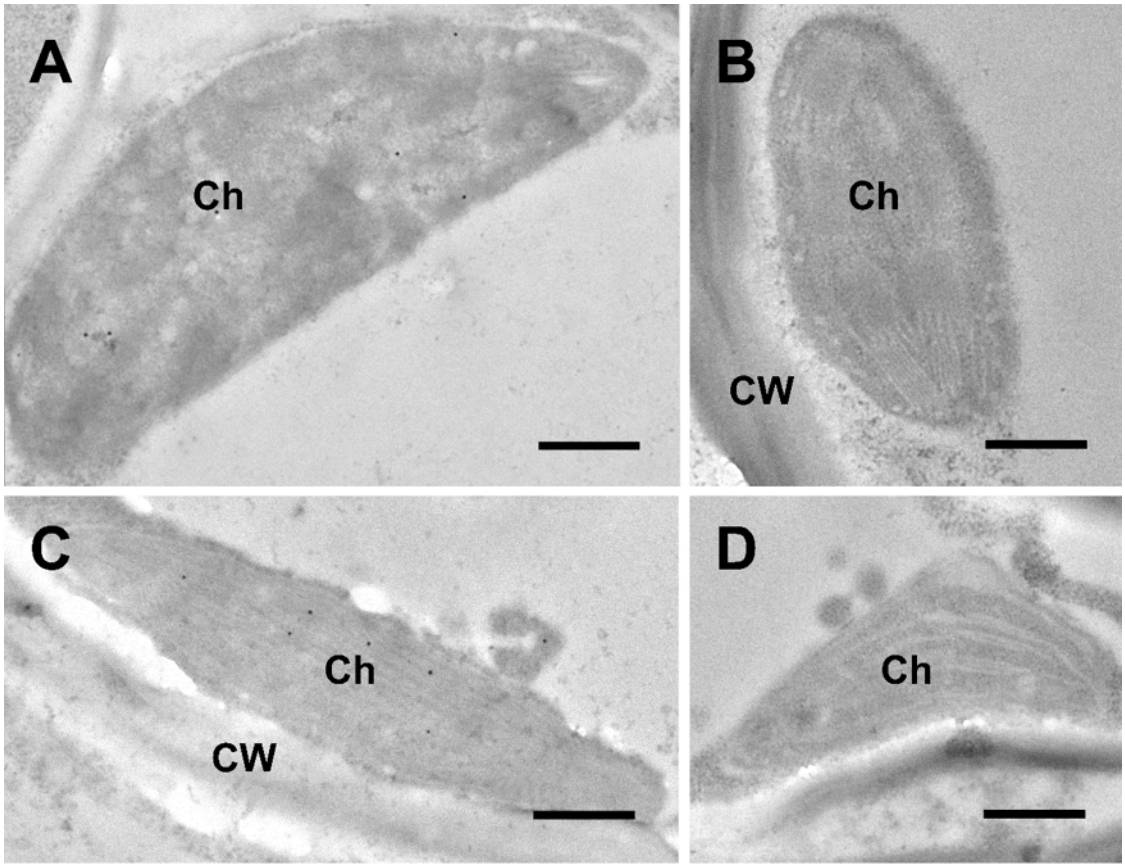


Figure 2.10. Non-specific immunogold labeling of chloroplasts in rice leaf cells using eGFP polyclonal cross-adsorbed and non-cross-adsorbed antibodies (diluted 1:500). A. Wild-type leaf cells treated with the non-cross-adsorbed antibody (1:500) show the labeling of chloroplasts. B. Wild-type leaf cells treated with the cross-adsorbed antibody showing no labeling of chloroplasts. C. Leaf cells expressing SU9-eGFP treated with the non-cross-adsorbed antibody (1:500). D. Leaf cells expressing SU9-eGFP treated with the cross-adsorbed antibody (bars = 500nm; gold particle size = 15nm).

2.5. Discussion

Proteins targeted for mitochondrial import may be directed to the inner matrix, the inner or outer membranes, or the intermembrane space (Chacinska *et al*., 2009; Weis *et al*., 2013). Most mitochondrial proteins are encoded by the nuclear genome, synthesized on cytosolic ribosomes and ultimately translocated to the appropriate sub-mitochondrial location (Taylor and Pfanner, 2004; Dolezal *et al.*, 2006). The mechanism that controls protein import into mitochondria involves an N-terminal pre-sequence or targeting peptide which directs the protein to the correct internal location, and the peptide is removed by proteolytic cleavage when the protein reaches its destination (Huang *et al*., 2009a; Carrie *et al.*, 2015).

Mitochondrial targeting peptides tend to feature a high content of hydrophobic and positively charged amino acid residues, a near absence of negatively charged residues, and a very low abundance of acidic amino acids (Berglund *et al*., 2009). Mitochondrial pre-sequences have the tendency to form an amphiphilic α-helix, which interacts with the import receptor Tom20 (Pfanner and Geissler, 2001; Endo and Kohda, 2002), a 20 kDa subunit of the outer mitochondrial membrane complex translocase (Saitoh *et al.,* 2007). Tom20 recognizes the same motifs in plants and yeast, conventionally represented as φxxφφ (where φ is a hydrophobic amino acid and x is any amino acid). However, there are significant differences in the topology, localization, amino acid variations, binding motifs and number of Tom20 proteins in different organisms (Obita *et al*., 2003; Saitoh *et al*., 2007).

The Tom20 protein in fungi and animals is anchored via its N-terminus to the mitochondrial outer membrane and therefore exposes its C-terminal domain to the cytosol. In contrast, Tom20 in plants is anchored to mitochondria via its C-terminus and the N-terminal domain is exposed. Accordingly, plant Tom20 proteins are not orthologous to those of fungi and animals, and plant mitochondria also lack the other two receptor components that have been functionally characterized in yeast and mammalian systems, namely Tom70 and Tom22 (Macasev *et al*., 2004; Lister *et al*., 2007; Huang *et al.,* 2009a; Rimmer *et al,*. 2011). Previous studies have shown that plant Tom20 proteins interact with the amino acid domains LRTLA and LRRFV in rice superoxide dismutase and Arabidopsis threonyl tRNA synthetase (Zhang *et al.,* 2010). Even though Tom20 recognition motifs have been well characterized, they are not essential for mitochondrial protein import (Mukhopadhyay *et al.,* 2006; Lister *et al*., 2007; Lee *et al.,* 2012). For example, when the Tom20 recognition site of pFA was deleted, this reduced the efficiency of mitochondrial import in Arabidopsis by only 20% (Lee *et al.,* 2012). However, there are four Tom20 paralogs in the Arabidopsis genome, three of which are known to be functional, whereas rice (like yeast) has only a single Tom20 gene (Lister *et al*., 2007). Despite the major differences between Tom20 proteins in plants and fungi, mitochondrial targeting peptides from yeast and filamentous fungi are functional in plants. I tested three such peptides, namely ATPA and COX4 from *S. cerevisiae* and SU9 from *N. crassa*. Previously, all three peptides have been shown to successfully direct recombinant proteins to mitochondria in yeast, tobacco and Arabidopsis (Nelson *et al*., 2007; Burén *et al.,* 2017; Pérez-González *et al*., 2017) and similarly I found that all three were able to

direct eGFP to the mitochondria in rice. Interestingly, only one of these peptides features a Tom20-recognition motif but all three contain N-terminal hexamers. This provides more evidence that Tom20-recognition motifs are not strictly required for mitochondrial import in plants.

I also tested three plant-derived targeting peptides, including the Arabidopsis pFA sequence, which was mentioned briefly above. The others were the OsSCSb peptide from rice and the MTS2 sequence from *N. plumbaginifolia*. Lee *et al*. (2012) showed that Arabidopsis pFA contains multiple sequence motifs that target different mitochondrial compartments. In addition to the IAARP Tom20-recognition motif, the sequences DQEEG and VVRNR are involved in translocation across the mitochondrial membrane, and the sequences RLLPS and SISTQ pull proteins into the matrix. In another study, the Arabidopsis pFA peptide efficiently imported 16 nitrogenase protein components into the mitochondrial matrix of *Nicotiana benthamiana* in transient expression experiments, as well as a GFP fusion protein. A ∼30 kDa band was observed corresponding to the correctly processed GFP as well as a less abundant ∼28 kDa band representing an alternatively processed form or a degradation product (Allen *et al*., 2017). Similarly, I found that the Arabidopsis pFA targeting peptide was sufficient for the mitochondrial import of eGFP in rice, yielding three different bands of ∼28, ∼30 and ∼32 kDa. The bioinformatics analysis suggested that the 30 and 32 kDa bands represented the expected processed forms of the pFA-eGFP fusion and the remaining band was an alternatively processed form or a degradation product. This strongly suggests that the pFA peptide works as effectively in rice (this study) as it does in dicot species (Lee *et al*., 2012; Allen *et al.,* 2017). I tested the OsSCSb peptide based on an earlier report with tentative bioinformatics and mass spectrometry data predicting its mitochondrial targeting ability (Huang *et al.*, 2009b). The data supported this prediction by confirming the mitochondrial import of eGFP, the first time this sequence has been shown directly to function as a mitochondrial import peptide. Notably, both pFA and OsSCSb contain Tom20-binding motifs and N-terminal hexamers.

The *N. plumbaginifolia* MTS2 sequence was one of the first mitochondrial targeting peptides to be characterized in plants and has been shown to work in tobacco, although the authors did not present detailed microscopic analysis (Boutry and Chua, 1985; Chaumont *et al*., 1994). Surprisingly, we found that the MTS2 sequence appeared

insufficient for the mitochondrial import of eGFP in rice, and immunoblot experiments provided evidence that the protein was not cleaved in rice cells. Among the six peptides I tested, MTS2 was the only one that contained at least one Tom20-binding motif but no N-terminal hexamers, indicating that the latter may be necessary for mitochondrial import in rice. Similarly, mutations in the N-terminal hexamers of potato formate dehydrogenase were previously shown to block the mitochondrial import of GFP fusion proteins (Ambard-Bretteville *et al*., 2003). In contrast, Gnanasambandam *et al.* (2008) showed in a transient expression system that the MTS2 peptide efficiently targeted GFP to mitochondria in several plants: five monocots (sugarcane, wheat, corn, sorghum and onion) and seven dicots (cucumber, cauliflower, tomato, capsicum, pumpkin, coriander and sunflower). The MTS2 peptide therefore appears to be strongly influenced by the choice of cargo protein and host species, and detailed microscopy should be carried out to confirm the correct targeting of recombinant proteins using this sequence.

Confocal microscopy alone may not be sufficient to draw definitive conclusions about the efficiency or accuracy of targeting peptides in stably transformed intact plants. The confocal microscopy data indicated efficient mitochondrial import in rice callus and leaf tissues, but detailed analysis by immuno-electron microscopy revealed a certain level of non-specific labeling in the cytoplasm (**Figure 2.7**). This may reflect the formation of protein body-like structures in response to the high levels of recombinant protein synthesis, as discussed by Gutiérrez *et al*. (2013). Many previous studies have reported that protein bodies can be induced in rice and wheat callus and seeds (Arcalis *et al*., 2004; Saito *et al*., 2009; Takaiwa *et al*., 2009; Shigemitsu *et al*., 2012). Protein bodies do not usually form naturally in leaves, but they can be induced by the high-level expression of recombinant proteins, for example in *N. benthamiana* (Saberianfar *et al*. 2016). Protein bodies form mainly in the endoplasmic reticulum but may bud off into the cytosol as unattached organelles or may be removed to the vacuole (Levanony *et al*., 1992). In wheat, storage proteins aggregate into protein bodies and are then transported to the vacuoles without passing through the Golgi complex (Levanony *et al*., 1992). I also observed non-specific labeling in the nucleus (**Figure 2.9**) and in the chloroplasts (**Figure 2.10**). The non-specific labeling in the nucleus is likely to be genuine, and probably occurs because eGFP is small enough (~27 kDa) to enter the nucleus by free bidirectional diffusion through the nuclear pore complex, which has a size exclusion limit of 40–60 kDa (Köhler *et al*., 1997; Wei *et al*., 2003; Seibel *et al*., 2007). The non-specific labeling

in the chloroplast is more likely to be an experimental artifact. Berglund *et al*. (2009) previously reported that mitochondrial and chloroplast targeting peptides are similar enough that mitochondrial proteins can be inadvertently directed to the plastids. However, I found that the erroneous labeling of plastids could be eliminated by cross-adsorbing the polyclonal antibody with wild-type leaf extract (**Figure 2.10**) suggesting that the fidelity of mitochondrial and plastid targeting *in vivo* is preserved. The absence of labeling in the plastids was confirmed by using a monoclonal antibody, although the signal was not as intense as that generated by the cross-adsorbed polyclonal antibody (**Figure 2.5**).

2.6. Conclusions

Mitochondrial protein import is a complex, multistep process which is dependent on the recognition of targeting peptides before translocation and their subsequent cleavage. N-terminal hexamer motifs and Tom20-recognition motifs are thought to enhance mitochondrial targeting efficiency and accuracy. However, detailed analysis of the sequence motifs in each targeting peptide is necessary to understand the interactions between targeting peptides and translocator components during protein import. Here I demonstrated that mitochondrial import of nuclear-encoded eGFP in rice requires only a single N-terminal hexamer motif whereas the Tom20-binding motifs are not strictly required, at least based on the evidence from the six targeting peptides were tested. Only one of six targeting peptides that did not function in rice in my experiments features a single Tom20-binding motif but lacks an N-terminal hexamer. The efficiency of mitochondrial import in rice is difficult to quantify but can be determined by combining stable gene expression with subcellular localization by immunogold labeling. Finally, the presence of particular functional motifs, specifically the N-terminal hexamer, is likely to be more important for successful mitochondrial import in rice than the phylogenetic origin of the targeting peptide.

2.7. References

Akopian, D., Shen, K., Zhang, X., & Shan, S. O. (2013). Signal recognition particle: an essential protein-targeting machine. *Annual Review of Biochemistry*, 82, 693–721.

Allen, R. S., Tilbrook, K., Warden, A. C., Campbell, P. C., Rolland, V., Singh, S. P., & Wood, C. C. (2017). Expression of 16 Nitrogenase Proteins within the Plant Mitochondrial Matrix. *Frontiers in Plant Science*, 8, 287.

Ambard-Bretteville, F., Small, I., Grandjean, O., & Colas des Francs-Small, C. (2003). Discrete mutations in the presequence of potato formate dehydrogenase inhibit the in vivo targeting of GFP fusions into mitochondria. *Biochemical and Biophysical Research Communications*, 311, 966–971.

Arcalis, E., Marcel, S., Altmann, F., Kolarich, D., Drakakaki, G., Fischer, R., Christou, P., & Stoger, E. (2004). Unexpected deposition patterns of recombinant proteins in post-endoplasmic reticulum compartments of wheat endosperm. *Plant Physiology*, 136, 3457–3466.

Atkin, O. K., & Macherel, D. (2009). The crucial role of plant mitochondria in orchestrating drought tolerance. *Annals of Botany*, 103, 581–597.

Bardel, J., Louwagie, M., Jaquinod, M., Jourdain, A., Luche, S., Rabilloud, T., Macherel, D., Garin, J., & Bourguignon, J. (2002). A survey of the plant mitochondrial proteome in relation to development. *Proteomics*, 2, 880–898.

Berglund, A. K., Spånning, E., Biverståhl, H., Maddalo, G., Tellgren-Roth, C., Mäler, L., & Glaser, E. (2009). Dual targeting to mitochondria and chloroplasts: characterization of Thr-tRNA synthetase targeting peptide. *Molecular Plant*, 2, 1298–1309.

Boutry, M., & Chua, N. H. (1985). A nuclear gene encoding the beta subunit of the mitochondrial ATP synthase in *Nicotiana plumbaginifolia*. *The EMBO Journal*, 4, 2159–2165.

Brix, J., Dietmeier, K., & Pfanner, N. (1997). Differential recognition of preproteins by the purified cytosolic domains of the mitochondrial import receptors Tom20, Tom22, and Tom70. *The Journal of Biological Chemistry*, 272, 20730–20735.

Burén, S., Jiang, X., López-Torrejón, G., Echavarri-Erasun, C., & Rubio, L. M. (2017). Purification and *In Vitro* Activity of Mitochondria Targeted Nitrogenase Cofactor Maturase NifB. *Frontiers in Plant Science*, 8, 1567.

Carrie, C., Venne, A. S., Zahedi, R. P., & Soll, J. (2015). Identification of cleavage sites and substrate proteins for two mitochondrial intermediate peptidases in *Arabidopsis thaliana*. *Journal of Experimental Botany*, 66, 2691–2708.

Chacinska, A., Koehler, C. M., Milenkovic, D., Lithgow, T., & Pfanner, N. (2009). Importing mitochondrial proteins: machineries and mechanisms. *Cell*, 138, 628–644.

Chaumont, F., Silva Filho, M., Thomas, D., Leterme, S., & Boutry, M. (1994). Truncated presequences of mitochondrial F1-ATPase beta subunit from Nicotiana plumbaginifolia transport CAT and GUS

proteins into mitochondria of transgenic tobacco. *Plant Molecular Biology*, 24, 631–641.

Chiu, W., Niwa, Y., Zeng, W., Hirano, T., Kobayashi, H., & Sheen, J. (1996). Engineered GFP as a vital reporter in plants. *Current Biology: CB*, 6, 325–330.

Christou, P., Ford, T. & Kofron, M. (1991). Production of transgenic Rice (*Oryza sativa* L.) plants from agronomically important indica and japonica varieties via electric discharge particle acceleration of exogenous DNA into tmmature zygotic embryos. *Nature Biotechnology*, 9, 957–962.

Claros, M. G., & Vincens, P. (1996). Computational method to predict mitochondrially imported proteins and their targeting sequences. *European Journal of Biochemistry*, 241, 779–786.

Curatti, L., & Rubio, L. M. (2014). Challenges to develop nitrogen-fixing cereals by direct nif-gene transfer. *Plant Science*, 225, 130–137.

Deena, E., & Fletcher, J. (1993). Production of monospecific polyclonal antibodies against aster yellows mycoplasma like organism-associated antigen. *American Phytopathological Society*, 83, 1279–1282.

Dolezal, P., Likic, V., Tachezy, J., & Lithgow, T. (2006). Evolution of the molecular machines for protein import into mitochondria. *Science*, 313, 314–318

Dudek, J., Rehling, P., & van der Laan, M. (2013). Mitochondrial protein import: common principles and physiological networks. *Biochimica et Biophysica Acta*, 1833, 274–285.

Egea, P. F., Stroud, R. M., & Walter, P. (2005). Targeting proteins to membranes: structure of the signal recognition particle. *Current Opinion in Structural Biology*, 15, 213–220.

Endo, T., & Kohda, D. (2002). Functions of outer membrane receptors in mitochondrial protein import. *Biochimica et Biophysica Acta*, 1592, 3–14.

Endo, T., Yamano, K., & Kawano, S. (2011). Structural insight into the mitochondrial protein import system. *Biochimica et Biophysica Acta*, 1808, 955–970.

Fukasawa, Y., Tsuji, J., Fu, S. C., Tomii, K., Horton, P., & Imai, K. (2015). MitoFates: improved prediction of mitochondrial targeting sequences and their cleavage sites. *Molecular & Cellular Proteomics*, 14, 1113–1126.

Gnanasambandam, A., Anderson, D. J., Purnell, M. P., Nielsen, L. K., & Brumbley, S. M. (2008). The N-terminal presequence from F_1-ATPase β-subunit of Nicotiana plumbaginifolia efficiently targets green fluorescent fusion protein to the mitochondria in diverse commercial crops. *Functional Plant Biology*, 35, 166–170.

Gutiérrez, S. P., Saberianfar, R., Kohalmi, S. E., & Menassa, R. (2013). Protein body formation in stable transgenic tobacco expressing elastin-like polypeptide and hydrophobin fusion proteins. *BMC Biotechnology*, 13, 40.

Heazlewood, J. L., Tonti-Filippini, J. S., Gout, A. M., Day, D. A., Whelan, J., & Millar, A. H. (2004). Experimental analysis of the Arabidopsis mitochondrial proteome highlights signaling and regulatory components, provides assessment of targeting prediction programs, and indicates plant-specific mitochondrial proteins. *The Plant Cell*, 16, 241–256.

Hochholdinger, F., Guo, L., & Schnable, P. S. (2004). Cytoplasmic regulation of the accumulation of nuclear-encoded proteins in the mitochondrial proteome of maize. *The Plant Journal*, 37, 199–208

Huang, J., Hack, E., Thornburg, R. W., & Myers, A. M. (1990). A yeast mitochondrial leader peptide functions in vivo as a dual targeting signal for both chloroplasts and mitochondria. *The Plant Cell*, 2, 1249–1260.

Huang, S., Taylor, N. L., Narsai, R., Eubel, H., Whelan, J., & Millar, A. H. (2009a). Experimental analysis of the rice mitochondrial proteome, its biogenesis, and heterogeneity. *Plant Physiology*, 149, 719–734.

Huang, S., Taylor, N. L., Whelan, J., & Millar, A. H. (2009b). Refining the definition of plant mitochondrial presequences through analysis of sorting signals, N-terminal modifications, and cleavage motifs. *Plant Physiology*, 150, 1272–1285.

Huang, S., Millar, A.H., & Taylor, N.L. (2011). The plant mitochondrial proteome composition and stress response: conservation and divergence between monocots and dicots. In: Kempken F (ed) *Advances in Plant Biology*. Springer, New York, p 533.

Kim, D. H., & Hwang, I. (2013). Direct targeting of proteins from the cytosol to organelles: the ER versus endosymbiotic organelles. *Traffic*, 14, 613–621.

Köhler, R. H., Zipfel, W. R., Webb, W. W., & Hanson, M. R. (1997). The green fluorescent protein as a marker to visualize plant mitochondria in vivo. *The Plant Journal*, 11, 613–621.

Kruft, V., Eubel, H., Jänsch, L., Werhahn, W., & Braun, H. P. (2001). Proteomic approach to identify novel mitochondrial proteins in Arabidopsis. *Plant Physiology*, 127, 1694–1710.

Lee, S., Lee, D. W., Yoo, Y. J., Duncan, O., Oh, Y. J., Lee, Y. J., Lee, G., Whelan, J., & Hwang, I. (2012). Mitochondrial targeting of the Arabidopsis F1-ATPase γ-subunit via multiple compensatory and synergistic presequence motifs. *The Plant Cell*, 24, 5037–5057.

Levanony, H., Rubin, R., Altschuler, Y., & Galili, G. (1992). Evidence for a novel route of wheat storage proteins to vacuoles. *The Journal of Cell Biology*, 119, 1117–1128.

Lill, R., & Mühlenhoff, U. (2008). Maturation of iron-sulfur proteins in eukaryotes: mechanisms, connected processes, and diseases. *Annual Review of Biochemistry*, 77, 669–700.

Lister, R., Carrie, C., Duncan, O., Ho, L. H., Howell, K. A., Murcha, M. W., & Whelan, J. (2007). Functional definition of outer membrane proteins involved in preprotein import into mitochondria. *The Plant Cell*, 19, 3739–3759.

López-Torrejón, G., Jiménez-Vicente, E., Buesa, J. M., Hernandez, J. A., Verma, H. K., & Rubio, L. M. (2016). Expression of a functional oxygen-labile nitrogenase component in the mitochondrial matrix of aerobically grown yeast. *Nature Communications*, 7, 11426.

Luirink, J., & Sinning, I. (2004). SRP-mediated protein targeting: structure and function revisited. *Biochimica et Biophysica Acta*, 1694, 17–35.

Maćasev, D., Whelan, J., Newbigin, E., Silva-Filho, M. C., Mulhern, T. D., & Lithgow, T. (2004). Tom22', an 8-kDa trans-site receptor in plants and protozoans, is a conserved feature of the TOM complex that appeared early in the evolution of eukaryotes. *Molecular Biology and Evolution*, 21, 1557–1564.

Mukhopadhyay, A., Yang, C. S., & Weiner, H. (2006). Binding of mitochondrial leader sequences to Tom20 assessed using a bacterial two-hybrid system shows that hydrophobic interactions are essential and that some mutated leaders that do not bind Tom20 can still be imported. *Protein Science*, 15, 2739–2748.

Nelson, B. K., Cai, X., & Nebenführ, A. (2007). A multicolored set of in vivo organelle markers for co-localization studies in Arabidopsis and other plants. *The Plant Journal*, 51, 1126–1136.

Obita, T., Muto, T., Endo, T., & Kohda, D. (2003). Peptide library approach with a disulfide tether to refine the Tom20 recognition motif in mitochondrial presequences. *Journal of Molecular Biology*, 328, 495–504.

Pérez-González, A., Kniewel, R., Veldhuizen, M., Verma, H. K., Navarro-Rodríguez, M., Rubio, L. M., & Caro, E. (2017). Adaptation of the GoldenBraid modular cloning system and creation of a toolkit for the expression of heterologous proteins in yeast mitochondria. *BMC Biotechnology*, 17, 80.

Pfanner, N., & Geissler, A. (2001). Versatility of the mitochondrial protein import machinery. *Nature reviews. Molecular Cell Biology*, 2, 339–349.

Pierrel, F., Cobine, P. A., & Winge, D. R. (2007). Metal Ion availability in mitochondria. *Biometals*, 20, 675–682.

Rasmusson, A. G., Soole, K. L., & Elthon, T. E. (2004). Alternative NAD(P)H dehydrogenases of plant mitochondria. *Annual Review of Plant Biology*, 55, 23–39.

Rimmer, K. A., Foo, J. H., Ng, A., Petrie, E. J., Shilling, P. J., Perry, A. J., Mertens, H. D., Lithgow, T., Mulhern, T. D., & Gooley, P. R. (2011). Recognition of mitochondrial targeting sequences by the import receptors Tom20 and Tom22. *Journal of Molecular Biology*, 405, 804–818

Saberianfar, R., Sattarzadeh, A., Joensuu, J. J., Kohalmi, S. E., & Menassa, R. (2016). Protein Bodies in Leaves Exchange Contents through the Endoplasmic Reticulum. *Frontiers in Plant Science*, 7, 693.

Saito, Y., Kishida, K., Takata, K., Takahashi, H., Shimada, T., Tanaka, K., Morita, S., Satoh, S., & Masumura, T. (2009). A green fluorescent protein fused to rice prolamin forms protein body-like structures in transgenic rice. *Journal of Experimental Botany*, 60, 615–627.

Saitoh, T., Igura, M., Obita, T., Ose, T., Kojima, R., Maenaka, K., Endo, T., & Kohda, D. (2007). Tom20 recognizes mitochondrial presequences through dynamic equilibrium among multiple bound states. *The EMBO Journal*, 26, 4777–4787.

Savojardo, C., Martelli, P. L., Fariselli, P., & Casadio, R. (2015). TPpred3 detects and discriminates mitochondrial and chloroplastic targeting peptides in eukaryotic proteins. *Bioinformatics*, 31, 3269–3275.

Seibel, N. M., Eljouni, J., Nalaskowski, M. M., & Hampe, W. (2007). Nuclear localization of enhanced green fluorescent protein homomultimers. *Analytical Biochemistry*, 368, 95–99.

Shigemitsu, T., Masumura, T., Morita, S., & Satoh, S. (2013). Accumulation of rice prolamin-GFP fusion proteins induces ER-derived protein bodies in transgenic rice calli. *Plant Cell Reports*, 32, 389–399

Sluse, F. E., Jarmuszkiewicz, W., Navet, R., Douette, P., Mathy, G., & Sluse-Goffart, C. M. (2006). Mitochondrial UCPs: new insights into regulation and impact. *Biochimica et Biophysica Acta*, 1757, 480–485.

Snapp E. (2005). Design and use of fluorescent fusion proteins in cell biology. *Current Protocols in Cell Biology*, 27, 21.4.1–21.4.13.

Sudhakar, D., Duc, L.T., Bong, B.B., Tinjuangjun, P., Maqbool, S.B., Valdez, M., Jefferson, R., & Christou, P. (1998). An efficient rice transformation system utilizing mature seed-derived explants and a portable, inexpensive particle bombardment device. *Transgenic Research*, 7, 289–294.

Takaiwa, F., Hirose, S., Takagi, H., Yang, L., & Wakasa, Y. (2009). Deposition of a recombinant peptide in ER-derived protein bodies by retention with cysteine-rich prolamins in transgenic rice seed. *Planta*, 229, 1147–1158.

Taylor, R. D., & Pfanner, N. (2004). The protein import and assembly machinery of the mitochondrial outer membrane. *Biochimica et Biophysica Acta*, 1658, 37–43.

Van Steeg, H., Oudshoorn, P., Van Hell, B., Polman, J. E., & Grivell, L. A. (1986). Targeting efficiency of a mitochondrial pre-sequence is dependent on the passenger protein. *The EMBO Journal*, 5, 3643–3650.

Weber, E., Engler, C., Gruetzner, R., Werner, S., & Marillonnet, S. (2011). A modular cloning system for standardized assembly of multigene constructs. *PloS ONE*, 6, e16765.

Wei, X., Henke, V. G., Strübing, C., Brown, E. B., & Clapham, D. E. (2003). Real-time imaging of nuclear permeation by EGFP in single intact cells. *Biophysical Journal*, 84, 1317–1327.

Weis, B. L., Schleiff, E., & Zerges, W. (2013). Protein targeting to subcellular organelles via MRNA localization. *Biochimica et Biophysica Acta*, 1833, 260–273.

Wiedemann, N., & Pfanner, N. (2017). Mitochondrial Machineries for Protein Import and Assembly. *Annual review of biochemistry*, 86, 685–714.

Yamamoto, H., Itoh, N., Kawano, S., Yatsukawa, Y., Momose, T., Makio, T., Matsunaga, M., Yokota, M., Esaki, M., Shodai, T., Kohda, D., Hobbs, A. E., Jensen, R. E., & Endo, T. (2011). Dual role of the receptor Tom20 in specificity and efficiency of protein import into mitochondria. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 91–96.

Zhang, Y., Baaden, M., Yan, J., Shao, J., Qiu, S., Wu, Y., & Ding, Y. (2010). The molecular recognition mechanism for superoxide dismutase presequence binding to the mitochondrial protein import receptor Tom20 from Oryza sativa involves an LRTLA motif. *The Journal of Physical Chemistry*, 114, 13839–13846.

Chapter 3

Functional expression of O_2-sensitive nitrogenase Fe protein (NifH) in rice mitochondria-a critical step towards the engineering of nitrogen fixation in cereals

3.0. Abstract

Biological nitrogen fixation is a microbial process that converts nitrogen gas to ammonia making it accessible for synthesis of nitrogenous compounds. This process is carried out by free-living diazotrophs as well as symbiotic bacteria in the root nodules of legumes and some non-leguminous plants. The integration of the nitrogen fixation machinery into cereal crops has been a long-term goal of agricultural development, but this has been frustrated by the inability to express the nitrogenase protein complex (NifHDK) in an active form, partly due to the extreme oxygen-sensitivity of the Fe protein encoded by *nifH*. I overcame this challenge by expressing NifH from the extremophile *Hydrogenobacter thermophilus* (*Aquificae*) in rice mitochondria together with a putative peptidyl prolyl cis-trans isomerase (NifM). Substantial amounts of *H. thermophilus* NifH were purified from rice callus. The isolated NifH carried out the fundamental roles of NifH protein required to engineer nitrogen fixation, including electron transfer to NifDK and FeMo-co biosynthesis. The activity of NifH confirmed its ability to incorporate endogenous rice mitochondrial [Fe-S] clusters. Furthermore, the reconstitution of NifH by the *in vitro* transfer of [4Fe-4S] from NifU increased NifH activity nine-fold compared to the as-isolated NifH protein. This is the first demonstration of the stable expression and activity of a nitrogenase component in a major staple crop.

3.1. Introduction

Nitrogen (N) is one of the essential elements and primary nutrients that limits plant growth, development and reproduction (Ohyama, 2010). Plants utilize nitrogen (N) in two different forms, nitrate (NO_3-) and ammonium (NH_4+). N is a fundamental nutrient which shows its significant effects directly as increase in the yield and enhances plant quality by playing a vital role in many biochemical, developmental and physiological functions (Lawlor *et al*., 2001). In addition, N is one of the major components of chlorophyll, energy-transfer compounds such as ATP, nucleic acids and proteins. Therefore, nitrogen availability in plants limits photosynthesis, cell growth, metabolism, and protein synthesis (Ohyama, 2010). Thus, crop production totally depends on an adequate N supply especially in cereals (Chapin *et al*., 2011). Even though crop production requires high amounts of nitrogen, plants do not assimilate more than half of the applied N and it is a

major source of pollution in agricultural systems (Cui *et al*., 2013; Kronzucker and Coskun, 2015). Therefore, it is important to explore strategies to reduce N fertilizer-dependence in crop production especially in cereals. Direct transfer of nif genes constitutues one of the most important biotechnological efforts to achieve biological nitrogen-fixation in plants (Curatti and Rubio, 2014; Burén *et al.,* 2018).

Biological nitrogen fixation (BNF) is the reduction of inert N_2 gas to ammonia (NH_3) through a reaction catalyzed by nitrogenase (Dos Santos *et al*., 2012). Nitrogenases are complex and extremely oxygen-sensitive enzymes that require multiple sub-units encoded by many different genes (Rubio and Ludden, 2008; Burén *et al*., 2017). BNF is carried out by a group of prokaryotes belonging to the bacteria or archaea domains. Eukaryotes are not capable of converting N_2 into a biologically useful form (Burén *et al.*, 2018).

There are three structurally and functionally similar but genetically distinct nitrogenases distinguished, in part, by the metal composition of their corresponding active-site cofactors namely; molybdenum (Mo) dependent, vanadium (V) dependent and iron (Fe) only nitrogenase (Burén *et al*., 2020) (**Figure 3.1**). The most abundant and well characterized Mo nitrogenase is an extremely oxygen sensitive metalloenzyme with two interacting subunits, dinitrogenase reductase and dinitrogenase (Rubio *et al*., 2008). The Mo dinitrogenase reductase is also known as Fe protein, because Fe is the only metal present in its cluster. It is a homodimer of the *nifH* product (López-Torrejón *et al.,* 2016). Fe protein carries a single [4Fe-4S] cluster linking the two subunits of nitrogenase and serves to mature the metalloclusters of the MoFe protein (which is the product of heterotetramer *nifD* and *nifK)* (Burén *et al.,* 2017). The most important function of Fe protein is to provide electrons to the MoFe protein during N_2 reduction. Therefore, correct processing of Fe protein (NifH), its maturation and activity are the first essential requirements for engineering N_2 fixation in plants. (Burén *et al.,* 2018).

The total number of genes required for having a functional nitrogenase enzyme differs among different organisms, but it is usually estimated to ca. 10–20 genes (Temme, Zhao and Voigt, 2012; Echavarri-Erasun and Rubio, 2015). The transfer of the entire prokaryotic *nif* gene cluster to eukaryotes is currently beyond the state of the art due to serious gaps in our understanding of the biochemistry and enzymology of how such

prokaryotic genes function in higher eucaryotes, including plants and the uncertainty of interactions among nif genes. Therefore, determining and transferring the minimum gene complement that is necessary to obtain active Fe protein (which is also the most abundant component of nitrogenase) is of paramount importance.

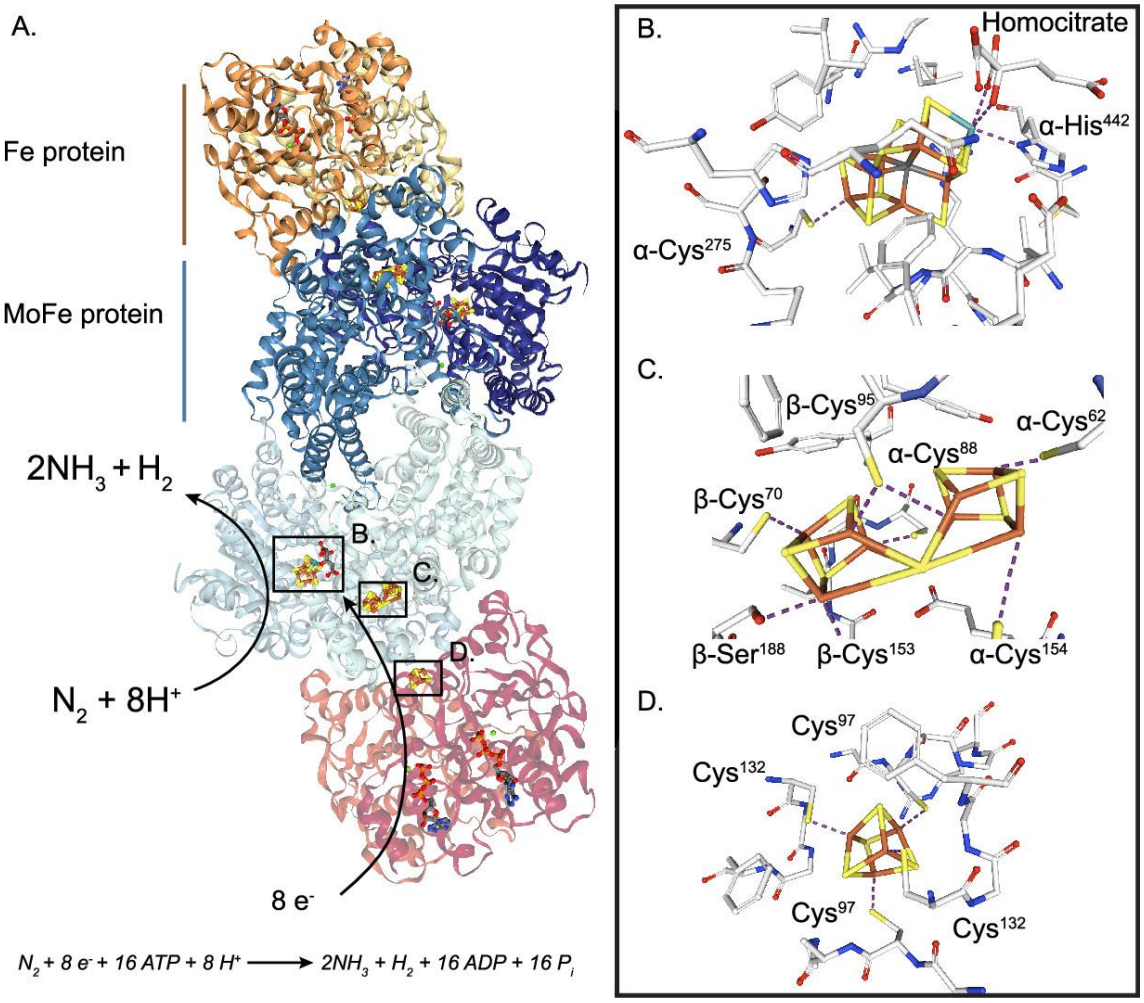


Figure 3.1. Structure of Mo nitrogenase. (A) *Azotobacter vinelandii* MoFe protein and Fe protein complex. Top half (panel A) shows polypeptide secondary structure. Partial transparency has been applied to the polypeptide chains of the bottom half (panel A) to visualize metal clusters and nucleotides. Ligands and surrounding environment for (B) FeMo-co, (C) P-cluster and (D) [4Fe-4S] cluster are shown. Mo nitrogenase reaction is shown at the bottom (Burén *et al*., 2020).

Three accessory genes (NifM, NifS and NifU) are required to render active Fe protein in *Azotobacter vinelandii* (López-Torrejón *et al.*, 2016). NifM, is a putative peptidyl-prolyl cis-trans isomerase that assists in the proper folding and is essential for the maturation of the Fe protein (Gavini *et al*., 2006). NifS catalyzes the desulfurization of L-cysteine responsible for the mobilization of S for [Fe-S] cluster synthesis on NifU where [Fe-S]

clusters are synthesized and later donated to the Fe protein for activity of nitrogenase (Dos Santos *et al*., 2004; Yuvaniyama *et al*., 2000). Nitrogenase is a complex and extremely oxygen-sensitive enzyme that requires specific cellular compartments for its assembly (Curatti and Rubio, 2014). Mitochondria and chloroplasts have been considered as potential locations to engineer nitrogen fixation in plants. Mitochondria, because of the low O_2 environment and presence of bacterial type [Fe-S] cluster biosynthesis machinery; chloroplasts because of the availability of the essential photosynthesis products NADPH and ATP to satisfy the high energy costs of BNF (Merrick and Dixon, 1984; López-Torrejón *et al*., 2016; Burén *et al*., 2018; Eseverri *et al*., 2020). Furthermore, the presence of glutamine synthetase/glutamate synthase pathway in chloroplasts is thought to facilitate the assimilation of ammonia into amino acids, produced during nitrogen fixation (Eseverri *et al*., 2020).

One of the most successful examples of engineering active nitrogenase components in eukaryotes to date is the production of active NifU and NifH within the mitochondria of aerobically growing- yeast, overcoming the O_2 sensitivity of Nif proteins (López-Torrejon *et al*., 2016). This study confirmed that to have active Fe protein in yeast, co-expression of only NifH and NifM in mitochondrial matrix is required.

The first (and only) active Fe protein in higher plants prior to the work described in this thesis was obtained by stable expression of the *Klebsiella pneumoniae* NifH with its maturase NifM in the tobacco plastid genome (Ivleva *et al.,* 2016). Functional activity of the plant-produced NifH protein was demonstrated *in vitro* by combining bacteria-produced MoFe protein. Fe protein was active only after when plants were incubated at low oxygen concentration. In addition, *Klebsiella* Fe protein activity was at best marginal, bearly over background levels (Ivleva *et al.,* 2016). Transient expression of *A. vinelandii* NifH, M, U and S proteins within the stroma of mesophyll cells confirmed Fe protein activity. However, Fe protein was only active active when it was co expressed with NifS and NifU (Eseverri *et al.*, 2020).

Recently, 32 distinct *nifH* genes were screened for expression levels and solubility in mitochondria of *N. benthamiana*. The *H. thermophilus* NifH was shown to be superior to the *A. vinelandii* homologue in terms of expression levels, solubility, and functionality. Transient expression of *H. thermophilus* NifH in *N. benthamiana* leaves showed lower

nitrogenase activity as-isolated. However, reconstituted NifH by *in vitro* transfer of [Fe4S4] clusters from NifU increased the activities up to 10-fold revealing that presence and assembly of plant mitochondrial [Fe-S] cluster is the major bottleneck for engineering nitrogenase (Jiang *et al.*, 2021).

Therefore, the first crucial steps to generate active nitrogenase in higher eukaryotes is to transfer bacterial nitrogenase iron protein encoding *nifH* with its maturase *nifM*, (with or without *nifS* and *nifU)* and to obtain correctly-folded and matured NifH which can provide electrons to activate (in vitro) the MoFe protein NifDK.

I transformed rice with a minimum *nif* gene combination (*nifH* and *nifM*) that is necessary for iron protein expression, its maturation and activity. I targeted nuclear-encoded Nif proteins (NifH and NifM) to the rice mitochondrial matrix to minimize the oxygen damage and to obtain stably expressed, correctly folded, clustered and enzymatically active NifH capable of donating electrons for N_2 reduction.

3.2. Aims

- Transfer the key nitrogenase genes and confirm their expression at mRNA and protein levels in rice dedifferentiated cells and plants.
- Identify the specific (trans)gene combination required to recover rice callus and plant lines accumulating high levels of NifH protein.
- Isolate Fe protein from rice callus and demonstrate enzymatic activity
- Identify constraints in obtaining enzymatically active NifH protein.

3.3. Material and Methods

3.3.1. *Genetic constructs*

S. cerevisiae codon optimized *H. thermophilus NifH* and *A. vinelandii NifM were* used for rice transformations. *NifH* and *NifM* sequences were optimized using the GeneOptimizer tool (ThermoFisher, Cambridge, MA, U.S.) and synthesized by ThermoFisher via Engineering Nitrogen Symbiosis for Africa (ENSA) project. To generate pUC57 (GenScript, Piscataway, New Jersey, U.S.) vector containing promoter maize ubi1+1sti:cox4-twinstrep-*NifH*^{Ht}:tNos, firstly the empty pUC57 vector was

digested with BamHI and PstI. The maize ubi1+1sti promoter was synthesized and cloned into the empty pUC57 vector with BamHI and PstI restriction sites. The cox4-twinstrep-*NifH*^{Ht}:tNOS was synthesized and cloned into pUC57 with PstI and SphI restriction sites flanking the maize ubi1+1sti promoter. To generate pUC57 containing p*OsActin*:su9-*NifM*^{Av}:tNos, firstly the empty pUC57 vector was digested with ACC65I and XBaI to clone the *OsActin* promoter. *OsActin* promoter was synthesized and cloned into the empty pUC57 vector with ACC65I and XbaI restriction sites. The su9-*NifM*^{Av}:tNos plasmid was synthesized and cloned into pUC57 with XbaI and SaII restriction sites flanking the *OsActin* promoter. All DNA digestions were performed using enzymes from Promega. Ligated products (T4 ligase, Promega) were introduced into *E. coli DH5α* chemically competent cells and selected on LB (Lysogenic broth) supplemented with ampicillin. Plasmid integrity was confirmed by Sanger sequencing (Stabvida, Caparica/Portugal). Plant expression vectors and sequences of all genetic constructs are listed in **Table 3.1.** Primers used for vector construction are listed in **Table 3.2.**

Table 3.1. Plant expression vectors used (A) and sequences of genetic constructs (B).

A) Plant expression vectors used				
Plasmid backbone	**Expressed proteins**	**Promoter / Terminator**	**Size (kDa) FL.**	**P.**
PUC57	Cox4-TS-NifH^{Ht}	p*ZmUbi1*+1st intron / tNos	37.0	33.5
PUC57	SU9-NifM^{Av}	p*OsActin* / tNos	40.4	33.0

FL, full-length; **P,** processed by removal of mitochondria targeting peptide.
TS, tween-strep
COX4, *Saccharomyces cerevisiae c*ytochrome c oxidase subunit 4 mitochondrial targeting peptide.
SU9, *Neurospora crassa* subunit 9 mitochondrial targeting peptide.

B) Sequences of genetic constructs
COX4: MLSLRQSIRFFKPATRTLCSSRYLLQQKP
SU9: MASTRVLASRLASQMAASAKVARPAVRVAQVSKRTIQTGSPLQTLKRTQMTSIVNATTRQAFQKRAYSS
TS: WSHPQFEKGGGSGGGSGGSAWSHPQFEK
NifH^{Ht}: MRQIAIYGKGGIGKSTTTQNTVAALAEAGRKCFIVGCDPKADSTRLILHVKAQSTVMHLAAERGAVEDLDLDEVMLV GFGGIKCVESGGPEPGVGCAGRGVITAINFLEENGAFDDDLDYVFYDVLGDVVCGGFAMPIREGKAQEIYIVTSGEMM AMYAANNISKGILKYAHSGGVRLGGLICNSRNVDNERELIEALAEKLGTQMIHFLPRNNIVQEAELRRMTVIEYAPDH PMADEYRTLAKKIEENRKLSIPTPLTMDELEQLLVEYGIMKPEEVA*
NifM^{Av}: MASERLADGDSRYYLLKVAHEQFGCAPGELSEEQLQQADRIIGRQRHIEDAVLRSPDAIGVVIPPSQLEEAWAHIASRY ESPEALQQALDAQALDAAGMRAMLARELRVEAVLDCVCAGLPEISDTDVSLYYFNHAEQFKVPAQHKARHILVTINE DFPENTREAARTRIETILKRLRGKPERFAEQAMKHSECPTAMQGGLLGEVVPGTLYPELDACLFQMARGELSPVLESPI GFHVLYCESVSPARQLTLEEILPRLRDRLQLRQRKAYQRKWLESLLQQNATLENLAHG*

3.3.2. *Gene expression analysis by quantitative real-time PCR*

Total RNA was isolated from rice callus and corresponding regenerated plant leaves using the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany). First-strand cDNA was synthesized from 1 µg total RNA using Ominiscript Reverse Transcriptase (Qiagen) and quantitative real-time RT-PCR (qRT-PCR) was carried out as previously described (Baysal *et al.,* 2020) using the gene-specific primers listed in **Table 3.2.** The identity of the PCR products was confirmed by sequencing. Expression levels were normalized against *OsActin* mRNA. Three technical replicates were used for each gene.

Table 3.2. Primers used for vector construction and quantitative Real-Time PCR analysis.

Genes/promoters	**Primers used for quantitative Real-Time PCR analysis**
HtNifH – (F)	5'-ACGCTCATTCTGGTGGTGTT-3'
HtNifH – (R)	5'-CCATTGGATGATCTGGGGCA-3'
AvNifM – (F)	5'-TGCAAGGTGGGTTTGTTAGGTG-3'
AvNifM – (R)	5'-GTCTTGCTGGACTGACGGAT-3'
OsActin – (F)	5'-TCATGTCCCTCACAATTTCC-3'
OsActin – (R)	5'-GACTCTGGTGATGGTGTCAGC-3'
	Primers used for vector construction
OsActin – (F)	5'-TAAGCAGGTACCTAGCTAGCATACTCGAGGT -3'
OsActin – (R)	5'-TGCTTATCTAGACTTCTACCTACAAAAAAGC-3'
$NifM^{Av}$ – (F)	5'-TAAGCATCTAGAATGGCCTCCACTCGTGTCC-3'
$NifM^{Av}$ – (R)	5'-TGCTTAGTCGACGATCTAGTAACATAGATGA-3'
pZmUbi1+1st*i* – (F)	5'-TAAGCAGGATCCGGAGTGCAGTGCAGCGTGA-3'
pZmUbi1+1st*i* – (R)	5'-TGCTTACTGCAGAAGTAACACCAAACAACAG-3'
$NifH^{Ht}$ – (F)	5'-TAAGCACTGCAGAATGCTTTCACTTAGACAA-3'
$NifH^{Ht}$ – (R)	5'-TGCTTAGCATGCACATACAAATGGACGAACG-3'

3.3.3. *Transformation of rice callus and regeneration of transgenic plants*

Six-day-old mature rice zygotic embryos (*Oryza sativa* cv. Nipponbare) were isolated from surface sterilized seeds and transferred to Murashige and Skoog (MS) osmoticum medium [4.4g/L MS powder (Duchefa Biochemie BV, Haarlem, Netherlands) supplemented with 0.3g/L casein hydrolysate, 0.5g/L proline, 72.8 g/L mannitol and

30g/L sucrose, 2.5mg/L 2,4-dichlorophenoxyacetic acid], 4h before bombardment and transformed with 10mg gold particles coated with the single *nif* gene carrying constructs and a construct carrying the selectable marker hygromycin phosphotransferase (*hpt*) in a 3:1 molar ratio (Baysal *et al.,* 2016). The embryos were returned to osmoticum MS medium for 16h before selection on MS medium (4.4g/L MS powder, 0.3g/L casein, 0.5g/L proline and 30g/L sucrose) supplemented with 50mg/L hygromycin and 2.5mg/L 2,4-dichlorophenoxyacetic acid in the dark for 2–3 weeks. One half of each callus clone was maintained in an undifferentiated state, and the other half was transferred to regeneration medium as previously described (Farré *et al*., 2012). The resulting transgenic plantlets were transferred to soil and grown in the greenhouse or growth chamber at 28/25ºC day/ night temperature with a 12h photoperiod and 80% relative humidity.

3.3.4. *Protein extraction and immunoblot analysis*

Total rice protein extracts were prepared by grinding 0.1–0.2g callus or leaf tissue in liquid nitrogen and thawing the powder in 0.2–0.4mL of extraction buffer: 20mM Tris–HCl pH 7.5, 5mM ethylenediaminetetraacetic acid (EDTA), 0.1% Tween-20, 0.1% sodium dodecyl sulfate (SDS), 2mM phenylmethanesulfonylfluoride (PMSF). The mixture was vortexed for 1h at 4°C. Cell debris was removed by centrifugation at 15,000 × g for 20 min at 4°C, and the supernatant was collected and stored at − 80°C. The protein concentration in the supernatants was determined using the Bradford method (AppliChem, Darmstadt, Germany). I fractionated 80µg of total rice protein by denaturing SDS-PAGE in polyacrylamide gels containing 10% SDS at 200 V for 60 min, and then electro-transferred the protein to an Immobilon FL polyvinylidene difluoride (PVDF) membrane (Merck, Darmstadt, Germany) using a semidry transfer apparatus (Bio-Rad, Hercules, CA, U.S.) at 20V for 45 min. The membrane was immersed in 5% non-fat milk in Tris-buffered saline with Tween-20 (TBST) solution (0.2M Tris–HCl pH 7.6, 1.37M NaCl, 0.1% Tween-20) for 1h at room temperature. Membranes were incubated with the appropriate antibodies and their dilutions for immunoblotting were as follows: Strep-tag II (“Strep-MAB”, 2-1507-001, IBA Lifesciences, 1:2,000 in 5% BSA), Strep-Tactin conjutaged to HRP (“StrepHRP”, 2-1502-001, IBA Lifesciences, 1:50,000 in TBS-T) monoclonal antibodies detecting strep tagged *Os*NifH^{Ht} and polyclonal antibody detecting NifM^{Av} (used at 1:2,000 in 5% BSA) (generated in CBGP).

3.3.5. *Rice Fe protein purification*

Purifications *Os*NifH^{Ht} was performed at O_2-levels below 1ppm inside anaerobic chambers (Coy systema or MBraun) at CPGB, in collaboration with Dr S. Burén. Typically, 100-200g of callus material were harvested and immediately frozen in liquid N_2. The material was then transferred into an anaerobic chamber while still frozen and disrupted in 2x (w/v) of lysis buffer (100mM Tris-HCl pH 8.6, 200mM NaCl, 10% glycerol, 3mM DTH, β-Mercaptoetanol 5mM, 1mM PMSF, 1µg/ml leupeptin, 5µg/ml DNaseI, biotin blocking 1µg/ml) using a blender (Oster Classic 4655) operating at maximum power and maintained at 4°C using a circulating water bath. The total extract (TE) was transferred to centrifuge tubes equipped with sealing closures (Beckman Coulter, Brea, California, U. S.) and centrifuged at 50,000g for 1h at 4°C (Avanti J-26 XP, Brea, California, U.S.). The supernatant was filtered using filtering cups with a pore size of 0.2 µm, rendering cell-free extract (CFE) of soluble proteins which was loaded at 2.5ml/min onto a 5ml Strep-Tactin XP column (IBA LifeSciences, Göttingen, Germany) attached to an ÄKTA FPLC (GE Heathcare, Chicago, Illinois, United States).

The column was washed using 75ml washing buffer (100mM Tris-HCl pH 8.0, 200mM NaCl, 10% glycerol, 2mM DTH). Strep-Tactin XP column-bound proteins were eluted with 15ml washing buffer supplemented with 50mM biotin (IBA LifeSciences). The elution fraction was concentrated, and biotin removed, by passing the protein through PD-10 desalting columns (GE Healthcare). Desalted eluate was further concentrated using centrifugal filters (Amicon, Millipore, Merck, Darmstadt, Germany) with 30 kDa cutoff. Finally, the concentrated protein was snap-frozen in cryovials (Nalgene, Merck,) and stored in liquid N_2.

All essential steps of the strep-tag affinity chromatography (STAC) purification procedure of mitochondria targeted *Os*NifH^{Ht} is shown in **Figure 3.2**.

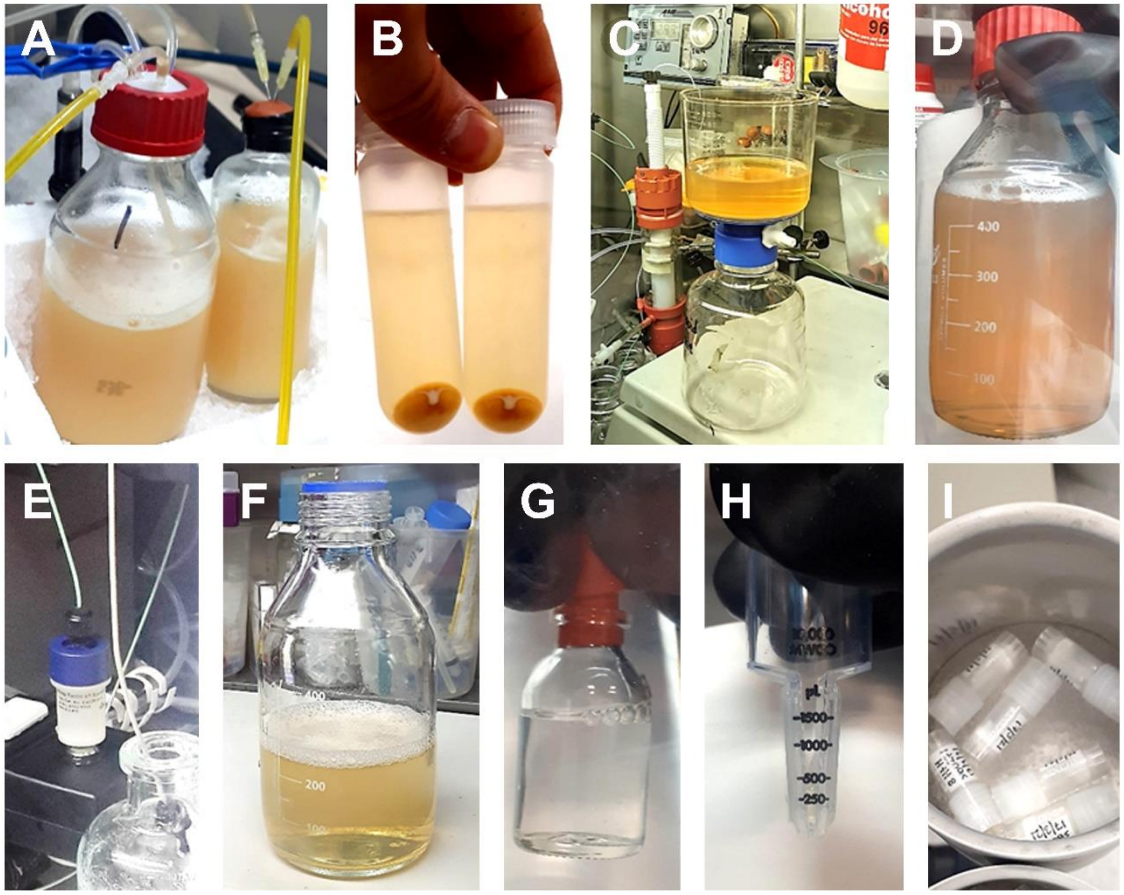


Figure 3.2. Strep-tag affinity chromatography (STAC) purification procedure of mitochondria targeted *Os*NifH^{Ht}. A) Rice callus total extract. B) Rice callus total extract after centrifugation to separate soluble (supernatant) and insoluble (pellet) fractions. C) Supernatant filtration. D) Resulting extract after supernatant filtration (denoted cell free extract), E) Streptactin XT column loaded with *Os*NifH^{Ht} after passing of CFE. F) Flow-through fraction. G) Biotin-eluted *Os*NifH^{Ht} (approx. 20ml). H) Concentrated elution of *Os*NifH^{Ht} I) Pure *Os*NifH^{Ht} final preparation (transfer to cryogenic vial for liquid nitrogen storage).

3.3.6. *Rice Fe protein activity determination*

*Os*NifH^{Ht} activity was determined as described by Jiang *et al.* (2021). Reactions were prepared inside anaerobic chambers. Purified *Os*NifH^{Ht} protein were analysed by acetylene reduction assay (ARA) after addition of NifDK^{Av} and ATP regenerating mixture (1.23mM ATP, 18mM phosphocreatine, 2.2mM MgCl2, 3mM DTH and 46µg/ml of creatine phosphokinase, 22mM Tris-HCl pH 7.5) in a final volume of 400µl inside 9ml serum vials under Argon (Ar) atmosphere containing 500µl of acetylene (1atm). The ratio of *Os*NifH^{Ht} to NifDK^{Av} in the assays was 20:1 unless otherwise indicated. The ARA were performed at 30°C in a shaking water bath for 15min. Reactions were stopped by adding 100µl of 8M NaOH. Positive control reactions for acetylene reduction were carried out with NifH^{Av}. Ethylene formed was measured in 50µl gas phase samples using a Porapak N 80/100 column in a gas chromatograph (Shimadzu, Duisburg, F.R. Germany).

3.3.7. *In vitro [Fe-S] cluster reconstitution and* reconstituted *Rice NifH activity*

In vitro [Fe-S] cluster reconstitution of *Os*NifH^{Ht} was performed in anaerobic chambers. *E. coli* NifU^{Av} (20µM) was added to 22mM Tris-HCl (pH 7.5) buffer supplemented with 8mM 1,4-dithiothreitol (DTT) in a final volume of 100µl and incubated at 37°C for 30min. Reactions were then supplemented with 1mM L-cysteine, 1mM DTT, 800µM $(NH_4)_2Fe(SO_4)_2$, and 225nM NifS^{Av} purified from *E. coli* (*Ec*NifS^{Av}), and incubated at 37°C overnight. Finally, the proteins were diluted 1000-fold in 22mM Tris-HCl (pH 7.5) buffer, and concentrated using centrifugal filters (Amicon, Millipore) with 30kDa cut off to remove excess reagents. Isolated *Os*NifH^{Ht} protein was mixed with [Fe-S] cluster reconstituted on *Ec*NifU^{Av} (NifU-mediated reconstitution) and then immediately used for ARA.

3.3.8. *In vitro FeMo-co synthesis and apo-NifDK^{Av} reconstitution*

NifB-co-dependent FeMo-co synthesis assays were performed inside anaerobic chambers as described by Curatti *et al*., (2007) with slight modifications. One hundred µl reactions contained 3.0µM *Sc*NifH^{Ht}, GST-NifX-NifB-co (20.4 µM Fe), 1.5µM apo-NifEN^{Av}, 0.3µM apo-NifDK^{Av}, 17.5µM Na_2MoO_4, 175µM R-homocitrate, 1mg/ml BSA, and ATP regenerating mixture (1.23mM ATP, 18mM phosphocreatine disodium salt, 2.2mM $MgCl_2$, 3mM DTH, 46µg/ml creatine phosphokinase, final concentrations in 22mM Tris-HCl (pH 7.5) buffer at 30°C for 60min. NifB-dependent FeMo-co synthesis assays were performed as described above for NifB-co-dependent assay replacing GST-NifX-NifB-co by 1µM *Sc*NifB^{Mt} monomer, 125µM $FeSO_4$, 125µM L-cys, and 125µM SAM. Following *in vitro* synthesis of FeMo-co, 17.5µM $(NH_4)_2MoS_4$ was added to prevent further FeMo-co incorporation into apo-NifDK^{Av}, and incubated for 10min at 25°C. Activation of apo-NifDK^{Av} was analyzed by addition of 500µl ATP regenerating mixture and NifH^{Av} (2.0µM final concentration) in 9ml vials containing Ar and 500µl acetylene. The ARA were performed at 30°C for 20min. Ethylene formed was measured in 50µl gas phase samples using a Porapak N 80/100 column in a gas chromatograph (Shimadzu).

3.3.9. *Statistical analysis*

The distinct samples used for *in vitro* activity measurements and sample sizes are indicated by n, where each sample was measured at least two times. For each assay, the standard deviation (SD) of the mean was calculated based on at least three technical replicates.

3.4. Results

3.4.1. *Vector construction and targeting Nif proteins to rice mitochondria*

The gene sequences encoding the NifH^{Ht}, NifM^{Av} were cloned into separate vectors and used for stable rice transformation with the hygromycin phosphotransferase selectable marker gene (*hpt*). The *nifH*^{Ht} and *nifM*^{Av} sequences were previously codon optimized for *S. cerevisiae* and successfully expressed in *S. cerevisiae* and *N. benthamiana* (Jiang *et al*., 2021). The same genes were used in rice transformations because of codon-usage similarity. The *nifH*^{Ht} gene was driven by the strong constitutive *ZmUbi1*+1sti promoter.

I previously analyzed a number of targeting peptides from diverse phylogenetic origins (higher and lower eukaryotes) and I found that COX4 successfully directed GFP into rice mitochondria (Baysal *et al.*, 2019). Therefore, amino-terminal COX4 was added to NifH^{Ht}. COX4 is the 29 amino acid transit peptide of the *S. cerevisiae* mitochondrial protein cytochrome c oxidase subunit IV, whereas Twinstrep-tag (TS-tag) denotes the 28 amino acid. TS-tag was used to enable immunoblot detection of NifH^{Ht} and its purification. Importantly, the TS-tag was shown not to have any significant effect on NifH^{Ht} functionality in yeast (stable transformation) and *N. benthamiana* (transient expression) (Burén *et al.*, 2018, Eseverri *et al.*, 2020, Jiang *et al.*, 2021).

An auxiliary vector was constructed to co-express *NifM*^{Av} and target its protein product to rice mitochondria via N-terminal SU9 extension. Similar to COX4, the mitochondrial presequence of subunit 9 of *Neurospora crassa* (SU9) has been confirmed to deliver GFP to rice (Baysal *et al.,* 2019) and Nif proteins to *N. benthamiana* mitochondria (Jiang *et al.*, 2021).

3.4.2. *Co-expression of plant-produced OsNifH^{Ht} and OsNifM^{Av}*

Expressions of *Os*NifH^{Ht} and *Os*NifM^{Av} at mRNA and protein levels were confirmed in rice callus and the corresponding regenerated plants (**Figure 3.3**). NifH is the most abundant Nif protein required for nitrogen fixation (Poza-Carrión *et al.,* 2014). *NifH^{Ht}* was driven by the strong constitutive *ZmUbi1*+1sti promoter to achieve high expression levels and consequently accumulate substantial amounths for its purification and activity assays. *NifM^{Av}* which is responsible of the maturation of NifH was driven by rice actin promoter to have relatively lower accumulation levels as required for the correct stoichiometry of the two enzymes.

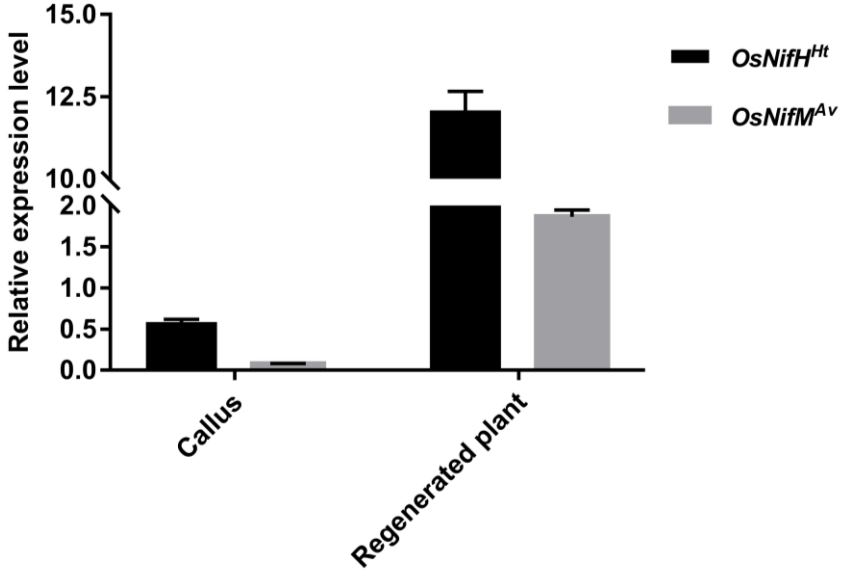


Figure 3.3. Relative *Os*NifH^{Ht} and *Os*NifM^{Av} mRNA expression in rice callus and corresponding regenerated plants (data normalized to *OsActin*).

To demonstrate the accumulation of the *Os*NifH^{Ht} and *Os*NifM^{Av} in the mitochondria of rice callus and plants, immunoblot analysis was carried out using antibodies raised against NifH^{Ht} and NifM^{Av} proteins (**Figures 3.4** and **3.5)**. Both *Os*NifH^{Ht} and *Os*NifM^{Av} produced immunoreactive bands of the expected size (ca. 33kDa). Side by side comparison of yeast mitochondria-expressed *Sc*NifH^{Ht} vs. the rice counterpart confirmed the correct processing of *Os*NifH^{Ht} protein in the rice mitochondrial matrix as evidenced by their identical molecular weight (ca. 33kDa) in Coomassie gels (**Fig. 3.4 D**). I thus demonstrated that *nifH^{Ht}* and *nifM^{Av}* were integrated into the rice genome and successfully expressed in rice mitochondria.

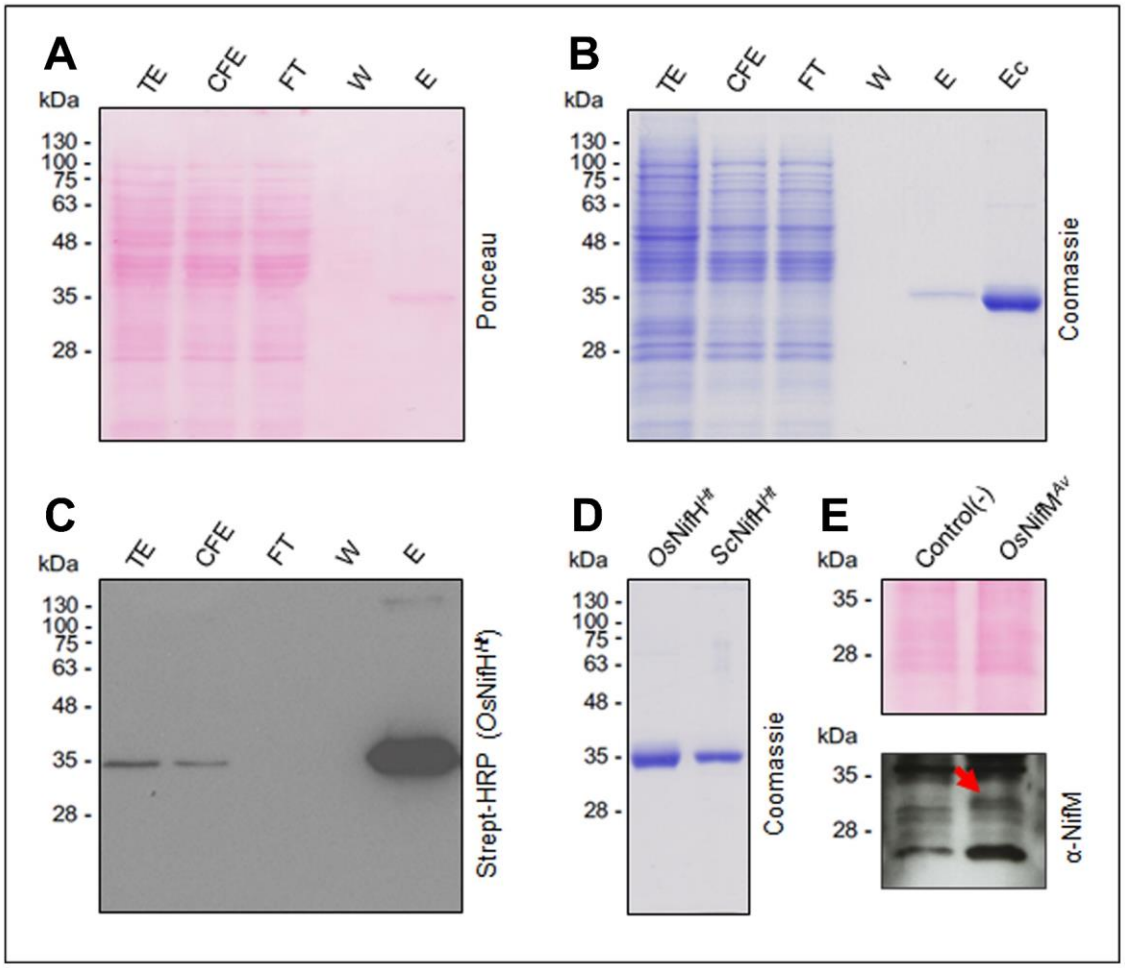


Figure 3.4. Strep-tag affinity chromatography (STAC) purification of mitochondria-targeted *Os*NifH^{Ht} protein from rice. (A) Ponceau staining showing the purification fractions. TE = total extract, CFE = soluble cell free extract, FT = flow through fraction, W = wash fraction, E = elution fraction. Ec = concentrated elution fraction. The purified *Os*NifH^{Ht} protein was soluble. B) Fractions were analyzed by SDS-PAGE by Coomassie staining. C) Immunoblot analysis of fractions against strep-tactin antibody. D) Side by side comparison of pure *Sc*NifH^{Ht} vs. *Os*NifH^{Ht} (E). Immunoblot analysis of *Os*NifM^{Av}: red arrow indicates *Os*NifM^{Av} protein based on it is correctly processed molecular weight 33kDa.

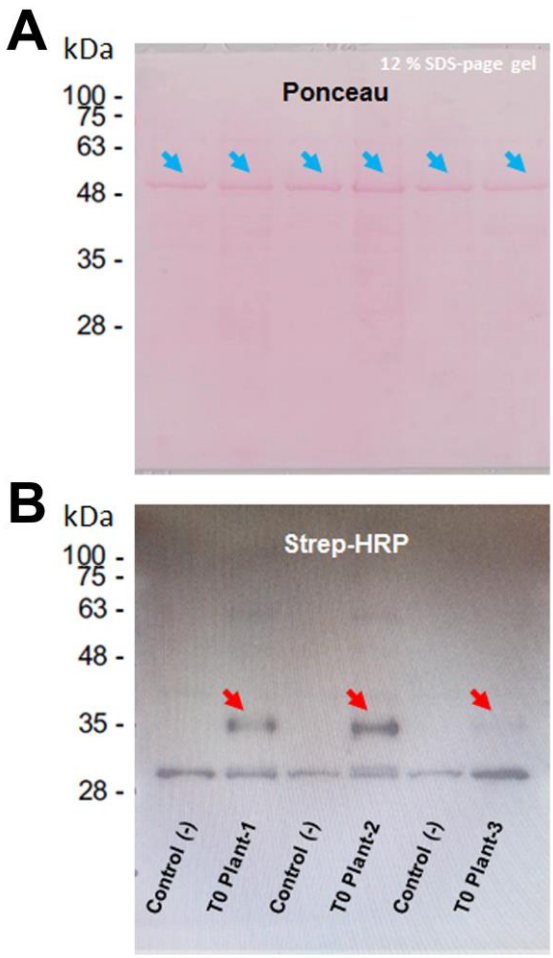


Figure 3.5. Expression of *Os*NifH^{Ht} protein in three independent regenerated rice plants. A) Rubisco protein in ponceau staining confirms equal sample loading (blue arrows). B) Total protein extracts of rice leaves were analysed by western blot analysis to detect strep tag conjugated *Os*NifH^{Ht} that was immunoreactive to antibody raised against the strep tag. Red arrows indicate correctly processed strep-tag conjugated *Os*NifH^{Ht}.

3.4.3. *Purification of mitochondria-expressed rice Fe protein*

N-terminal strep-tagged *Os*NifH^{Ht} was purified from rice callus using strep-tag affinity chromatography (STAC). Purified *Os*NifH^{Ht} was totally soluble and there was no protein lost in the pellet fraction when preparing the soluble cell-free extract (CFE) explaining its high purification yield (ca. 2mg per kg of rice callus). The purified *Os*NifH^{Ht} protein in the elution fraction is shown as a single band corresponding the correctly processed NifH^{Ht} protein in rice callus (**Figure 3.4 B**).

3.4.4. *Activity and reconstitution analysis of Fe protein*

Activity of purified *Os*NifH^{Ht} was determined using the *in vitro* acetylene reduction assay (ARA) and compared with NifH purified from *A. vinelandii* (NifH^{Av}). In all activity assays NifDK purified from *A. vinelandii* (NifDK^{Av}) was used as MoFe protein component. *Os*NifH^{Ht} activities varied among preparations, with a maximum specific activity of 54.5 ± 11.6 nmol nmol of ethylene formed $min^{-1} \cdot mg^{-1}$. These results constitute

the first-time demonstration of a monocot purified Fe protein, capable of donating electrons to bacterial MoFe protein (NifDK^{Av}). In addition, the activity of *Os*NifH^{Ht} confirmed its ability to incorporate endogenous rice mitochondrial [Fe-S] clusters (**Figure 3.6**).

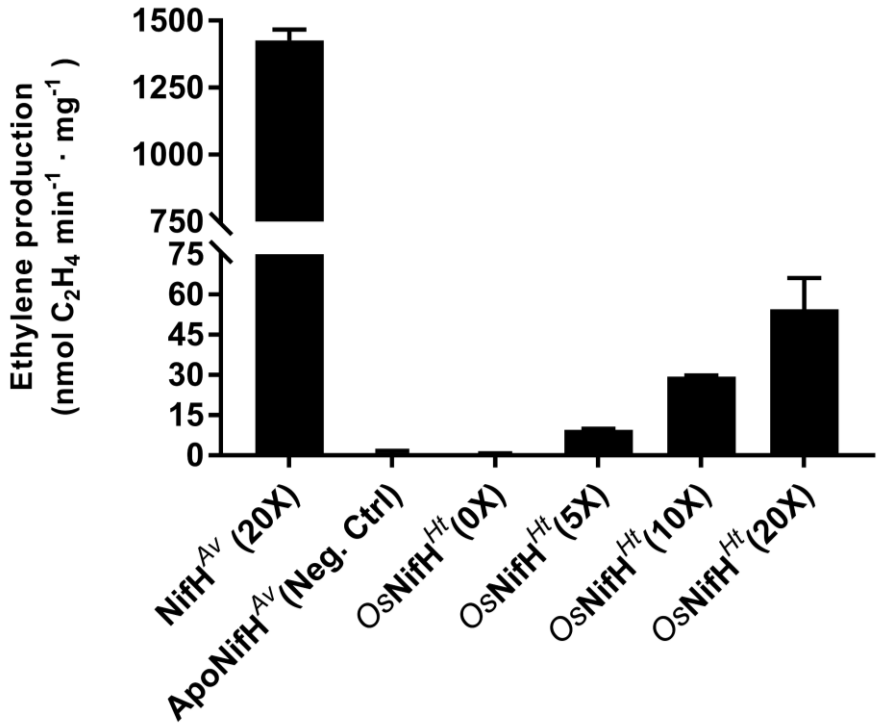


Figure 3.6. Enzymatic activity of *Os*NifH^{Ht} protein. Acetylene reduction assay (ARA) of STAC-purified *Os*NifH^{Ht}. Positive control reactions performed with NifH^{Av} and NifDK^{Av} at 20:1 molar ratio was 1426 ± 40 units. The negative control (without NifH^{Av}) was 1 ± 0.1 units. The maximum activity using *Os*NifH^{Ht} (as isolated) obtained at 20:1 molar ratio of *Os*NifH^{Ht} to NifDK^{Av} was 54.5 ± 11.6 units nmol ethylene formed per min and mg of NifDK^{Av}. Data represent mean values (n = 3 technical replicates).

Furthermore, the reconstitution of isolated rice Fe protein by the *in vitro* transfer of [4Fe-4S] from A. *vinelandii* NifU (NifU^{Av}) increased *Os*NifH^{Ht} activity nine-fold (54.5 to 465 nmol) compared to the isolated *Os*NifH^{Ht} protein (**Figure 3.7 A**). The measured activity of the positive control NifH^{Av} was ca. 1500 nmol and activities of *Sc*NifH^{Ht} and *Os*NifH^{Ht} were identical (ca. 465 nmol of ethylene formed $min^{-1} \cdot mg^{-1}$). These results demonstrate for the first time that the activities of reconstituted rice and yeast Fe protein are comparable and identical in substrate (N_2) reduction.

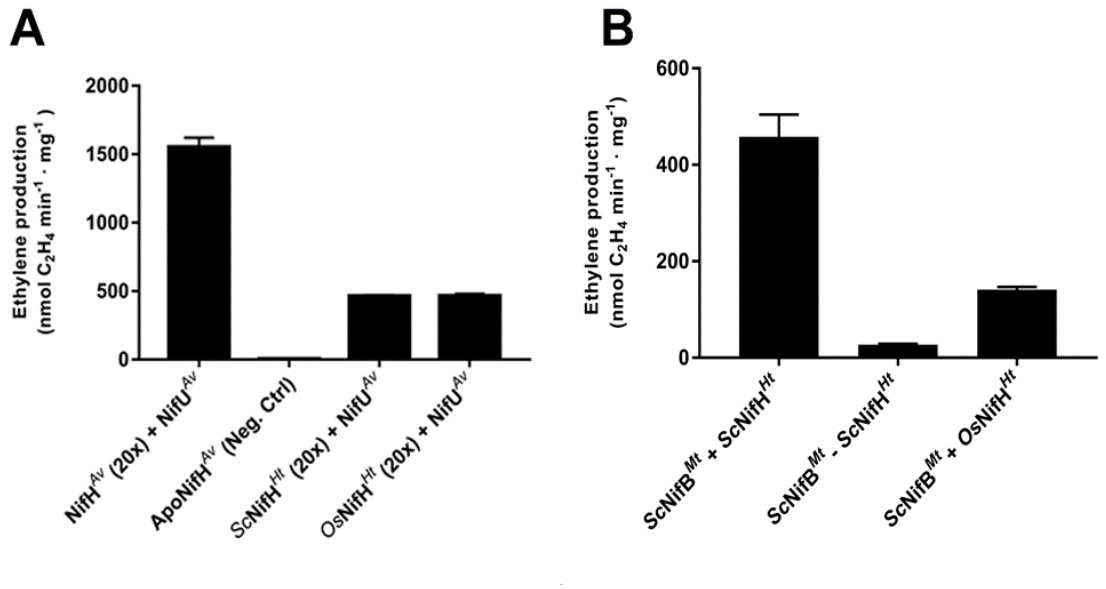


Figure 3.7. Reconstitution of active rice Fe protein. A) Reconstitute activities of *Os*NifH^{Ht} and *Sc*NifH^{Ht} with [4Fe4S] cluster loaded *Ec*NifU^{Av} (NifU-mediated reconstitution). Control reactions using holo-NifH^{Av} (NifU-mediated reconstitution) was 1500 ± 67 units (nmol ethylene formed per min and mg of NifDK^{Av}). Negative control reactions using non-reconstituted apo-NifH was 4 units. *Os*NifH^{Ht} and *Sc*NifH^{Ht} with [4Fe4S] cluster loaded NifU-mediated reconstitution was 465 ± 6.9 nmol ethylene formed per min and mg of NifDK^{Av}. B) NifB-dependent in vitro FeMo-co synthesis using NifH^{Av}. 20:1 molar ratio of *Os*NifH^{Ht} to NifDK^{Av} was used in the ARA. Data represent mean values (n = 3 technical replicates).

3.4.5. *In vitro FeMo-co biosynthesis*

The isolated rice Fe protein, in addition to electron transfer for substrate reduction, also supported FeMo-co biosynthesis *in vitro* which is one of the fundamental roles of Fe protein required for engineering nitrogen fixation. *Os*NifH^{Ht} and *Sc*NifB^{Mt} catalysed jointly the NifB-dependent *in vitro* FeMo-co biosynthesis. Activities were ca. 150 nmol in rice vs. 450 nmol in yeast (**Figure 3.7 B**). This critical result confirmed compatibility of two essential proteins (NifH and NifB) for N_2 fixation. It also demonstrated interspecies compatibility with NifDK^{Av} and NifEN^{Av} *(*scaffold proteins for NifB-co maturation into FeMo-co), altogether constituting the conserved biochemical core of nitrogenase.

3.5. Discussion

Engineering nitrogen fixation in plants, improving the ability of plant and plant-associated organisms to fix and assimilate atmospheric nitrogen has been a major long-standing goal of plant biotechnology for decades. Strategies to reduce the global dependence on nitrogen fertilizers are urgently needed in view of the excessive use of N fertilizers and their concomitant environmental problems in industrialized countries on one hand, on the other hand to overcome the lack of N fertilizers in most developing countries especially in sub-Saharan Africa. Early strategies to remedy the situation mostly concentrated on increasing nitrogen fixing bacterial colonization in close proximity to the root in order to fix atmospheric nitrogen and improve release and assimilation of NH_3 produced in plant cells (Stoltzfus *et al*., 1997; Mueller *et al.*, 2014., Mus *et al*., 2016).

The first strategy aimed to develop novel symbiosis mechanisms in non-legume plants, i.e., to engineer cereals to associate with N_2-fixing bacteria to form nodules (similarly to legumes) (Charpentier and Oldroyd, 2010; Rogers and Oldroyd, 2014; Mus *et al.,* 2016). Two main bottlenecks still remain. Firstly, the plant has to be able to interact with N-fixing bacteria and the bacteria have to recognize the plant as a suitable host. The second bottleneck is the formation of nodules or nodule-like structure to provide the low-O_2 environment necessary to enable interchange of nutrients, mainly C, N and metals between bacteria and plant (Burén *et al*., 2018). It is recognized that Myc factors, involved in signaling between soil fungi and most plants (including cereals) when forming symbiotic arbuscular mycorrhiza are similar to Nod factors secreted by symbiotic N_2-fixing bacteria. As Myc-factors are already recognized by most plants, engineering cereals capable of perceiving Nod-factors may be one approach to engineering N_2 fixation (Maillet *et al*., 2011, Rogers and Oldroyd, 2014). To date there has not been any further update on the status of this approach.

A more straightforward strategy to achieve BNF in plants is to transfer prokaryotic *nif* genes into the plant genome itself and engineer the N_2-fixing machinery to make plants able to fix their own N without bacterial interactions (Curatti and Rubio, 2014). However, the genetic complexity of the *nif* regulon (set of a number of operons used to regulate nitrogen fixation), the extreme O_2 sensitivity of nitrogenase, the requirement of multiple accessory proteins for metal cluster formation and maturation of the nitrogenase

components are the major barriers that make this process unachievable to date (Dixon and Kahn, 2004; Temme, Zhao and Voigt, 2012; Poza-Carrión *et al*., 2014, Burén *et al*., 2018). However, a number of studies have demonstrated the potential to overcome each barrier, to varying degrees (López-Torrejón *et al.,* 2016; Ivleva *et al.,* 2016; Burén *et al*., 2017; Eseverri *et al*., 2020, Jiang *et al.,* 2021).

The first study reporting the production of active Fe protein in yeast demonstrated that mitochondria were a suitable sub-cellular organelle for expressing O_2-sensitive Nif proteins under aerobic growth conditions. The measured activity of *A. vinelandii* NifH in yeast was 400 nmol of ethylene formed $min^{-1} \cdot mg^{-1}$ (López-Torrejón *et al*., 2016). Although this study demonstrated Fe protein activity for the first time, reproducibility of active *A. vinelandii* NifH was limited by low yields because only a very small portion of the enzyme was soluble in the mitochondrial matrix. The first active Fe protein in higher plants was obtained by stable expression of the *Klebsiella pneumoniae* NifH in the tobacco plastid genome (Ivleva *et al.,* 2016). Similar solubility issues were reported. Moreover, active Fe protein was obtained only when plants were incubated at low oxygen concentration (10% oxygen). In addition, the levels of activities were bearly above the negative control levels (max. 10 nmol of ethylene formed $min^{-1} \cdot mg^{-1}$) (Ivleva *et al.,* 2016). Allen *et al*. (2017) reported the individual expression of 16 *nif*- or *nif*-related *Klebsiella oxytoca* genes and the corresponding protein accumulation in the mitochondrial matrix of *N. benthamiana*, using transient expression. However, no enzymatic activity was shown for any of the sixteen proteins.

NifH is also essential for the assembly of both NifDK cofactors, namely the P-cluster and the FeMo-co besides electron transfer to NifDK (Burén *et al.,* 2020). Therefore, Fe protein (dinitrogenase reductase) for engineering nitrogen fixation in plants has to accumulate at high levels, be stable and soluble in the subcellular location where expressed, and needs to be compatible with the NifDK (dinitrogenase) component for N_2 reduction. In *A. vinelandii* and other well studied diazotrophs NifM is involved in NifH folding or dimerization prior to [4Fe4S] cluster acquisition (Burén *et al*., 2020). NifM co-expression was found to be necessary for soluble NifH accumulation in the stroma cells of *N. benthamiana* (Eseverri *et al.,* 2020). However, Jiang *et al*. (2021) demostrated that of *H. thermofilus* NifH was soluble and accumulated at high levels in the mitochondrial matrix of *S. cerevisiae* and *N. benthamiana* with or without NifM co-expression.

In my experiments, I co-expressed NifH^{Ht} with NifM^{Av}. In yeast, NifH^{Ht} was highly active (ca. 450 nmol) as isolated, however in *N. benthamiana* it was active only after reconstitution. These results confirmed that the presence of [Fe-S] clusters in rice and *N. benthamiana* mitochondria was relatively low compared to yeast. I found that the reconstituted activity of rice NifH^{Ht} was nine-fold higher than as isolated (50 vs. 450 nmol). This confirms rice mitochondria provides [Fe-S] clusters up to some extend but in order to improve the system in the future it will be crucial to transfer additional genes for [Fe-S] cluster formation, mainly *nifS* and *nifU*. While they were not essential for functional NifH^{Av} in *S. cerevisiae* mitochondria, they were required to generate high amounts of active NifB in yeast (Burén *et al.*, 2019). Eseverri *et al*. (2020) demonstrated the transient expression and accumulation of soluble *A. vinelandii* NifH, M, U and S proteins within the stroma of mesophyll cells and the activity of NifH and NifU proteins in *N. benthamiana*. NifH activity was detectable when NifH was used as-isolated and reconstituted as well [isolated activity was 189 nmol of ethylene $min^{-1} \cdot mg^{-1}$ (16% of *A. vinelandii* NifH activities)]. In rice NifH^{Ht} enzymatic activity was ca. 50 nmol of ethylene $min^{-1} \cdot mg^{-1}$ (as isolated) and 465 nmol of ethylene production $min^{-1} \cdot mg^{-1}$ (reconstituted). NifH^{Av} purified from *N. benthamiana* chloroplasts was active when it was co-expressed with NifU^{Av} and NifS^{Av}. This shows the importance of mitochondrial [Fe-S] clusters because in rice only expression of NifH^{Ht} and NifM^{Av} without NifS and NifU resulted in NifH activity. Eseverri *et al*. (2020) performed the same experiments in *N. benthamiana* leaves expressing only NifH^{Av} and NifM^{Av}. In the absence of NifU and NifS, no *Nb*NifH^{Av} activity could be detected, suggesting that the chloroplast assembly and transfer factors are not able to provide enough Fe-S clusters to result in a functional NifH, unlike the results I obtained in rice.

H. thermophilus NifH (NifH^{Ht}) was shown to be superior after screening 32 distinct origin NifH proteins in terms of expression levels, solubility, and functionality. About 6mg of *Nb*NifH^{Ht} was consistently isolated per kg of *N. benthamiana* leaves. Transient expression of *H. thermophilus* NifH in *N. benthamiana* showed lower nitrogenase activity as isolated (max. 25 nmol of ethylene formed $min^{-1} \cdot mg^{-1}$). However, reconstituted NifH by *in vitro* transfer of [4Fe4S] clusters from NifU increased the activity to 250 nmol of ethylene formed $min^{-1} \cdot mg^{-1}$ demonstrating that presence and assembly of plant mitochondrial [Fe-S] cluster is a major bottleneck for further engineering the entire nitrogenase enzyme complex (Jiang *et al*., 2021).

In rice, I demonstrated the expression of mitochondria-targeted NifH^{Ht} and NifM^{Av} at the mRNA and the protein level in stably transformed callus and regenerated plants. I also demonstrated enzymatic activity for the first time in any major staple crop. I purified relatively low amounts of NifH^{Ht} (2mg of *Os*NifH^{Ht} per kg). However, in rice activity was twice as high as in *N. benthamiana* as isolated (ca. 25 vs. 50 (rice) nmol of ethylene formed $min^{-1} \cdot mg^{-1}$) and reconstituted (ca. 250 vs. 465 (rice) nmol of ethylene formed $min^{-1} \cdot mg^{-1}$). Furthermore, reconstituted activity of *Os*NifH^{Ht} was found identical to reconstituted activity of yeast-expressed NifH^{Ht} (both 465 nmol ethylene formed $min^{-1} \cdot mg^{-1}$). The formation of FeMo-co is one of the most important outcomes of my study which confirms the funcionality of *Os*NifH^{Ht} and compatibility of two essential proteins (NifH^{Ht} and NifB^{Mt}) even though they are from totally distinct origins.

3.6. Conclusions

Rice mitochondria are capable of accumulating active Fe protein (NifH) which is one of the two structural components of nitrogenase, despite its extreme O_2 sensitivity.The endogenous mitochondrial [Fe-S] cluster assembly machinery is able to provide metal clusters to NifH and the activity of NifH as isolated enzyme confirmed its ability to incorporate endogenous rice mitochondrial [Fe-S] clusters.The reconstitution of NifH by the *in vitro* transfer of [4Fe-4S] from NifU increased NifH activity nine-fold compared to the as-isolated NifH protein. This is the first demonstration of the stable expression and activity of a nitrogenase component in any cereal. NifH carried out the fundamental roles of Fe protein required to engineer nitrogen fixation, including electron transfer to NifDK for N_2 reduction and FeMo-co biosynthesis. FeMo-co biosynthesis, confirmed that rice Fe protein also formed functional interspecies interactions with NifB, NifEN, and NifDK proteins constituting the core of diazotrophy.

3.7. References

Allen, R. S., Tilbrook, K., Warden, A. C., Campbell, P. C., Rolland, V., Singh, S. P., & Wood, C. C. (2017). Expression of 16 Nitrogenase Proteins within the Plant Mitochondrial Matrix. *Frontiers in Plant Science*, 8, 287.

Baysal, C., Pérez-González, A., Eseverri, Á., Jiang, X., Medina, V., Caro, E., Rubio, L., Christou, P., & Zhu, C. (2019). Recognition motifs rather than phylogenetic origin influence the ability of targeting peptides to import nuclear-encoded recombinant proteins into rice mitochondria. *Transgenic Research*, 29, 37–52.

Baysal, C., He, W., Drapal, M., Villorbina, G., Medina, V., Capell, T., Khush, G. S., Zhu, C., Fraser, P. D., & Christou, P. (2020). Inactivation of rice starch branching enzyme IIb triggers broad and unexpected changes in metabolism by transcriptional reprogramming. *Proceedings of the National Academy of Sciences of the United States of America*, 117, 26503–26512.

Burén, S., Young, E. M., Sweeny, E. A., Lopez-Torrejón, G., Veldhuizen, M., Voigt, C. A., & Rubio, L. M. (2017). Formation of Nitrogenase NifDK Tetramers in the Mitochondria of Saccharomyces cerevisiae. *ACS Synthetic Biology*, 6, 1043–1055.

Burén, S., López-Torrejón, G., & Rubio, L. M. (2018). Extreme bioengineering to meet the nitrogen challenge. *Proceedings of the National Academy of Sciences of the United States of America*, 115, 8849–8851.

Burén, S., Pratt, K., Jiang, X., Guo, Y., Jimenez-Vicente, E., Echavarri-Erasun, C., Dean, D. R., Saaem, I., Gordon, D. B., Voigt, C. A., & Rubio, L. M. (2019). Biosynthesis of the nitrogenase active-site cofactor precursor NifB-co in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America*, 116, 25078–25086.

Burén, S., Jimenez-Vicente, E., Echavarri-Erasun, C. & Rubio, L. M. (2020). Biosynthesis of nitrogenase cofactors. *Chemical Reviews*, 120, 4921–4968.

Chapin, F.S.III., Matson, P.A., & Vitousek, P. (2011). *Principles of Terrestrial Ecosystem Ecology*. New York, NY: Springer Science & Business Media.

Charpentier, M., & Oldroyd, G. (2010). How close are we to nitrogen-fixing cereals? *Current Opinion in Plant Biology*, 13, 556–564.

Cui, S., Shi, Y., Groffman, P. M., Schlesinger, W. H., & Zhu, Y. G. (2013). Centennial-scale analysis of the creation and fate of reactive nitrogen in China (1910-2010). *Proceedings of the National Academy of Sciences of the United States of America*, 110, 2052–2057.

Curatti, L., Hernandez, J. A., Igarashi, R. Y., Soboh, B., Zhao, D., & Rubio, L. M. (2007). In vitro synthesis of the iron-molybdenum cofactor of nitrogenase from iron, sulfur, molybdenum, and homocitrate using purified proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 17626–17631.

Curatti, L., & Rubio, L. M. (2014). Challenges to develop nitrogen-fixing cereals by direct nif-gene transfer. *Plant Science*, 225, 130–137.

Dixon, R., & Kahn, D. (2004). Genetic regulation of biological nitrogen fixation. *Nature Reviews. Microbiology*, 2, 621–631.

Dobermann, A., & Cassman, K. G. (2004). Environmental dimensions of fertilizer nitrogen: What can be done to increase nitrogen use efficiency and ensure global food security? In: A. R. Mosier, J. K. Syers, and J. R. Freney (eds.) *Agriculture and the Nitrogen Cycle: Assessing the Impacts of Fertilizer Use on Food Production and the Environment*, pp. 261–278.

Dos Santos, P. C., Dean, D. R., Hu, Y., & Ribbe, M. W. (2004). Formation and insertion of the nitrogenase iron-molybdenum cofactor. *Chemical Reviews*, 104, 1159–1173.

Dos Santos, P. C., Fang, Z., Mason, S. W., Setubal, J. C., & Dixon, R. (2012). Distribution of nitrogen fixation and nitrogenase-like sequences amongst microbial genomes. *BMC Genomics*, 13, 162.

Eady, R. R., Smith, B. E., Cook, K. A., & Postgate, J. R. (1972). Nitrogenase of *Klebsiella pneumoniae*. Purification and properties of the component proteins. *The Biochemical Journal*, 128, 655–675.

Echavarri-Erasun, C., Gonzalez, E., & y Rubio Herrero, L.M. (2015). *A novel method to isolate native NiFB-cofactor*. In: "*XV Congress of the Spanish Society of Nitrogen Fixation, and the IV Portuguese-Spanish Congress on Nitrogen Fixation*", 16/06/2015-18/06/2015, León. p. 1.

Eseverri, Á., López-Torrejón, G., Jiang, X., Burén, S., Rubio, L. M., & Caro, E. (2020). Use of synthetic biology tools to optimize the production of active nitrogenase Fe protein in chloroplasts of tobacco leaf cells. *Plant Biotechnology Journal*, 18, 1882–1896.

Farré, G., Sudhakar, D., Naqvi, S., Sandmann, G., Christou, P., Capell, T., & Zhu, C. (2012). Transgenic rice grains expressing a heterologous ρ-hydroxyphenylpyruvate dioxygenase shift tocopherol synthesis from the γ to the α isoform without increasing absolute tocopherol levels. *Transgenic Research*, 21, 1093–1097

Gavini, N., Tungtur, S., & Pulakat, L. (2006). Peptidyl-prolyl cis/trans isomerase-independent functional NifH mutant of Azotobacter vinelandii. *Journal of Bacteriology*, 188, 6020–6025.

Ivleva, N. B., Groat, J., Staub, J. M., & Stephens, M. (2016). Expression of Active Subunit of Nitrogenase via Integration into Plant Organelle Genome. *PLOS ONE*, 11, e0160951.

Jiang, X., Payá-Tormo, L., Coroian, D., García-Rubio, I., Castellanos-Rueda, R., Eseverri, Á., López-Torrejón, G., Burén, S., & Rubio, L. M. (2021). Exploiting genetic diversity and gene synthesis to identify superior nitrogenase NifH protein variants to engineer N_2-fixation in plants. *Communications Biology*, 4, 4.

Kronzucker, H. J., and Coskun, D. (2015). "Bioengineering nitrogen acquisition in rice: promises for global food security," In: F. J. de Bruijn (ed.), *Biological Nitrogen Fixation* (Hoboken, NJ: JohnWiley & Sons, Inc.), 47–56.

Lawlor D.W., Lemaire G., Gastal F. (2001) Nitrogen, Plant Growth and Crop Yield. In: Lea P.J., Morot-Gaudry JF. (eds.), *Plant Nitrogen*. Springer, Berlin, Heidelberg.

López-Torrejón, G., Jiménez-Vicente, E., Buesa, J. M., Hernandez, J. A., Verma, H. K., & Rubio, L. M. (2016). Expression of a functional oxygen-labile nitrogenase component in the mitochondrial matrix of aerobically grown yeast. *Nature Communications*, 7, 11426.

Maillet, F., Poinsot, V., André, O., Puech-Pagès, V., Haouy, A., Gueunier, M., Cromer, L., Giraudet, D., Formey, D., Niebel, A., Martinez, E. A., Driguez, H., Bécard, G., & Dénarié, J. (2011). Fungal lipochitooligosaccharide symbiotic signals in arbuscular mycorrhiza. Nature, 469, 58–63. Merrick, M., & Dixon, R. (1984). Why don't plants fix nitrogen? *Trends in Biotechnology*. 2, 162.

Mueller, N.D., West, P.C., Gerber, J.S., MacDonald, G.K., Stephen Polasky, S., & Foley, J.A. (2014). A tradeoff frontier for global nitrogen use and cereal production. *Environmental Research Letters*, 9, 5.

Mus, F., Crook, M. B., Garcia, K., Garcia Costas, A., Geddes, B. A., Kouri, E. D., Paramasivan, P., Ryu, M. H., Oldroyd, G., Poole, P. S., Udvardi, M. K., Voigt, C. A., Ané, J. M., & Peters, J. W. (2016). Symbiotic nitrogen fixation and the challenges to its extension to nonlegumes. *Applied and Environmental Microbiology*, 82, 3698–3710.

Ohyama, T. (2010). Nitrogen as a major essential element of plants. In: Ohyama, T. and Sueyoshi, K., (eds), *Nitrogen Assimilation in Plants*, Research Singpot, Kerala, 1-18.

Paustian, T.D., Shah, V.K., &Roberts, G.P. (1989). Purification and characterization of the *nifN* and *nifE* gene products from *Azotobacter vinelandii* mutant UW45. *Proceedings of the National Academy of Sciences of the United States of America*, 86, 6082-6086.

Poza-Carrión, C., Jiménez-Vicente, E., Navarro-Rodríguez, M., Echavarri-Erasun, C., & Rubio, L. M. (2014). Kinetics of *nif* gene expression in a nitrogen-fixing bacterium. *Journal of Bacteriology*, 196, 595–603.

Rogers, C., & Oldroyd, G. E. (2014). Synthetic biology approaches to engineering the nitrogen symbiosis in cereals. *Journal of Experimental Botany*, 65, 1939–1946.

Rubio, L. M., & Ludden, P. W. (2008). Biosynthesis of the iron-molybdenum cofactor of nitrogenase. *Annual Review of Microbiology*, 62, 93–111.

Shah, V.K. & Brill, W.J. (1977). Isolation of an iron-molybdenum cofactor from nitrogenase (nitrogenfixation). *Proceedings of the National Academy of Sciences, USA*, 74, 3249 3253

Stoltzfus, J.R., So, R., Malarvithi, P.P., Ladha, J.K., & de Bruijn, F.J. (1997). Isolation of endophytic bacteria from rice and assessment of their potential for supplying rice with biologically fixed nitrogen. *Plant and Soil*, 194, 25–36).

Sutton, M.A., Oenema, O., Erisman, J.W., Leip, A., van Grinsven, H., & Winiwarter, W. (2011) Too much of a good thing. *Nature*, 472, 159–161.

Temme, K., Zhao, D., & Voigt, C. A. (2012). Refactoring the nitrogen fixation gene cluster from Klebsiella oxytoca. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 7085–7090.

Tilman, D., Cassman, K. G., Matson, P. A., Naylor, R., & Polasky, S. (2002). Agricultural sustainability and intensive production practices. *Nature*, 418, 671–677.

Yuvaniyama, P., Agar, J. N., Cash, V. L., Johnson, M. K., & Dean, D. R. (2000). NifS-directed assembly of a transient [2Fe-2S] cluster within the NifU protein. *Proceedings of the National Academy of Sciences of the United States of America*, 97, 599–604.

Zehr, J. P., Jenkins, B. D., Short, S. M., & Steward, G. F. (2003). Nitrogenase gene diversity and microbial community structure: a cross-system comparison. *Environmental Microbiology*, 5, 539–554.

Chapter 4

Inactivation of rice starch branching enzyme IIb triggers broad and unexpected changes in metabolism by transcriptional reprogramming

4.0. Abstract

Starch properties can be modified by mutating genes responsible for the synthesis of amylose and amylopectin in the endosperm. However, little is known about the effects of such targeted modifications on the overall starch biosynthesis pathway and broader metabolism. Here I investigated the effects of mutating the *OsSBEIIb* gene encoding starch branching enzyme IIb, which is required for amylopectin synthesis in the endosperm. As anticipated, homozygous mutant plants, in which OsSBEIIb was completely inactivated by abolishing the catalytic center and C-terminal regulatory domain, produced opaque seeds with depleted starch reserves. Amylose content in the mutant increased from 19.6 to 27.4% and resistant starch (RS) content increased from 0.2% to 17.2%. Many genes encoding isoforms of AGPase, soluble starch synthase and other starch branching enzymes were upregulated, either in their native tissues or in an ectopic manner, whereas genes encoding granule bound starch synthase, debranching enzymes, pullulanases and starch phosphorylases were largely downregulated. There was a general increase in the accumulation of sugars, fatty acids, amino acids and phytosterols in the mutant endosperm, suggesting that intermediates in the starch biosynthesis pathway increased flux through spillover pathways causing a profound impact on the accumulation of multiple primary and secondary metabolites. These results provide insights into the broader implications of perturbing starch metabolism in rice endosperm and its impact on the whole plant, which will make it easier to predict the effect of metabolic engineering in cereals for nutritional improvement or the production of valuable metabolites.

4.1. Introduction

Starch is the major storage polysaccharide in higher plants. It is a mixture of linear amylose, in which glucose residues are linked by α(1→4) glycosidic bonds, and amylopectin, in which amylose-like chains feature additional branches formed by α(1→6) glycosidic bonds every 24–30 glucose units. Rice starch typically consists of ~20% amylose and ~80% amylopectin (Tetlow, 2010; Crofts *et al*., 2015). Amylose synthesis requires only two enzymes: ADP-glucose pyrophosphorylase (AGPase) and granule-bound starch synthase (GBSS). In contrast, the synthesis of amylopectin involves the combined activity of AGPase, soluble starch synthases (SS), starch branching enzymes (SBE), debranching enzymes (DBE), pullulanase (PUL) and phosphorylases (Pho), the

latter forming a complex with disproportionating enzyme (Tian *et al*., 2009; Qu *et al*., 2018). The rice genome encodes seven AGPase subunits, ten isoforms of SS, three isoforms of SBE, four isoforms of DBE and two isoforms of Pho, with different expression profiles and activities, underpinning the complex regulation of starch metabolism in this species (Ohdan *et al*., 2005).

SBE is the only enzyme that introduces α(1→6) glycosidic bonds into α-polyglucans, and it is therefore essential for amylopectin biosynthesis (Nakamura, 2002). The three isoforms in rice (OsSBEI, OsSBEIIa and OsSBEIIb) play distinct roles, with OsSBEI favoring predominantly linear amylose-like substrates and the others favoring amylopectin substrates and thus acting to increase the density of branching (Satoh *et al*., 2003a; Nakamura *et al*., 2010). OsSBEIIb is the major isozyme found in rice seeds due to the strong endosperm-specific expression profile of the *OsSBEIIb* gene. OsSBEIIb also has the unique ability to transfer short chains to the crystalline lamellae of amylopectin, whereas OsSBEIIa lacks this function (Nishi *et al*., 2001; Nakamura, 2002; Ohdan *et al*., 2005). *OsSBEIIa* is expressed in all rice tissues including the developing endosperm, and is the major isoform found in leaves (Nishi *et al*., 2001; Nakamura, 2002; Ohdan *et al*., 2005).

The inactivation of OsSBEI and/or OsSBEIIa does not affect seed morphology in rice (Satoh *et al*., 2003b; Abe *et al*., 2014; Sawada *et al*., 2018) whereas the inactivation of OsSBEIIb generates *amylose extender* (*ae*) mutants containing starch molecules with longer amylopectin chains and fewer branches, increasing the amylose content and producing an opaque seed phenotype (Nishi *et al*., 2001; Abe *et al*., 2014; Asai *et al*., 2014). In rice cultivar Kinmaze, the *ae* mutation was shown to reduce the dry seed weight by 35% compared to wild-type plants, and to increase the amylose content of the seeds from 15.7% to 26.5% (Nishi *et al*., 2001).

The inactivation of individual starch biosynthesis enzymes in rice can have a wider effect on general starch metabolism, causing the modulation of genes encoding other enzymes in the pathway (Nishi *et al*., 2001; Asai *et al*., 2014). Interestingly, even when the target gene is endosperm-specific, the ripple effects can be evident in vegetative tissues, as shown for *OsGBSSI* (Pérez *et al*., 2019) and *OsAPL2*, encoding the large subunit of cytosolic AGPase (Soto *et al*., 2018). This may reflect the accumulation of intermediates

that trigger feedback inhibition and/or the effect of enzyme depletion or structural disruption on the activity of metabolons (enzyme consortia that function as supramolecular complexes) (Crofts *et al*., 2015; 2018). The broader effects of metabolic interventions can be determined by comparing wild-type and mutant plants at the transcriptomic and metabolomic levels (Decourcelle *et al*., 2015; Galland *et al*., 2017), but this form of systematic analysis is rarely carried out and therefore the broader implications of perturbing starch metabolism in the endosperm and its impact on the whole plant are unclear.

4.2. Aims

- Evaluate the impact of mutating *OsSBEIIb* by analyzing the starch biosynthesis transcriptome and broader metabolome of a mutant line generated using the CRISPR/Cas9.

- Investigate the effect of the mutation on the structure and function of the OsSBEIIb protein, on the morphology of starch granules in the mutant endosperm, and on the properties of the extracted starch.

- Analyze the expression of entire starch biosynthesis pathway genes in the mutant endosperm and leaves.

- Determine how the inactivation of OsSBEIIb effect overall carbohydrate metabolism, as well as primary and secondary metabolism more broadly.

4.3. Materials and methods

4.3.1. *Plant material*

I used a homozygous *OsSBEIIb* mutant rice line (*Oryza sativa* cv Nipponbare) that we generated previously using the CRISPR/Cas9 system (Baysal *et al.,* 2016). The mutant is characterized by a 4-bp deletion immediately downstream of the catalytic domain of OsSBEIIb (Baysal *et al.,* 2016). I used wild-type segregants (azygous controls) from the original mutant line for comparison, as well as a Tos17 insertion line (NE9005) from the Rice Tos17 Insertion Mutant Database (https://tos.nias.affrc.go.jp/) as an alternative negative control for grain morphology and starch physiochemical properties. In the

NE9005 line, the transposon is integrated within *OsSBEIIb* intron 18 but displays no overt phenotype in terms of morphology or starch characteristics.

4.3.2. *Gene expression analysis by quantitative real-time PCR*

Total RNA was isolated from T2 leaf and T3 endosperm tissue using the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany). Mutant and control plants were grown in flooded trays for seed sampling, and spikelets were marked and harvested 15 DAF. On the same day, three leaf blades (the uppermost expanding leaf) were collected from each of three individual plants (biological replicates). First-strand cDNA was synthesized from 1 μg total RNA using Ominiscript Reverse Transcriptase (Qiagen) and quantitative real-time RT-PCR (qRT-PCR) was carried out as previously described by (Jin *et al*., 2019) using the gene-specific primers listed in **Table 4.1**. The identity of the PCR products was confirmed by sequencing. Expression levels were normalized against *OsActin* mRNA. Three biological replicates (leaves from siblings of homozygous T2 plants and the corresponding T3 seeds) each comprising three technical replicates were tested for each gene.

4.3.3. *Homology modeling and protein structure analysis*

The *OsSBEIIb* sequence was translated and the mature OsSBEIIb polypeptide (761 amino acids, omitting the 65 residues of the N-terminal transit peptide region) was aligned with the mutant version from line E15 and the maize ortholog ZmSBEIIb using the Unipro UGENE alignment platforms Kalign, MUSCLE, and ClustalW (Okonechnikov *et al.*, 2012). Automated homology modeling was carried out using Phyre2 (Kelly *et al.*, 2015). The model of the mutant protein was superimposed on the wild-type version using DS Visualizer (https://www.3dsbiovia.com/products/collaborative-science/biovia-discoverystudio/visualization.html).

Table 4.1. List of primers used for qRT-PCR analysis (obtained from supplementary Ohdan *et al.*, 2005).

Enzyme	Gene name	Accession number	Primer sequence	Amplicon size (bp)
1-ADP-glucose pyrophosphorylase small subunit 1	*OsAGPS1*	AK073146	[F] GTGCCACTTAAAGGCACCATT [R] CCCACATTTCAGACACGGTTT	97
2-ADP-glucose pyrophosphorylase small subunit 2a	*OsAGPS2a*	AK071826	[F] ACTCCAAGAGCTCGCAGACC [R] GCCTGTAGTTGGCACCCAGA	147
3-ADP-glucose pyrophosphorylase small subunit 2b*	*OsAGPS2b*	AK103906	[F] AACAATCGAAGCGCGAGAAA [R] GCCTGTAGTTGGCACCCAGA	186
4-ADP-glucose pyrophosphorylase large subunit 1	*OsAGPL1*	D50317	[F] GGAAGACGGATGATCGAGAAAG [R] CACATGAGATGCACCAACGA	140
5-ADP-glucose pyrophosphorylase large subunit 2*	*OsAGPL2*	U66041	[F] AGTTCGATTCAAGACGGATAGC [R] CGACTTCCACAGGCAGCTTATT	96
6-ADP-glucose pyrophosphorylase large subunit 3	*OsAGPL3*	AK069296	[F] AAGCCAGCCATGACCATTTG [R] CACACGGTAGATTCACGAGACAA	131
7-ADP-glucose pyrophosphorylase large subunit 4	*OsAGPL4*	AK121036	[F] TCAACGTCGATGCAGCAAAT [R] ATCCCTCAGTTCCTAGCCTCATT	77
8-Starch synthase I	*OsSSI*	D16202	[F] GGGCCTTCATGGATCAACC [R] CCGCTTCAAGCATCCTCATC	279
9-Starch synthase IIa	*OsSSIIa*	AF419099	[F] GCTTCCGGTTTGTGTGTTCA [R] CTTAATACTCCCTCAACTCCACCAT	54
10-Starch synthase IIb	*OsSSIIb*	AF395537	[F] TAGGAGCAACGGTGGAAGTGA [R] GTGAACGTGAGTACGTGACCAAT	89
11-Starch synthase IIc	*OsSSIIc*	AF383878	[F] GACCGAAATGCCTTTTTCTCG [R] GGGCTTGGAGCCTCTCCTTA	256
12-Starch synthase IIIa	*OsSSIIIa*	AY100469	[F] GCCTGCCCTGGACTACATTG [R] GCAAACATATGTACACGGTTCTGG	334
13-Starch synthase IIIb	*OsSSIIIb*	AF432915	[F] ATTCCGCTCGCAAGAACTGA [R] CAACCGCAGGATAACGGAAA	224
14-Starch synthase IVa	*OsSSIVa*	AY100470	[F] GGGAGCGGCTCAAACATAAA [R] CCGTGCACTGACTGCAAAAT	237
15-Starch synthase IVb	*OsSSIVb*	AY373258	[F] ATGCAGGAAGCCGAGATGTT [R] ACGACAATGGGTGCCAAGAT	75
16-Granule-bound starch synthase I	*OsGBSSI*	X62134	[F] AACGTGGCTGCTCCTTGAA [R] TTGGCAATAAGCCACACACA	218
17-Granule-bound starch synthase II	*OsGBSSII*	AY069940	[F] AGGCATCGAGGGTGAGGAG [R] CCATCTGGCCCACATCTCTA	246
18-Starch branching enzyme I	*OsSBEI*	D11082	[F] TGGCCATGGAAGAGTTGGC [R] CAGAAGCAACTGCTCCACC	191
19-Starch branching enzyme IIa	*OsSBEIIa*	AB023498	[F] GCCAATGCCAGGAAGATGA [R] GCGCAACATAGGATGGGTTT	128
20-Starch branching enzyme IIb	*OsSBEIIb*	D16201	[F] ATGCTAGAGTTTGACCGC [R] AGTGTGATGGATCCTGCC	261
21-Starch debranching enzyme: Isoamylase I	*OsISA1*	AB093426	[F] TGCTCAGCTACTCCTCCATCATC [R] AGGACCGCACAACTTCAACATA	132
22-Starch debranching enzyme: Isoamylase II	*OsISA2*	AC132483	[F] TAGAGGTCCTCTTGGAGG [R] AATCAGCTTCTGAGTCACCG	170
23-Starch debranching enzyme: Isoamylase III	*OsISA3*	AP005574	[F] ACAGCTTGAGACACTGGGTTGAG [R] GCATCAAGAGGACAACCATCTG	100
24-Starch debranching enzyme: Pullulanase	*OsPUL*	AB012915	[F] ACCTTTCTTCCATGCTGG [R] CAAAGGTCTGAAAGATGGG	202
25-Starch phosphorylase L	*OsPHOL*	AK063766	[F] TTGGCAGGAAGGTTTCGCT [R] CGAAGCCTGAAGTGAACTTGCT	66
26-Starch phosphorylase H*	*OsPHOH*	AK103367	[F] CACCAAGACGAAGCTCATCAAG [R] TTCACTCGTTGCTGGGTTCTC	126
* Cytosolic type				

4.3.4. *Amylose and resistant starch content*

The relative amylose content in 25mg of whole-grain flour was determined using the AMYLOSE/AMYLOPECTIN kit (Megazyme International, Wicklow, Ireland) according to the manufacturer's instructions. The resistant starch content in 100mg of whole-grain flour was determined using the RESISTANT STARCH kit (Megazyme International, AOAC Official Method 2002.02 and AACC Method 32-40.01). The quantities of solubilized and resistant starch were calculated on a dry weight basis incorporating relative moisture values based on the recommendations provided in the kit

(McClary *et al.*, 2002). The resistant starch content was added to the amount of solubilized (non-resistant) starch to determine the total starch content.

4.3.5. *Light microscopy and scanning electron microscopy*

Rice seeds were dried in an oven at 37°C for one week before analysis. The opacity or translucency of complete and transversely-cut rice seeds were observed using incident and transmitted light and a black background under a Leica MZ8 stereo microscope coupled to a Leica DFC digital camera. The morphology of starch grains in transversely-cut rice seeds was observed by scanning electron microscopy. Three seeds (three independent biological replicates) were mounted on aluminum stubs coated with double-stick carbon tape and sputter coated with gold using a Baltec SDC 050 sputter coater. Images were captured using a JEOL JSM-6360LV instrument.

4.3.6. *Metabolite analysis and identification*

Freeze-dried rice seeds (~200mg) were ground to a fine powder using a TissueRuptor (Qiagen) and samples of 10±0.5mg were transferred to plastic tubes for methanol–chloroform extraction as previously described (Drapal *et al*., 2019). For GC/MS analysis, 100-µl aliquots of the polar phase were dried with the internal standard d4-succinic acid (5µg) and 700-µl aliquots of the non-polar phase were dried with d27-myristic acid (10µg). The samples were derivatized immediately before injection and were introduced into the GC/MS device using splitless mode. Metabolites were identified based on their retention times, retention indices and mass spectra, and the area quantified relative to the internal standards (Fraser *et al*., 2000; Nogueira *et al*., 2013).

4.3.7. *Data analysis and statistical tests*

Grain morphology and starch physiochemical properties in the *OsSBEIIb* mutant and Tos17 insertion line were compared to wild-type rice by one-way analysis of variance (ANOVA). Significant differences were determined using Fisher's least significant difference (LSD) test ($p < 0.05$ and $p < 0.01$). For each treatment, the standard deviation (SD) of the mean (SEM) was calculated based on at least three biological replicates. Student's t-test was used for gene expression analysis with thresholds $p < 0.05$ and $p < 0.01$. For metabolomic datasets, principal component analysis (PCA) was applied after auto-scaling using MetaboAnalyst v4.0 (Xia and Wishart, 2016). Changes in metabolites

were visualized in heat maps and were overlaid with biochemical pathways constructed using BioSynLab (Royal Holloway University of London). Significant differences in the relative abundance of different metabolites were determined using Student's t-test ($p < 0.05$).

4.4. Results

4.4.1. *The mutation site in OsSBEIIb abolishes the catalytic center of the enzyme.*

The *OsSBEIIb* gene was targeted by using the CRISPR/Cas9 system to introduce a double-strand break in exon 12, resulting in a 4-bp deletion that generated a frameshift and premature stop codon (Baysal *et al*., 2016). T0 transformants heterozygous for the mutant allele were self-pollinated over two generations to yield homozygous T2 progeny lacking the cas9 transgene. We used this line (E15) to investigate changes at the protein level that might explain the loss of OsSBEIIb activity. We translated the mutant sequence and aligned it with wild-type OsSBEIIb (**Figure. 4.1.A**) allowing us to identify key functional residues (**Figure 4.1.B**). Homology modeling was then used to determine the effect of the mutation on the structure and key functional residues of the enzyme (**Figure 4.2)**.

The catalytic domains of SBEs are highly conserved in plants (Tethlow and Emes, 2014; Que *et al*., 2018). Seven amino acids in the maize ortholog ZmSBEIIb are necessary for catalysis (Asp^{376}, His^{381}, Arg^{445}, Asp^{447}, Glu^{502}, His^{569}, and Asp^{570}) in addition to the catalytic triad (Asp447, Glu502, and Asp570) which was previously shown to reside in the central domain (Tethlow and Emes, 2014). The corresponding amino acids in the OsSBEIIb protein are Asp^{402}, His^{407}, Arg^{471}, Asp^{473}, Glu^{528}, His^{595}, and Asp^{596}, supporting the catalytic triad Asp^{473}, Glu^{528}, and Asp^{596}, indicating that the two sequences are offset by 26 residues (**Figure 4.2.B**). The 4-bp deletion in line E15 occurred immediately downstream of Gly^{418} and the frameshift generated a new sequence of 31 amino acids followed by a termination codon **(Figure 4.2.A**).

A) Amino acid alignment and comparison of amino acid sequences of wild type SBEIIb protein versus mutant SBEIIb protein.

>Rice SBEIIb Protein

MAAPASAVPGSAAGLRAGAVRFPVPAGARSWRAAAELPTSRSLLSGRRFPGAVRVGGSGGRVAVRAAGASGEVMIPEGESDGMPVSAGSDDLQLPALDDELSTEVGAEVEIESSGAS
DVEGVKRVVEELAAEQKPRVVPPTGDGQKIFQMDSMLNGYKYHLEYRYSLYRRLRSDIDQYEGGLETFSRGYEKFGFNHSAEGVTYREWAPGAHSAALVGDFNNWNPNADRMSKN
EFGVWEIFLPNNADGSSPIPHGSRVKVRMETPSGIKDSIPAWIKYSVQAAGEIPYNGIYYDPPEEEKYIFKHPQPKRPKSLRIYETHVGMSSTEPKINTYANFRDEVLPRIKKLGYNAVQI
MAIQEHAYYGSFGYHVTNFFAPSSRFGTPEDLKSLIDKAHELGLVVLMDVVHSHASNNTLDGLNGFDGTDTHYFHSGSRGHHWMWDSRLFNYGNWEVLRFLLSNARWWLEEYKFDG
FRFDGVTSMMYTHHGLQVAFTGNYSEYFGFATDADAVVYLMLVNDLIHGLYPEAITIGEDVSGMPTFALPVQDGGVGFDYRLHMAVPDKWIELLKQSDESWKMGDIVHTLTNRRWSEKC
VTYAESHDQALVGDKTIAFWLMDKDMYDFMALDRPATPSIDRGIALHKMIRLITMGLGGEGYLNFMGNEFGHPEWIDFPRAPQVLPNGKFIPGNNNSYDKCRRRFDLGDADYLRYRGM
LEFDRAMQSLEEKYGFMTSDHQYISRKHEEDKMIIFEKGDLVFVFNFHWSNSYFDYRVGCLKPGKYKVVLDSDAGLFGGFGRIHHTAEHFTADCSHDNRPYSFSVYSPSRTCVVYAPAEStop

>Rice SBEIIb Mutant (E15) Protein

MAAPASAVPGSAAGLRAGAVRFPVPAGARSWRAAAELPTSRSLLSGRRFPGAVRVGGSGGRVAVRAAGASGEVMIPEGESDGMPVSAGSDDLQLPALDDELSTEVGAEVEIESSGAS
DVEGVKRVVEELAAEQKPRVVPPTGDGQKIFQMDSMLNGYKYHLEYRYSLYRRLRSDIDQYEGGLETFSRGYEKFGFNHSAEGVTYREWAPGAHSAALVGDFNNWNPNADRMSKN
EFGVWEIFLPNNADGSSPIPHGSRVKVRMETPSGIKDSIPAWIKYSVQAAGEIPYNGIYYDPPEEEKYIFKHPQPKRPKSLRIYETHVGMSSTEPKINTYANFRDEVLPRIKKLGYNAVQI
MAIQEHAYYGSFGYHVTNFFAPSSRFGTPEDLKSLIDKAHELGLVVLMDVVHSHASNNTLDGLNGLVQIRITFIVVHAAIIGCGILAFSTMGIGKFStop

B) Amino acid residues shown as different colors represent different domains of rice SBEIIb protein. Sequence alignment and homology modelling was performed and modified from the data of Qu et al 2018.

OsSBEIIb 178 Y E 206 W 208 P 233 E 234 F 423 D T 448 W E 452 R — Maltopentaose Binding Sites

OsSBEIIb 222 W 258 K 269 D 288 N — Glucose Binding Sites I

OsSBEIIb 160 H 163 Y 481 T 498 G F 501 T 503 A — Glucose Binding Sites II

OsSBEIIb 367 Y 402 D 407 H 471 R 473 D 528 E 595 H — Catalytic residues

OsSBEIIb 368 H 428 H 596 D 598 A — Active Sites I

OsSBEIIb 324 S T - 663 W I D F P 680 P — *CD Binding Sites I

OsSBEIIb 621 D 636 K 685 S Y — CD Binding Sites II

OsSBEIIb 538 L P 545 V G F D 587 K — CD Binding Sites III

OsSBEIIb - 259 V 282 A 286 P — CD Binding Sites IV

OsSBEIIb 632 I 634 L H 755 W 758 S 786 F — CD Binding Sites V

OsSBEIIb 574 D 577 H T 580 T N 729 Y — CD Binding Sites VI

OsSBEIIb 225 P 329 I N - - — CD Binding Sites VII

*cyclodextrin (CD)-binding sites

Figure 4.1. Characterization of the mutant *OsSBEIIb* sequence. (A) Alignment of the wild-type OsSBEIIb protein with the mutant version from line E15. The N-terminal transit peptide region is shown in brown (65 amino acids). The frameshift mutation generated 31 new amino acids and a termination codon (red). (B) Amino acid residues in different colors represent the various important regions of the OsSBEIIb protein. Sequence alignment and homology modelling was performed and modified from data shown in reference 10

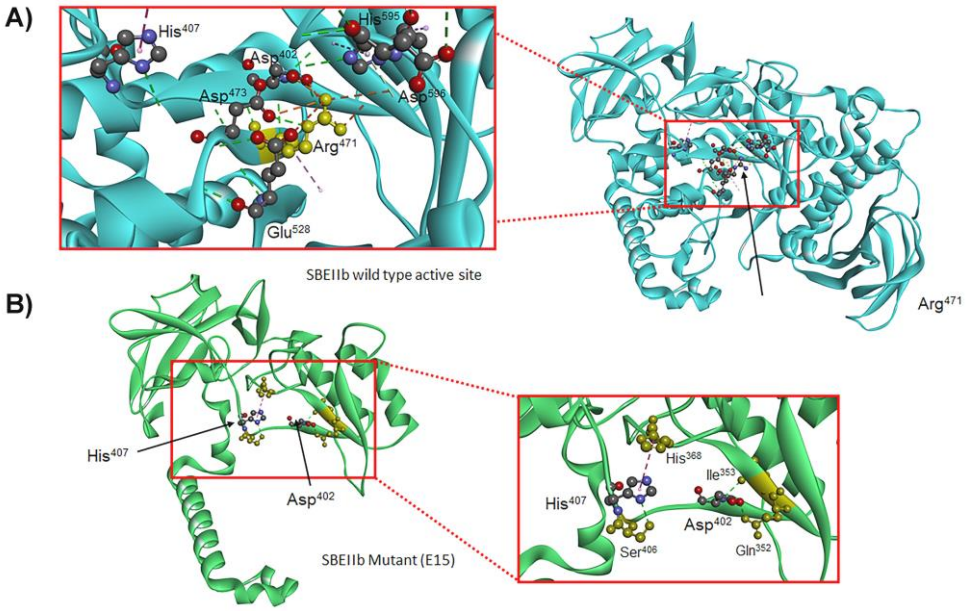


Figure 4.2. Structural comparison of wild-type OsSBEIIb and the mutated version in line E15. The protein is represented as a ribbon, whereas all of the labeled residues are represented as scaled ball and stick models. (A) Wild-type OsSBEIIb showing the position of Arg471. The Inset box shows all of the key amino acids in the central catalytic domain (Asp402, His407, Arg471 [yellow], Asp473, Glu528, His595, and Asp596). Dashed lines show the intramolecular interactions (orange, salt bridge; green, hydrogen bond; purple, π-interactions). (B) Mutant OsSBEIIb in line E15 showing the position of His407 and Asp402. The Inset box shows that His407 has the same interactions with His368 and Ser406 as shown for the wild-type protein, but Asp402 no longer interacts with Arg471 but retains the normal interactions with lle353 and Gln352. Amino acids that interact with His407 and Asp402 are colored yellow. Dashed lines show the intramolecular interactions (green, hydrogen bond; purple, π-interactions).

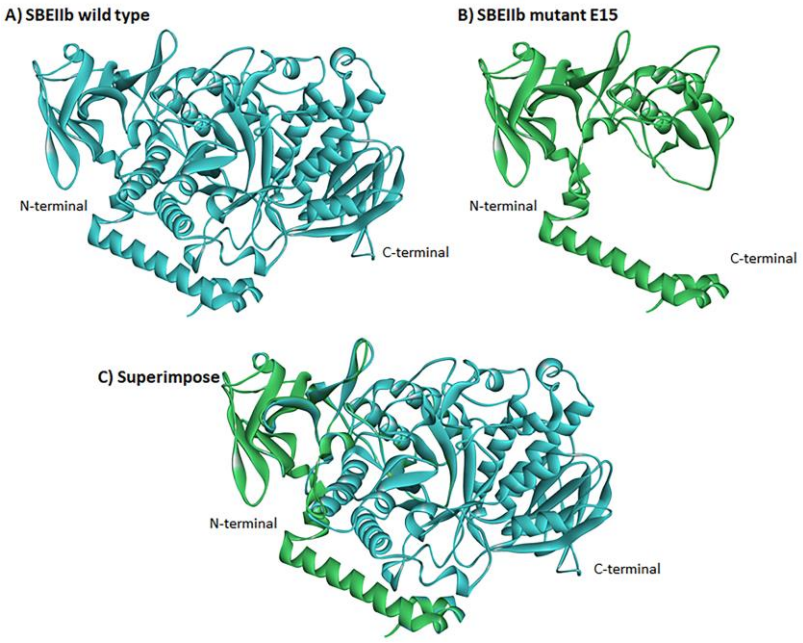


Figure. 4.3. Comparison of the three-dimensional models of wild-type OsSBEIIb and the mutant version in line E15. (A) Wild-type OsSBEIIb. (B) Mutant version in line E15. (C) Superimposition of the wild-type and mutant versions.

The comparative models of wild-type OsSBEIIb and the mutant version revealed that the deletion and frameshift caused the complete loss of the catalytic triad (**Figure 4.3.A**) as well as the Arg471 residue that plays a key role in stabilizing the catalytic center (**Figure 4.2.A** and **B**). The E15 mutant retains only residues Asp402 and His407 of the central catalytic domain and these are insufficient to stabilize the enzyme structure (**Figure 4.3.B**). The loss of OsSBEIIb activity in line E15 therefore reflects the disruption of the catalytic center and loss of the regulatory C-terminal domain that determines substrate specificity.

4.4.2. *Loss of OsSBEIIb activity alters the starch grain content and structure.*

The endosperm amylose content of the T3 seeds in line E15 was 27.4%, 1.4-fold higher than that of wild-type seeds (19.6%) and a Tos17 insertion line (NE9005) with the transposon integrated into intron 18 of the *OsSBEIIb* gene (19.5%). The mutation also substantially increased the resistant starch content from 0.2% in wild-type and NE9005 seeds to 17.2% in line E15 (**Figure 4.4.A** and **Figure 4.5**). However, the total starch content was ~26% lower in line E15, resulting in significantly smaller seeds with lower dry seed and dehulled grain weights. The dry weight of the E15 seeds was 14.7 ± 1.9 mg (compared to 24.9 ± 2.8 mg for wild-type seeds) and the dehulled grain weight was 13.1 ± 1.4 mg (compared to 18.2 ± 2.4 mg for wild-type seeds) (**Table 4.2**.).

Table 4.2. Phenotypic characteristics and starch composition of wild-type seeds, Tos17 insertion line NE9005 and OsSBEIIb mutant line E15.

Sample	Seed weight (mg)	Kernel weight (mg)	Kernel length (mm)	Kernel width (mm)	Kernel area (mm^2)	L/W Ratio	Amylose Content (%, w/w)	Resistant Starch Content (%, w/w)	Total Starch Content (%, w/w)
Wild type	24.9±2.8	18.2±2.4	4.9±0.3	2.8±0.1	10.4±0.8	1.8±0.1	19.6±0.7	0.2±0	87±0.5
NE9005	25.9±2.6	19.9±2.4	5.1±0.2	2.9±0.1	11.1±1	1.8±0.1	19.5±0.7	0.2±0.2	85.4±0.5
E15	14.7±1.9	13.5±1.4	4.7±0.2	2.3±0.1	7.8±0.6	2.1±0.1	27.4±0.8	17.2±0.5	67,7±3,2

Morphological analysis by light microscopy revealed that the E15 grains were opaque, whereas those of the wild-type and NE9005 lines were uniformly translucent (**Figure 4.4.B, i–vi**). Scanning electron microscopy revealed that the loss of OsSBEIIb activity had a profound influence on the structure of starch grains. In wild-type lines, the starch grains were homogeneous, compact, and angular, with few interstitial spaces, whereas those in the E15 mutant varied considerably in size, were rounded rather than angular, and separated by large gaps (**Figure 4.4.B, vii–ix**).

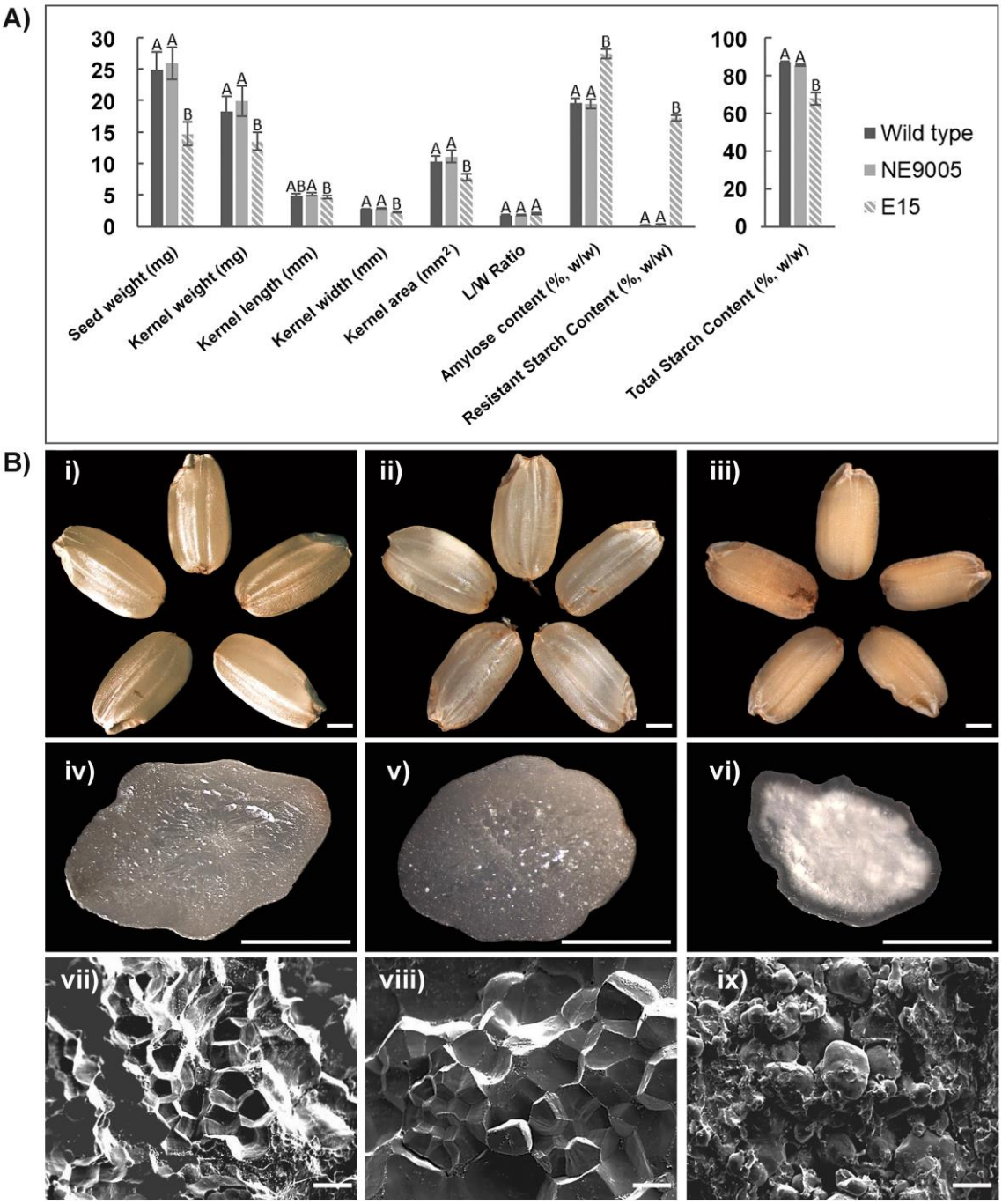


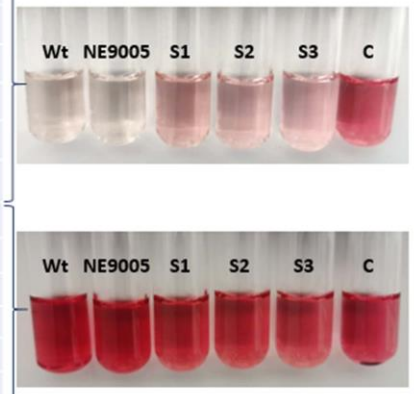
Figure 4.4.. Starch properties and grain morphology of wild-type rice, Tos17 insertion line NE9005, and mutant line E15. (A) Properties of the starch in all three genotypes. Samples annotated with different letters are significantly different from one another as determined by ANOVA ($P < 0.01$). (B) Gross morphology (i–iii) and transverse sections (iv–vi) of the seeds, and scanning electron microscopy images (vii–ix) of starch grains in the endosperm. (Scale bars, 1 mm in i–vi; 5 μm in vii–ix.)

Absorbance values for 100 micrograms of D-glucose standard

Rep. 1	Rep. 2	Rep. 3	Rep. 4	Average Abs
1,1060	1,1200	1,1430	1,1240	1,1233

89,0274 Factor [=100 (micrograms of D-glucose)/Absorbance for 100 micrograms of D-glucose]

Sample identifier	Absorbance values (510 nm) R1	R2	Δ Abs	Sample weight (mg)	Extract volume (mL)	Resistant Starch (g/100 g) "as is"	Non Resistant Starch (g/100 g) "as is"	Moisture Content %	Resistant Starch (g/100 g) "dwb"	Non Resistant Starch (g/100 g) "dwb"	Total Starch (g/100 g) "dwb"
Wild type	0.025	0.0230	0.024	100	10.3*	0.1981	-	10.4	0.2211	-	87.3656
NE9005	0.0240	0.0230	0.0235	100	10.3*	0.1939	-	10	0.2155	-	85.2366
E15 (S1)	0.2090	0.1920	0.2005	100	100	16.0650	-	9.7	17.7907	-	70.4529
E15 (S2)	0.2070	0.1820	0.1945	100	100	15.5842	-	8.9	17.1067	-	68.383
E15 (S3)	0.1970	0.1830	0.1900	100	100	15.2237	-	9.2	16.7662	-	64.2409
Control 39%	0.4320	0.4110	0.4215	100	100	33.7725	-	12	38.3779	-	89.1387
Wild type	0.9740	0.9750	0.9745	100	100	-	78.0815	10.4	-	87.1445	-
NE9005	0.9630	0.9470	0.9550	100	100	-	76.5190	10		85.0211	-
E15 (S1)	0.5750	0.6120	0.5935	100	100	-	47.5540	9.7	-	52.6622	-
E15 (S2)	0.6080	0.5580	0.5830	100	100	-	46.7127	8.9	-	51.2763	-
E15 (S3)	0.5560	0.5200	0.5380	100	100	-	43.1071	9.2	-	47.4747	-
Control 39%	0.5400	0.5750	0.5575	100	100		44.6695	12	-	50.7608	-



R1 and R2 are technical replicates
S1, S2, S3 are 3 biological replicates (T3 seeds from 3 independent siblings) of *BEIIb* mutant line E15
*Extract volume for samples containing < 10% RS (See the protocol)
Rep. 1,2,3,4 are the absorbance values for 100 ug of D-glucose standard

Figure 4.5. Analysis of total starch and resistant starch. R1 and R2 are two technical replicates. S1, S2, S3 are three biological replicates (T3 seeds from 3 independent siblings) of OsSBEIIb mutant line E15. Rep. 1,2,3 and 4 are the absorbance values for 100 ug of D-glucose standard. *Extract volume for samples containing < 10% RS (See the protocol)

4.4.3. *Loss of OsSBEIIb cctivity causes the broad transcriptional reprogramming of starch metabolism*

To determine whether the loss of OsSBEIIb activity caused a ripple effect on starch metabolism, we compared the expression profiles of 26 genes involved in starch biosynthesis representing five classes of enzymes (AGPase, SS, SBE, DBE, and Pho). We carried out a separate analysis of the isoforms expressed specifically in the endosperm 15 d after fertilization (DAF) and those expressed in the leaves at the same time. Heat maps comparing the expression of these genes in the endosperm and leaves are shown in **Figure 4.6.A**. Rice AGPase is a tetramer of two large subunits and two small subunits. There are four large subunit genes (*OsAGPL1–4*) and two small subunit genes (*OsAGPS1* and *OsAGPS2*, the latter producing mRNA variants *OsAGPS2a* and *OsAGPS2b* by alternative splicing). Some of them are predominantly expressed in the endosperm (e.g., *OsAGPL2* and *OsAGPS2b*) while others are predominantly expressed in the leaves (e.g., *OsAGPS2a* and *OsAGPL3*) (5). We found that *OsAGPL1* and *OsAGPS1* were minimally expressed in the endosperm and leaves of both genotypes. *OsAGPS2a* was expressed at low levels in the leaves but was 1.5-fold higher in the E15 line than wild-type plants. *OsAGPS2b* was strongly expressed in the seeds and was 2-fold higher in the E15 line than wild-type plants. *OsAGPL2* was expressed more strongly in the endosperm than leaves of wild-type plants and there was a slight but statistically nonsignificant increase in the endosperm of line E15 but a striking 1.7-fold increase in the leaves. *OsAGPL3* was predominantly expressed in the leaves of wild-type plants, but we observed a significant increase in expression in both the leaves (1.7-fold) and endosperm (1.6-fold) in line E15. *OsAGPL4* expression was not detected in either tissue.

I also observed a major transcriptional reprogramming of genes encoding starch synthases. *OsSSI* was strongly expressed in the developing endosperm and leaves of wild-type plants and was up-regulated in both tissues in line E15, although the change was only statistically significant (2.2-fold) in the endosperm. *OsSSIIa* was only expressed in the endosperm and was significantly (1.6-fold) up-regulated in line E15. *OsSSIIb* was only expressed in the leaves of wild-type plants and a similar (1.6-fold) up-regulation was observed in the mutant. *OsSSIIc* was expressed in both tissues and I observed a significant (1.8-fold) increase in the E15 leaf tissue but no change in the endosperm.

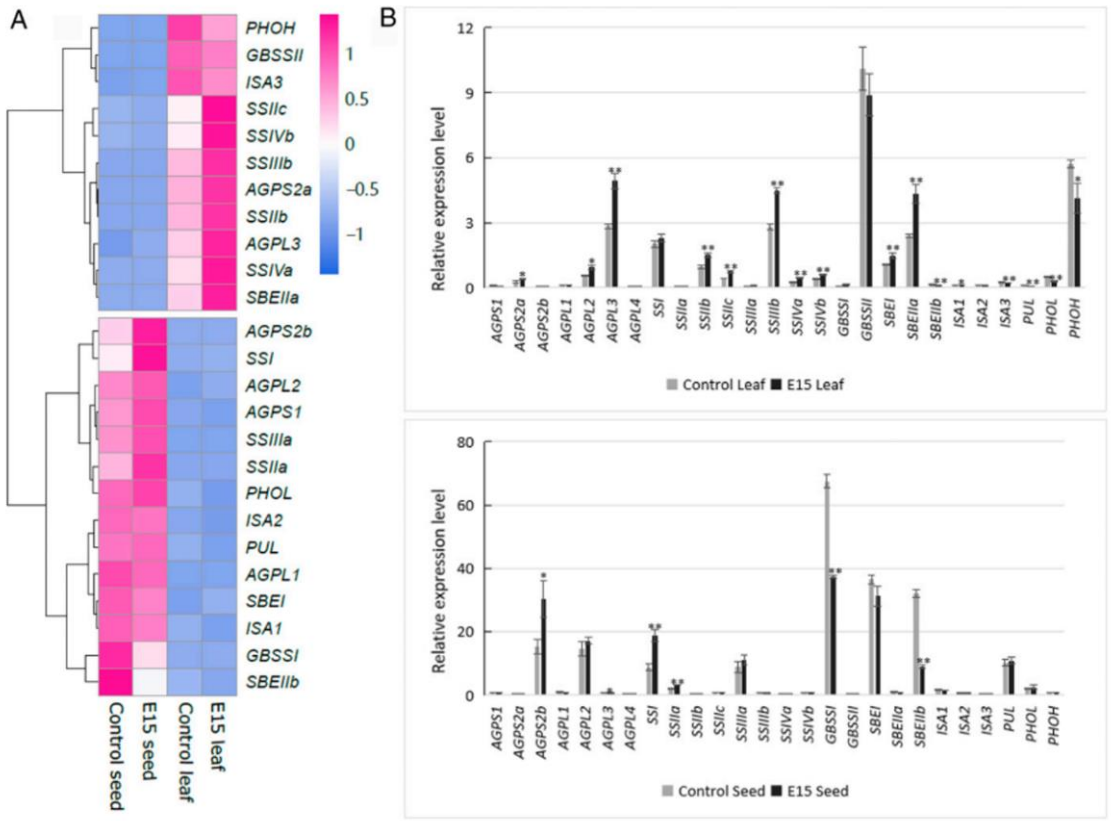


Figure 4.6. Expression analysis of 26 genes representing the starch biosynthesis pathway. (A) Heat map summarizing the transcriptional reprogramming of the starch biosynthesis pathway in the T2 leaves and T3 seeds of line E15. Fold changes are shown in red (increase) or blue (decrease) compared to wild type. (B) Expression level of all 26 genes relative to *OsActin* 15 DAF in the leaves (Top) and seeds (Bottom). Each value is the mean SD of at least three independent measurements with SEs, and significant differences were determined using Student's t test (*P < 0.05, **P < 0.01).

OsSSIIIa is endosperm specific and there was no significant difference between wild-type and E15 seeds. *OsSSIIIb*, *OsSSIVa* and *OsSSIVb* are expressed in both tissues but the only change we observed was in the leaf, where the expression levels of all three genes increased by 1.6-fold in the mutant. Finally, the endosperm-specific *OsGBSSI* gene was strongly down-regulated (1.8-fold) in the mutant, whereas the leaf-specific *OsGBSSII* gene was slightly down-regulated.

I observed the strong down-regulation of *OsSBEIIb* (3.5-fold) in the endosperm, possibly reflecting the direct effect of the truncated coding region on the mRNA quality control system via nonsense-mediated mRNA decay. *OsSBEIIb* was expressed at low levels in wild-type leaves, but even in this tissue we were able to detect a statistically significant down-regulation (1.5-fold) in line E15. *OsSBEI* was expressed in both tissues in wild-type plants, albeit at a higher level in the endosperm. In the mutant plants, there was no

significant change in the endosperm but a significant (1.4-fold) up-regulation in the leaves. *OsSBEIIa* was mainly expressed in the leaves of wild-type plants and it was significantly (1.8-fold) up-regulated in the leaves of mutant plants, whereas there was no change in the endosperm.

The DBE genes *OsISA1–3* were expressed at low levels in the endosperm and leaves of wild-type plants, and the significant changes were the down-regulations of *OsISA1* and *OsISA3* by 1.2-fold in the leaves. The pullulanase (*OsPUL*) and plastidial starch phosphorylase (*OsPHOL*) genes were expressed more strongly in the endosperm than the leaves, but there was no significant change in the endosperm of line E15, whereas *OsPUL* expression fell by 1.6-fold and *OsPHOL* by 1.9-fold in the mutant leaves. The cytosolic starch phosphorylase (*OsPHOH*) gene, which was mainly expressed in the leaves of wild-type plants, was also significantly (1.4-fold) down-regulated in the mutant leaves (**Figure 4.6.B** and **Table 4.3**).

Table 4. 3. Expression levels of 26 genes involved in starch biosynthesis in rice endosperm and leaves from *OsSBEIIb* mutant line E15 and a wild-type control. The data were normalized to the expression of *OsActin.*

	AGPS1	*AGPS2a*	*AGPS2b*	*AGPL1*	*AGPL2*	*AGPL3*	*AGPL4*	*SSI*	*SSIIa*	*SSIIb*	*SSIIc*	*SSIIIa*	*SSIIIb*
Control Leaf	0.02	0.21	0.00	0.05	0.55	2.86	0.01	1.86	0.00	0.90	0.40	0.02	2.68
	0.02	0.27	0.00	0.05	0.54	2.94	0.01	2.13	0.00	0.94	0.40	0.02	2.82
	0.04	0.29	0.00	0.05	0.53	2.74	0.01	2.06	0.00	1.04	0.41	0.02	2.92
E15 Leaf	0.02	0.42	0.00	0.06	0.97	5.00	0.01	2.46	0.00	1.62	0.80	0.02	4.48
	0.02	0.36	0.00	0.04	0.84	4.54	0.01	2.08	0.00	1.40	0.69	0.01	4.36
	0.02	0.38	0.00	0.04	1.00	5.22	0.01	2.33	0.00	1.50	0.74	0.02	4.63
	SSIVa	*SSIVb*	*GBSSI*	*GBSSII*	*SBEI*	*SBEIIa*	*SBEIIb*	*ISA1*	*ISA2*	*ISA3*	*PUL*	*PHOL*	*PHOH*
Control Leaf	0.23	0.40	0.07	8.99	1.05	2.33	0.14	0.06	0.09	0.24	0.13	0.48	5.50
	0.25	0.35	0.07	10.36	1.10	2.33	0.14	0.06	0.11	0.24	0.13	0.51	5.79
	0.27	0.40	0.07	10.92	1.04	2.49	0.15	0.06	0.12	0.25	0.14	0.49	5.87
E15 Leaf	0.47	0.59	0.07	7.92	1.62	4.40	0.10	0.05	0.09	0.22	0.08	0.29	4.01
	0.44	0.61	0.05	8.95	1.38	3.86	0.09	0.05	0.08	0.20	0.08	0.24	3.53
	0.48	0.58	0.07	9.84	1.43	4.75	0.09	0.05	0.08	0.20	0.08	0.28	4.89

	AGPS1	*AGPS2a*	*AGPS2b*	*AGPL1*	*AGPL2*	*AGPL3*	*AGPL4*	*SSI*	*SSIIa*	*SSIIb*	*SSIIc*	*SSIIIa*	*SSIIIb*
Control Seed	0.21	0.02	14.24	0.42	12.58	0.33	0.01	7.74	1.97	0.01	0.20	8.75	0.13
	0.28	0.02	13.75	0.38	13.85	0.30	0.01	8.73	1.94	0.01	0.21	7.19	0.12
	0.33	0.02	17.83	0.50	17.08	0.41	0.01	9.67	2.17	0.01	0.21	10.50	0.16
E15 Seed	0.37	0.02	36.05	0.40	18.22	0.61	0.01	18.67	3.33	0.01	0.23	12.82	0.13
	0.33	0.02	24.21	0.31	16.72	0.51	0.01	16.94	3.15	0.01	0.16	9.67	0.11
	0.35	0.02	30.68	0.33	16.28	0.50	0.01	20.67	3.07	0.01	0.19	11.13	0.12
	SSIVa	*SSIVb*	*GBSSI*	*GBSSII*	*SBEI*	*SBEIIa*	*SBEIIb*	*ISA1*	*ISA2*	*ISA3*	*PUL*	*PHOL*	*PHOH*
Control Seed	0.06	0.26	65.49	0.08	35.59	0.40	30.55	1.43	0.61	0.04	10.63	1.64	0.28
	0.06	0.25	67.36	0.08	35.84	0.37	32.95	1.62	0.59	0.04	8.82	1.89	0.25
	0.08	0.32	69.66	0.11	38.10	0.50	32.40	1.75	0.54	0.04	11.08	2.15	0.35
E15 Seed	0.07	0.22	37.59	0.09	27.65	0.40	9.61	1.54	0.67	0.05	12.00	1.81	0.35
	0.05	0.22	37.43	0.08	33.00	0.40	8.56	1.35	0.55	0.04	9.83	2.51	0.29
	0.06	0.22	36.66	0.08	33.25	0.40	8.91	1.44	0.45	0.04	10.69	3.15	0.31

4.4.4. *Loss of OsSBEIIb activity triggers changes in endosperm primary and secondary metabolism*

Having established that the loss of OsSBEIIb activity affected the expression of all classes of starch biosynthesis enzymes in seeds and also in leaves, we extracted polar and nonpolar metabolites from the endosperm of wild-type plants and line E15 in order to compare the broader metabolic profiles. Principal component analysis (PCA) revealed clear genotype-specific differences in metabolic composition between the wild-type and two biological replicates of the mutant. The first principal component (PC1) which is related to the metabolic differences mostly in sugars, fatty acids, sterols, amino acids, organic acids, and phenols explained 73.2% of total variance among the genotypes, while PC2 explained 5.9% between the genotypes (**Figure 4.7**). The comparison of extracts produced by gas chromatography/mass spectrometry (GC/MS) revealed 50 metabolites in the endosperm tissue (**Table 4.4**), of which 42 showed genotype-specific differences in the heat map (**Figure 4.8**) and pathway analysis (**Figure 4.9**).

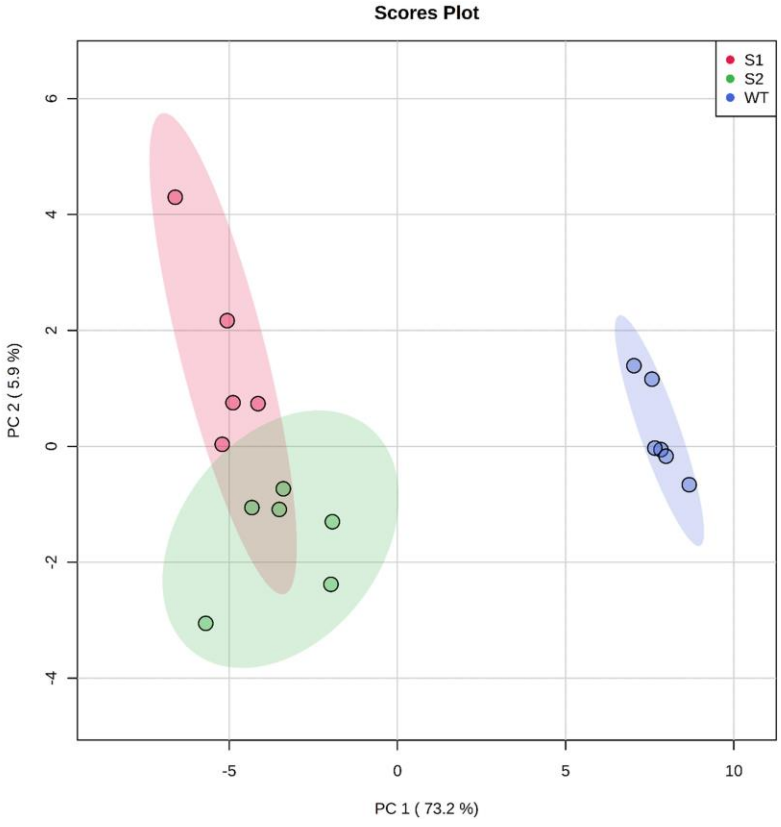


Figure 4.7. Principal component analysis (PCA) of polar and non-polar metabolites identified in rice endosperm. The wild-type seeds (blue circles) are compared to two biological replicates of line E15 mutant seeds (red and green circles). The score plots display the average of six technical replicates.

Table 4.4. Absolute values of all detected metabolites in seeds from *OsSBEIIb* mutant line E15 and a wild-type control (μg/g dry weight).

Genotype	Sample	Dodecanol	Alanine	Alkane-alcohol	Aspartic acid	a-Tocopherol	b-Sitosterol	C12_0	C14_0	C15_0	C16_0	C16_0-Glycerol	C18_0	C18_0-Glycerol	C18_2	C20_0	C22_0	Campesterol
WT	WT_1			26.48		23.69	397.88	9.87	43.3		2669.72	62.53	1389.14	41.1	1864.56	4.12		135.05
	WT_2	14.24				19.06	446.93	13.46	54.52		3414.23	78.89	1767.61	52.4	2324.18	4.24		172.53
	WT_3	12.61		34.93			462.49	12.53	59.96		3570.62	70.3	1669.88	46.65	2347.86	4.93	1.85	185.97
	WT_4	14.77		31.56			404.12	13.71	53.24		3604	72.22	1773.9	49.98	2411.19	4.36	2	164.98
	WT_5	12.37		37.08		19.14	412.52	14	52.84		3575.01	72.18	1723.31	47.81	2543.85	3.98		137.2
	WT_6	14.37		33.75		24.59	492.53	14.33	58.59		4128.9	82.4		53.7	2862.9	4.8	2.15	166.33
S1	S1_1	15.19	178.35	29.44	230.21	21.35	803.36	10.42	87.68		5908.2	101.51	2110.71		3547.04	9.7		245.96
	S1_2		221.11	33	276.97	19.63	799.53	10.9	97.83	2.19	5766.57	103.3	2118.32	58.61	3574.11	13.36		279.41
	S1_3	14.62	203.39	33.66	219.32	23.01	854.31	13.09	94.57	2.23	5619.54	98.94	2109.66	53.79	3458.38	13.06	3.74	277.41
	S1_4	16.56	160.47	33.72		19.91	785.68	13.33	100.98	2.33	5592.87	94.83	2122.88	54.95	3656.33	13.44	3.23	287.3
	S1_5	15.78	172	39.09			980.06		104.94	2.49	6606.35	108.29	2451.02	59.2	4295.19	14.07	5.1	328.84
S2	S2_1	14.35	140.01	31.7	245.51	17.89	762.86	11.78	92.78	2.27	5279.86	99.74	2063.35	47.35	3492.79	11.99	5.21	259.58
	S2_2		166.84	35.66	316.37	22.83	805.01	12.62	88.62	2.35	5312.93	104.7	2138.37	60.65	3421.7	11.1	3.56	278.41
	S2_3		134.06	30.49	222.07	17.51	730.05	12.64	96.06	2.16	5063.92	87.77	1906.09	50.33	3450.36	10.18	3.45	264.02
	S2_4	14.61	201.59	34.94		22.17	750.9	13.36	83.48	1.96	5445.54	96.57	2156.33	55.21	3226.66	9.6		288.51
	S2_5	13.46	221.36	33.84	330.99	25.52	798.67	13.11	92.64	2.1	5663.21	90.24	2108.64	50.46	3844.52	10.14		281.87
	S2_6	13.89	146.25		209.99	17.06	667.73	12.73	86.73	1.89	5331.94	86.17	2008.11	47.05	3553.12	10.85	2.83	235.83

Genotype	Sample	Citric acid	Disaccharide 50.7min	Dotriacontol	Eriostanyl-3,5-dinitro-benzoate	Fructose	Gluconic acid	Glucose	Glycerol	Glycerol-3-P	Glycerol-galactose	Glycine	Inositol	Lysine	Methyl-phosphate	N.C. Benzoic acid	N.C. Steroid	Phosphate
WT	WT_1		11.68	10.45	6.46	71.17		339.21		129.21			128.81		3.2		22.47	182.75
	WT_2	99.9	10.85	9.61	6.74	50.77		290.08	946.38	83.28			91.37		4.76	1.72	27.67	
	WT_3			9.37		66.52		300.75	859.84	102.73			121.89		4.78		26.56	153.25
	WT_4			9.6		62.73		324.56	1740.75	82.44			109.61		2.94	1.81		175.55
	WT_5	98.98	11.26	9.61		57.47		311.51	1182.65	72.2			104.34		3.41	1.75	27.05	224
	WT_6	81.36		10	7.03	57.67		262.06		87.24			112.29		6.14		29.98	222.09
S1	S1_1	168.02	33.45	6.31	15.68	1370.65	72.36	2982.62		294.7		35.5	211.94		23.03	1.75		
	S1_2	119.67	40.62	9.08	21.8	1583.6	73.21	3504.65	1348.19	243.7	1038.02	23.63	195.27	4.12			33.95	344.17
	S1_3	124.56			23.76				1293.7	283.52	1051.23	27.52	209.1	5.6	20.59	1.86	31.29	310.45
	S1_4	123.5	41.31	8.33	21.02	1399.56	69.56	3188.25	1039.3	257.51		25.97	206.54		13.28	1.78	28.22	216.8
	S1_5	180.19	27.65	9.81	23.81	1069.29		2653.71		314.46	1228.9	31.42	241.14	4.55	15.4			327.03
S2	S2_1	165.98	44.51	7.91	18.69	1311.56	51.88	2998.2		291.89		19.11	173.93	6.13		1.84	23.08	292.35
	S2_2	222.44		7.86	21.27	1614.81	66.92	3593.79	1387.71	292.86	980.2	28.78	198.47	8.37	13.99	1.76		207.22
	S2_3	141.01	22.22	6.91	16.71	1456.42	45.18	3335.11	1619.31	232.97	1023.35	22.37	171.6		13.02	1.67		254.68
	S2_4	203.95	30.61	7.03	19	1393.69	67.83	3170.45	1269.17	257.77	943.33	28.27	209.65	5.1	14.14		32.25	217.84

Table 4.4. cont.

	S2_5	160.94		6.71	19.91	1565.39	81.87	3545.32	1208.79	332.16	1002.98	38.3	253.54	7.17	17.36	1.93	34.36	208.71
	S2_6	171.44	29.25	5.71	17.08	1271.77	65.85	2967.77		274.21	926.77	26.61	186.31		14.64	1.79		
Genotype	Sample	Proline	Pentose 16.4min 1815	Disaccharide 30.7min 2926	Disaccharide 25.5min 3005	Disaccharide 33.3min 3820	Sedoheptulose	Serine	Stigmasterol	Sucrose	Fatty acid (unsaturated) 29.9min 2959	Sugar acid 24.2min	Triacontanol	C16_0-Glycerol isomer	Disaccharide 32.9min 3560	Valine	Sugar alcohol 24.4min 2066	
WT	WT_1		18.87		38.78	1216.39	24.26		124.81	76774.52		55.1			12.61			
	WT_2		13.67		25.58		17.48		141.22	94254.37	261.87	37.22	3.47		15.62			
	WT_3		14.86		33.47				138.28	88665.39	221.29	42.61	3.6					
	WT_4		13.58		25.51	1257.08	18.15		123.59	76006.76	252.02	38.47	3.49					
	WT_5		14.15		28.81	1175.7	19.37		125.68		213.59	40.97	3.64					
	WT_6				29.39		16.77		139.72			45.65	3.61		14.87			
S1	S1_1	142.2	21.78	30.94	75.94	456.39	5.14	17.8	251.85	135046.5	293.54	96.21		2.94	31.98		43.07	
	S1_2	93.77	28.15		57.84	399.5	6.29	34.68	239.15	117710.85	307.13	80.77	4.17	2.94	33.34	6.21	55.9	
	S1_3	165.42	23.12	26.01	68.25	485.06		30.92	253.36	178375.15	278.95	96.6		5.15	30.88	4.84	54.86	
	S1_4	135.93	25.95	24.04	62.32	473.82	6.31	17.57	206.96	191392.78	262.55	90.36	4.92	5.15	29.47	3.06	47.04	
	S1_5	148.41	33.36	26.81	79.31	480.99	10.15	17.69	261.94	173897.27	334.14	101.61	4.96	2.31	32.19	2.6	45.4	
S2	S2_1	80.78	19.24	21.53	62.64	574.32		34.8	222.22	103389.38	282.68	85.7		5.92	29.15	4.18	41.03	
	S2_2				57.8	332	3.67	38.55	225.78		241.59	88.01	4.12	2.49	32.89	5.5	34.47	
	S2_3	91.09	19.04		48.74	359.78	5.2	20.07	185.6	145222.7		71.48	4.27		28.51		37.72	
	S2_4	119.59	25.68		53.59	334.52	4.84	28.8	200.23	167166.56	252.36	84.39	3.92		27.76		35.41	
	S2_5	181.54	27.75	21.36	59.87	530.23	4.47	49.17	210.71	221359.24	263.77	99.88	5	4.19	27.48	7.6	36.9	
	S2_6	83.74	19.61	19.36	50.68	302.96	12.58	25.44	183.57	149067.11	237.16	80.94			24.7	3.9	37.48	

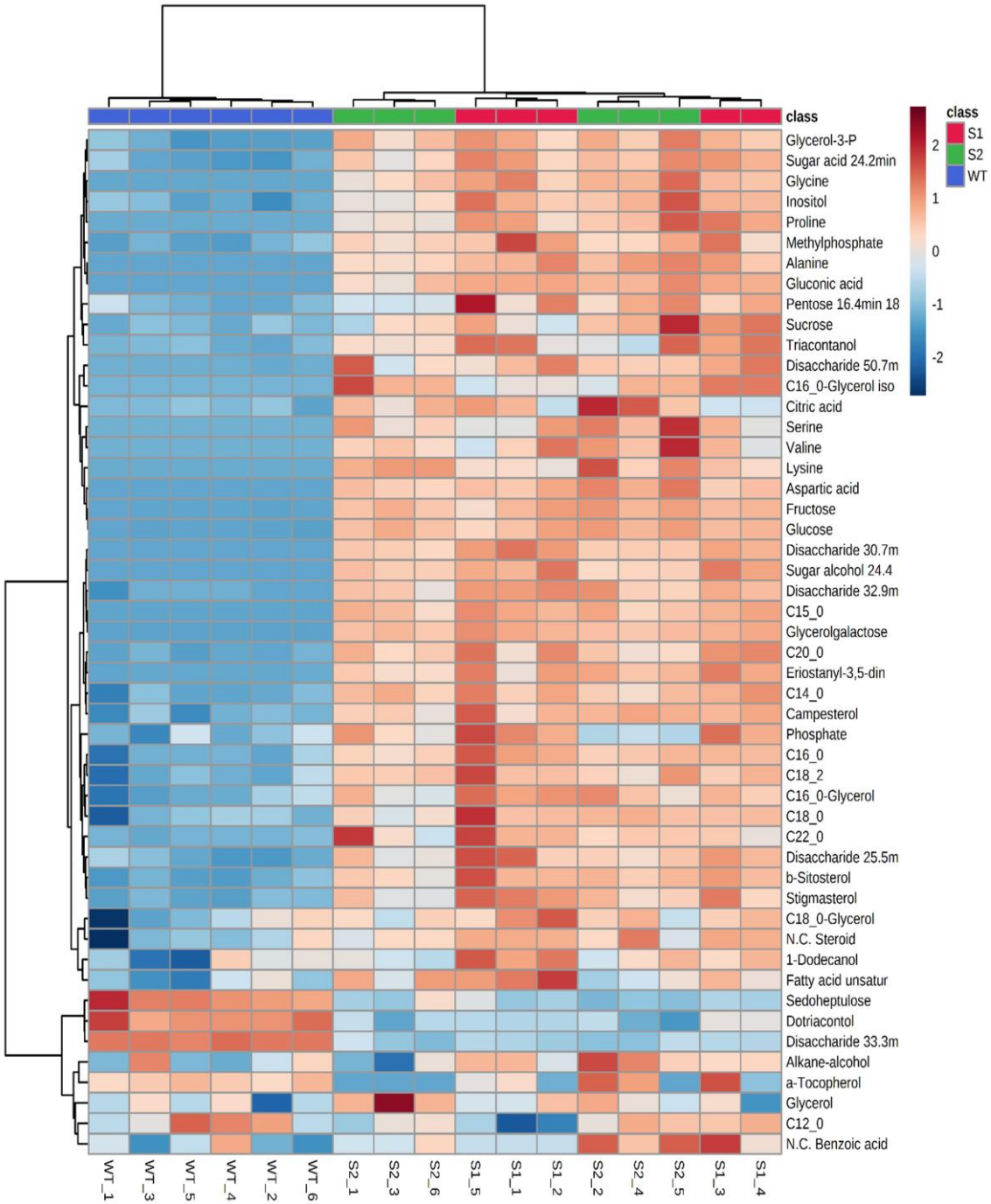


Figure 4.8. Heat map of polar and nonpolar metabolites identified in rice endosperm. S1 and S2 are two biological replicates of line E15 and WT is a wild-type segregant (azygous control). Individual columns represent technical replicates of WT, S1, and S2 (Bottom of the plot). Fold changes are indicated by varying shades of blue (decrease) and red (increase) compared to wild-type seeds. N.C., not confirmed.

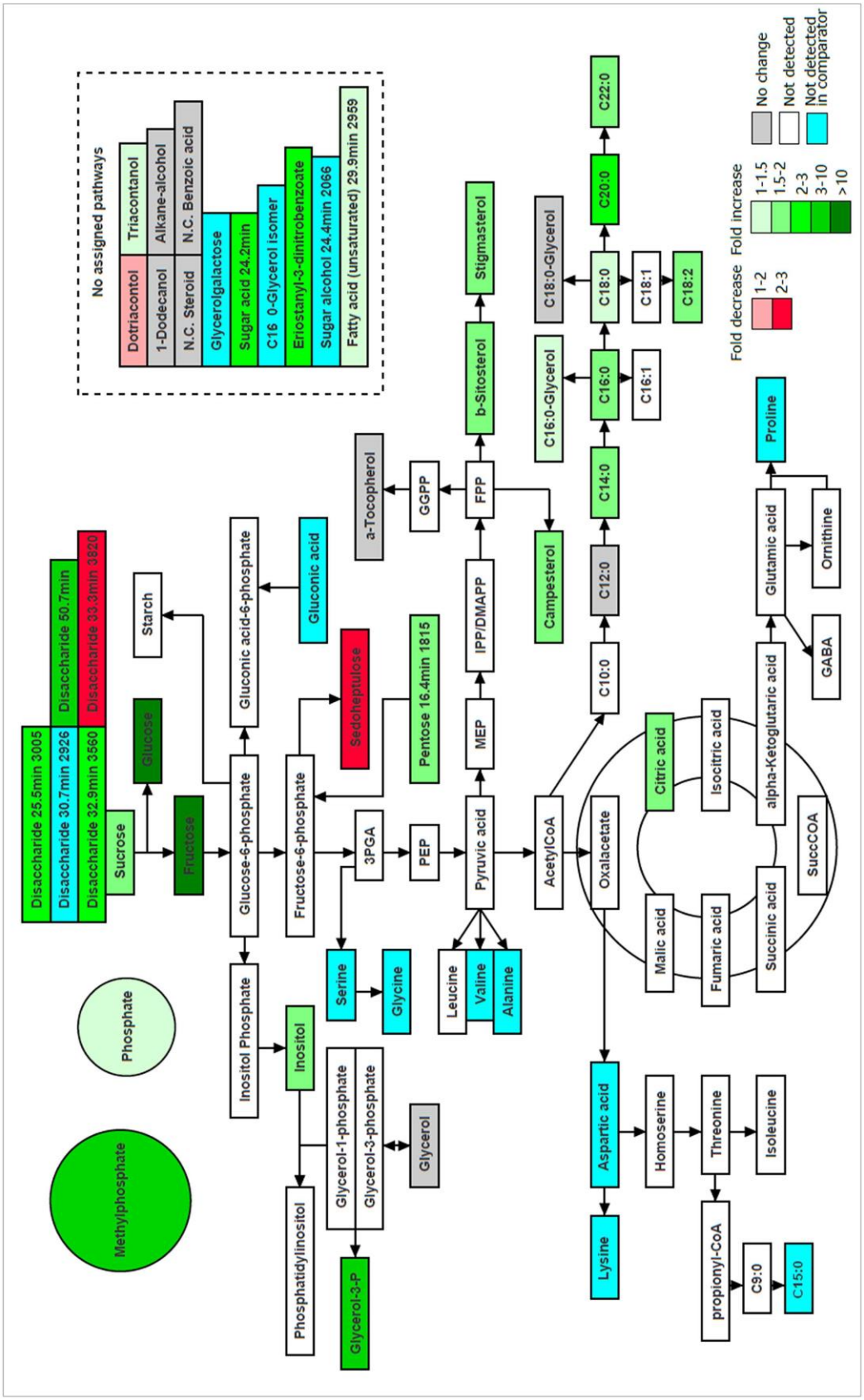


Figure 4.9. Metabolic pathway highlighting significant changes in the endosperm of line E15 compared to wild-type seeds. Significant changes are shown as gradients of red (higher levels in E15), blue (lower levels in E15), and green (detected only in E15). Metabolites with the same abundance in both genotypes are shown in gray and those undetected in either genotype are shown in white. The average of six technical replicates was used for all

Three sugars were present at significantly higher levels in the mutant endosperm compared to the wild-type plants. There was a 2-fold increase in sucrose, a 10-fold increase in glucose, and a 23-fold increase in fructose. Furthermore, glycerol–galactose was not detected in wild-type endosperm, but accumulated to a concentration of 1 mg/g in line E15. Myo-inositol levels were also significantly (1.8-fold) higher in the mutant. Glycerol was the only polyol present at the same concentration in both genotypes, and sedoheptulose was the only sugar that was significantly (3-fold) less abundant in the mutant. Several fatty acids also accumulated to higher levels in the mutant. There was a 1.3-fold increase in stearic acid (C18:0), a >1.5-fold increase in myristic acid (C14:0), palmitic acid (C16:0), linoleic acid (C18:2), and behenic acid (C22:0), and a 2.6-fold increase in arachidic acid (C20:0). Furthermore, pentadecanoic acid (C15:0) was not detected in wild-type endosperm but accumulated to a concentration of >2μg/g in line E15. The only fatty acid present at similar levels in both genotypes was lauric acid (C12:0).

Several amino acids were not detected in the wild-type endosperm (alanine, aspartic acid, glycine, lysine, proline, serine, and valine) but accumulated to detectable levels in the mutant, with alanine, aspartic acid, and proline all accumulating to levels in excess of 100μg/g. Citric acid was also significantly (1.7-fold) more abundant in the mutant, whereas benzoic acid accumulated to similar levels in both genotypes. Gluconic acid was not detected in wild-type endosperm but the concentration was >50 μg/g in the mutant line. The concentrations of campesterol, stigmasterol, and β-sitosterol were significantly (>1.7-fold) higher in the mutant, whereas the levels of two aliphatic alcohols (dodecanol and an alkene alcohol) were similar in both genotypes, and triacontanol was 1.3-fold more abundant in line E15.

The mutant line accumulated significantly higher levels of phosphate (1.4-fold), methyl phosphate (3.8-fold, from the derivatizing agent), and glycerol 3-phosphate (3-fold), whereas the levels of α-tocopherol were similar in both genotypes.

4.5. Discussion

Plants can thrive in a changing environment due to their phenotypic plasticity, which involves the modulation of gene expression to optimize fitness (Matesanz *et al*., 2010). One facet of phenotypic plasticity is adaptive metabolism, in which plants regulate their

metabolic processes in response to stress (Shaar-Moshe *et al*., 2019). Adaptive metabolism also influences the outcome of metabolic engineering because plants adjust their metabolism in response to targeted exogenous interventions such as the introduction of one or more transgenes (Capell *et al*., 2004; Naqvi *et al*., 2008; Zhu *et al*., 2008; Vamvaka *et al*., 2018) or the modification of endogenous genes by targeted (Baysal *et al*., 2016; Pérez *et al*., 2019) or random mutagenesis (Wu *et al*., 2005; Parry *et al*., 2009). Although the regulation of primary and secondary metabolism has been a key focus of plant science for decades, the broader impact of metabolic interventions has only been revealed since the availability of transcriptomics and metabolomics approaches for the high-throughput quantitative analysis of gene expression and small molecules.

Starch metabolism has been studied in detail because starch is the major storage carbohydrate in higher plants and is therefore the main source of carbon for all primary and secondary metabolites in storage tissues (Busi *et al*., 2015). However, starch also accumulates transiently in rapidly proliferating vegetative tissues, such as the cells immediately adjacent to the division zones of meristems (Jordy, 2004). Interest in starch metabolism also reflects its importance in human nutrition and as a feedstock for industrial applications, including the production of biofuels (Zeeman *et al*., 2010). Starch accounts for up to 75% of the mass of cereal grains (Laskowski *et al*., 2019), and rice in particular is a major source of both dietary and industrial starch (Ordonio and Matsuoka, 2016). The metabolic engineering of starch metabolism is partly driven by the health benefits of resistant starch, which has a high content of amylose and is more difficult to digest (Zhu *et al*., 2012). Resistant starch inhibits insulin release and thus lowers the risk of diabetes, obesity, and heart disease (Regina *et al*., 2006, Ordonio and Matsuoka, 2016), and also acts as a prebiotic to support a healthy colon microbiome (Yang *et al*., 2017). Normal rice starch consists of ∼20% amylose and ∼80% amylopectin (Tetlow, 2010; Crofts *et al*., 2015) and the content of resistant starch is typically <2% (Yang *et al*., 2012).

The amylose content of rice starch can be increased by suppressing the activity of OsSBEIIb, the main isoform of SBE responsible for amylopectin synthesis in the endosperm (Nishi *et al*., 2001; Tanaka *et al*., 2004; Sawada *et al*., 2009). I previously reported the mutation of *OsSBEIIb* in japonica cultivar Nipponbare using the CRISPR/Cas9 system to introduce a 4-bp deletion in exon 12, without any off-target effects, generating a truncated protein with no catalytic activity (Baysal *et al*., 2016). Here

I analyzed the T3 seeds of a homozygous mutant line (E15) and found that the amylose content was 27.4%. This was 1.4-fold higher than that of wild-type seeds (19.6%) and Tos17 insertionline NE9005 (19.5%), an additional negative control with the transposon integrated into *OsSBEIIb* intron 18. The E15 seeds also contained a much higher content of resistant starch (17.2%, compared to 0.2% in the wild-type and NE9005 seeds). This came at the expense of a ∼26% fall in the total starch content and significantly lower dry seed and dehulled grain weights. Morphological analysis also revealed that the mutant seeds were opaque and the starch grains were abnormal in size, shape, and distribution.

Similar outcomes have been reported in earlier studies involving the mutation of *OsSBEIIb*. The original ae mutant generated by chemical mutagenesis in japonica cultivar Kinmaze increased the amylose content from 15.7 to 26.5% and the grain weight decreased by 32% (Nishi *et al*., 2001). The knockdown of *OsSBEIIb* by hairpin RNA (RNAi) and microRNA expression in japonica cultivar Nipponbare increased the resistant starch content to 9.5% and the grain weight decreased by 30% (Butardo *et al*., 2011). In this case, the total starch content was shown to increase (albeit by only ∼1%), which together with the lower grain weight resulted in the starch representing up to 90% of the grain. Similarly, the targeted mutation of *OsSBEIIb* exon 3 in cultivar Kinmaze using CRISPR/Cas9 increased the amylose content to 25% and the resistant starch content to 9.8% while the grain weight fell by 30%, and there was a slight but statistically nonsignificant decrease in total starch (Sun *et al*., 2017). The simultaneous knockdown of *OsSBEIIb* and *OsSBEI* by RNAi in a high-amylose indica rice variety increased the amylose content from 27.2 to 64.8% and the resistant starch content from 0 to 14.6% while the grain weight fell by 38% (Zhu *et al*., 2012). Finally, chemical mutagenesis in japonica cultivar Jiangtangdao increased the amylose content from 16.2 to 31.1% and the resistant starch content from 0.4 to 11.7% (Yang *et al*., 2012; 2016). All of these studies also reported opaque rather than translucent seeds and abnormal starch grains, in broad agreement with my results. Furthermore, the inactivation of both homologs of *TaSBEIIa* in a tetraploid durum wheat line increased the amylose content by 22%, whereas the inactivation of single homologs had no effect (Hazard *et al*., 2012). The suppression of *TaSBEIIb* by RNAi in wheat had no effect on the amylose content, whereas the suppression of both *TaSBEIIa* and *TaSBEIIb* increased the amylose content to >70% (35) and similar results were reported for the barley genes *HvSBEIIa* and *HvSBEIIb* (Regina

et al., 2010). Note that in wheat and barley it is the SBEIIa isoform that is predominantly active in the endosperm in contrast to the SBEIIb isoform in rice and maize.

Subtle differences in the amylose and resistant starch content of rice endosperm in the different studies discussed above may reflect the different rice cultivars used in each case, the different suppression strategies (mutation or RNAi), and the different mutation sites in *OsSBEIIb*. Starch branching enzymes comprise three domains (MacGregor *et al*., 2001; Tethlow and Emes, 2014). The highly conserved central domain is responsible for catalytic activity, the N-terminal domain is selective for the size of the α-glucan chain, and the C-terminal domain regulates substrate preference and catalytic activity (20 Mizuno *et al*., 2001; Tethlow and Emes, 2014). Homology modeling and molecular dynamics simulations based on ZmSBEIIb indicated that phosphorylated Ser^{297} forms a salt bridge with Arg^{665} to stabilize the protein (Makhmoudova *et al.,* 2014). The mutation in line E15 introduces a frameshift immediately down-stream of Gly^{418} thus removing Arg^{471} (which plays a key role in the stabilization of the catalytic center) as well as the catalytic triad (Asp^{473}, Glu^{528}, and Asp^{596}). Furthermore, the mutation disrupts salt bridges involving residues Ser^{323} and Arg^{691}. Finally, I observed the strong down-regulation of *OsSBEIIb* gene expression in both leaves and seeds. The mutation completely abolishes OsSBEIIb activity by a combination of removing the catalytic triad, destabilizing the entire catalytic center, abolishing important intramolecular bonds, and feeding back to suppress *OsSBEIIb* gene expression (potentially) by inducing nonsense-mediated decay.

To evaluate the broad impact of the *OsSBEIIb* mutation on starch metabolism, we selected 26 genes representing the starch biosynthesis pathway and evaluated their expression levels in the endosperm and leaves. Previous studies of *OsSBEIIb* mutants and RNAi lines have mainly concentrated on the physicochemical properties of starch and the morphology of the endosperm, but have also considered the potential impact on other aspects of the starch biosynthesis pathway by looking at selected enzyme activities and/or gene expression levels. For example, analysis of the original ae mutant revealed that the loss of OsSBEIIb activity had no effect on the activities of OsSBEI, OsSBEIIa, isoamylase, OsPUL, and OsSSIII, but reduced the activity of OsSSI by 50% and increased the activities of AGPase and sucrose synthase by 25% (Nishi *et al*., 2001). Interestingly, the same study noted that while *OsSBEIIb* gene expression was down-regulated, the *OsSSI* and *OsSSIII* genes were up-regulated, which is broadly consistent with my results.

I also found that *OsSSI*, *OsSSIIIa*, and *OsSSIIIb* were up-regulated in the seeds, although the increase in *OsSSIIIa* expression between wild-type and mutant plants was not statistically significant. Furthermore, the up-regulation of *OsSSI* gene expression did not correlate with the 50% loss of OsSSI activity they detected (Nishi *et al*., 2001). However, a number of studies have shown that gene expression levels do not directly mirror changes in protein abundance or enzymatic activity (Ohdan *et al*., 2005; Geigenberger, 2011).

OsAGPase and OsGBSSI activity increased and OsSSI activity decreased in the triple mutant line generated by crossing the original ae mutant with lines carrying mutations in the *OsSSIIa* and *OsSSIIIa* genes, which Asai *et al*. (2004) proposed was a response to correct for the imbalance between branching and elongation in the mutant. The silencing of *OsSBEIIb* expression by RNAi led to the anticipated decrease in *OsSBEIIb* expression levels but did not affect the expression of *OsSBEI*, *OsSSI*, *OsSSIIa*, or *OsSSIIIa* (Butardo *et al*., 2011) all of which except *OsSSIIIa* were up-regulated in the E15 line. In another RNAi strategy targeting *OsSBEI* and *OsSBEIIb* simultaneously, the target genes were completely suppressed at the mRNA and protein levels, but the abundance of OsGBSSI protein was unaffected. I observed the strong down-regulation of *OsGBSSI* gene expression in the seeds of line E15, although in this case it was a single mutation and we measured mRNA levels, so the two experiments are not comparable.

Although previous studies of *OsSBEIIb* mutants have only considered a small selection of genes, making direct comparisons with this study largely uninformative, a broad panel of starch biosynthesis genes was evaluated in lines generated by mutating *OsGBSSI* (Pérez *et al*., 2019) and *OsAGPL2* (Soto *et al*., 2018). In both cases, the authors noted wide-ranging effects on gene expression in leaves and seeds even though the gene they targeted encoded an endosperm-specific enzyme. Also, in both cases, the response included the ectopic expression of normally compartmentalized enzymes, such as normally leaf-specific *OsGBSSII* expressed in the endosperm and normally endosperm-specific *OsAGPL2* and *OsAGPS2b* expressed in the leaves, generating an ectopic complete AGPase in the leaf cytosol. These phenomena highlight the complex responses that may follow metabolic interventions due to adaptive metabolism (Naqvi *et al.*, 2009). I followed a similar strategy by evaluating 26 genes representing five classes of enzymes. I used seeds harvested at 15 DAF, representing the midpoint of starch accumulation in

developing rice grains (Ohdan *et al*., 2005; Xu *et al*., 2008), and leaf tissue harvested at the same time.

The first set of genes encoded components of AGPase. I found that *OsAGPS2b* and *OsAGPL3* were significantly up-regulated in the mutant seeds (although the latter was expressed at minimal levels) whereas *OsAGPS2a*, *OsAGPL2*, and *OsAGPL3* were significantly up-regulated in the mutant leaves. The other genes were either not expressed at this stage or there was no significant difference between wild-type and mutant plants. In addition to amylopectin, AGPase is also required for amylose synthesis (providing a substrate for SBEs) so there is potential for feedback inhibition if the loss of OsSBEIIb activity causes the accumulation of amylose. However, the response I observed was positive rather than negative, suggesting a more complex regulatory interaction. Complex responses have been reported in earlier studies, where the mutation of one AGPase isoform results in compensatory changes among the others, including unanticipated ectopic expression profiles (Ohdan *et al*., 2005; Soto *et al*., 2018). The second set of genes encoded starch synthases, which elongate α-polyglucan chains. SSI elongates shorter chains than the various isoforms of SSII, SSIII, and SSIV (James *et al.*, 2003). I found that the expression of all detectable soluble SS isoforms was up-regulated in line E15 (*OsSSI* and *OsSSIIa* in the endosperm and *OsSSIIb/c*, *OsSSIIIb*, and *OsSSIVa/b* in the leaves). This may reflect the greater availability of longer α-polyglucan chains, which have been shown to bind SSI with greater affinity than shorter segments (James *et al*., 2003). The only starch synthase gene that was down-regulated in the mutant was *OsGBSSI (Waxy)*, which catalyzes the extension of long glucan chains primarily in amylose (Maddelein *et al*., 1994). The enzyme is restricted to starch grains, and the depletion of this enzyme may explain the abnormal morphology I observed. The third group of genes encoded SBEs. As discussed above, *OsSBEIIb* expression was significantly down-regulated, reflecting a combination of feedback mechanisms and, potentially, nonsense-mediated decay. *OsSBEI* expression in the endosperm was not significantly affected by the mutation, but both *OsSBEI* and *OsSBEIIa* (the main SBE isoform in the leaf) were up-regulated in leaf tissue, suggesting this is a direct compensatory change in response to the mutation. In maize, the lack of *Zm*SBEIIb can be partially complemented by *Zm*SBEIIa and *Zm*SBEI (Liu *et al.*, 2009). The remaining genes encoded DBEs, PUL, PHOL, and PHOH, which degrade starch and/or modify excessively or incorrectly branched chains (Ohdan *et al*., 2005). The main role of SBEIIb

in rice is to introduce new branches into α-polyglucans, so the loss of this enzyme reduces the need for DBEs, PUL, PHOL, and PHOH by removing many of their substrates. This may explain the suppression of the corresponding genes in the leaves of line E15.

The full impact of metabolic interventions can only be determined by metabolomic analysis, and I therefore carried out a broad survey of the endosperm metabolome to identify genotype-specific differences in the content of polar and non-polar metabolites. I anticipated changes in carbohydrate metabolism, which is directly linked to the synthesis of starch, but also evaluated the broader metabolic landscape to look for ripple effects caused by the spillover of intermediates into other pathways and/or the disruption of enzyme complexes. Protein–protein interactions between enzymes in the starch biosynthesis pathway have been demonstrated in rice (Crofts *et al*., 2018), wheat (Tetlow *et al*., 2004; 2008), maize (Hennen-Bierwagen *et al*., 2009; Liu *et al*., 2009; 2012a,b), and barley (Ahmed *et al*., 2015), and the resulting complexes may also include enzymes from linked pathways that regulate carbon partitioning between starch and lipids (Hennen-Bierwagen *et al*., 2009). The starch biosynthesis pathway also differs in source and sink tissues because the carbon is derived from fructose-6-phosphate (Calvin–Benson cycle) in source tissues but from sucrose (transported via the phloem) in sink tissues (Ohdan *et al*., 2005). The sucrose must then be phosphorylated or converted to ADP-glucose in the cytoplasm and translocated into the amyloplasts via the compound-specific hexose monophosphate translocator or the ADP-glucose translocator, respectively (James *et al.*, 2003).

One of the key metabolic effects in line E15 was the greater abundance of soluble sugars, which is a direct result of the ∼26% reduction in starch levels. Similar effects have been reported for other starch biosynthesis mutants in rice (Tang *et al*., 2016; Pérez *et al*., 2019; Soto *et al.*, 2018). The sucrose level in the mutant was twice that in wild-type endosperm, but the amount of glucose and fructose was 10 to 20 times higher.

This may reflect the balance of sucrose hydrolysis vs. amylose synthesis in the endosperm, combined with the strong down-regulation of OsGBSSI, which encodes the only enzyme that transfers glucosyl units from ADP-glucose and utilizes soluble malto-oligosaccharides as a substrate for amylose production (Zeeman *et al*., 2010). Only sedoheptulose was significantly depleted in the mutant, probably reflecting the diversion

of fructose-6-phosphate toward the synthesis of pyruvic acid, acetyl-CoA, and oxaloacetate for the synthesis of additional metabolites, as discussed below (**Figure 4.9**). The disruption of the normal metabolic balance between starch and soluble sugars is also likely to affect the activity of AGPases and the availability of ADP-glucose and pyrophosphate, which in turn may explain the higher inorganic phosphate levels I observed in the E15 endosperm.

Another notable metabolic effect in line E15 was the greater abundance of fatty acids. The dosage effect of ae mutants affects the fatty acid content as well as the starch content of maize seeds (South *et al.*, 1991) and there is strong evidence that amylose and lipid biosynthesis are interdependent (Morrison *et al*., 1993; Villwock *et al*., 1999). This is because both products compete for the same precursors during grain filling, so any deficiency in the starch biosynthesis pathway is likely to provide the lipid pathway with additional substrates (Kang *et al*., 2005). For example, starch deficiency in *Arabidopsis thaliana* resulted in the accumulation of sugars and significantly increased fatty acid synthesis via the modulation of acetyl-CoA carboxylase activity (Yu *et al*., 2018). The down-regulation of AGPase also increased fatty acid accumulation in sugarcane and potato (Zale *et al*., 2016; Xu *et al*., 2019). In the rice floury shrunken endosperm 1 (*fse1*) mutant, the reduction in total starch and amylose levels was complemented by a ∼48% increase in total lipids (Long *et al*., 2018). The ∼26% loss of total starch in line E15 is therefore consistent with the increase in fatty acid levels, although the levels of lauric acid (C12:0) remained stable, perhaps because amylose can form complexes with longer-chain fatty acids (Blazek *et al*., 2011).

I also observed a strong increase in the levels of multiple amino acids and phytosterols, indicating the profound redirection of metabolic flux as a means of carbon partitioning in the endosperm (Ohdan *et al*., 2005). This appears to reflect the greater abundance of sucrose, glucose, and fructose, which feed directly or indirectly into the glycolytic pathway, yielding higher quantities of glucose-6-phosphate (diverted to inositol biosynthesis), 3-phosphoglycerate (diverted to the synthesis of serine and glycine), pyruvate (diverted to the synthesis of valine and alanine, and also into the MEP pathway), acetyl-CoA (utilized for the synthesis of fatty acids), oxaloacetate (utilized for the synthesis of aspartic acid and lysine), and the remaining citric acid cycle organic acids (leading to the synthesis of proline). Environmental factors that affect core metabolism in

rice have previously been shown to modulate the synthesis of amino acids and other nitrogen-rich compounds such as polyamines (Glaubitz *et al*., 2015) and serotonin (Gupta and De, 2017), and the utilization of excess sugars to fuel nitrogen metabolism has been reported in other starch mutants (Kumari *et al*., 2016).

4.6. Conclusions

I found that the mutation of *OsSBEIIb* in line E15 caused the complete inactivation of the enzyme OsSBEIIb and resulted in the accumulation of amylose-rich resistant starch, although the total starch content of the grains fell by ∼26%. This triggered the broad transcriptional reprogramming of starch metabolism, up-regulating the expression of several AGPase subunits and genes encoding soluble SS and SBE, but suppressing the expression of *OsGBSSI* and multiple genes encoding DBEs and disproportionating enzymes. These responses are broadly in line with earlier reports and reflect the adjustment of metabolism to deal with the higher levels of amylose and lower levels of amylopectin. More broadly, the release of sucrose, glucose, and fructose resulted in the profound redirection of flux through core metabolism (glycolysis and the citric acid cycle) boosting the synthesis of multiple fatty acids and amino acids, as well as secondary products of the MEP pathway (phytosterols) and additional products such as gluconic acid, glycerol–galactose, and triacontanol. These results provide insight into the broader implications of perturbing starch metabolism in rice endosperm and its impact on the whole plant, which will make it easier to predict the effect of metabolic engineering in cereals for nutritional improvement or the production of valuable metabolites.

4.7. References

Abe, N., Asai, H., Yago, H., Oitome, N. F., Itoh, R., Crofts, N., Nakamura, Y., & Fujita, N. (2014). Relationships between starch synthase I and branching enzyme isozymes determined using double mutant rice lines. *BMC Plant Biology*, 14, 80.

Ahmed, Z., Tetlow, I. J., Ahmed, R., Morell, M. K., & Emes, M. J. (2015). Protein-protein interactions among enzymes of starch biosynthesis in high-amylose barley genotypes reveal differential roles of heteromeric enzyme complexes in the synthesis of A and B granules. *Plant Science* 233, 95–106.

Asai, H., Abe, N., Matsushima, R., Crofts, N., Oitome, N. F., Nakamura, Y., & Fujita, N. (2014). Deficiencies in both starch synthase IIIa and branching enzyme IIb lead to a significant increase in amylose in SSIIa-inactive japonica rice seeds. *Journal of Experimental Botany*, 65, 5497–5507.

Baysal, C., Bortesi, L., Zhu, C., Farre, G., Schillberg, S., & Christou, P. (2016). CRISPR/Cas9 activity in the rice OsBEIIb gene does not induce off-target effects in the closely related paralog OsBEIIa. *Molecular Breeding*, 36,108.

Blazek, J., Gilbert, E.P., & Copeland, L. (2011). Effects of monoglycerides on pasting properties of wheat starch after repeated heating and cooling in the RVA. *Journal of Cereal Science*, 54, 151–159.

Bläsing, O. E., Gibon, Y., Günther, M., Höhne, M., Morcuende, R., Osuna, D., Thimm, O., Usadel, B., Scheible, W. R., & Stitt, M. (2005). Sugars and circadian regulation make major contributions to the global regulation of diurnal gene expression in Arabidopsis. *The Plant Cell*, 17, 3257–3281.

Busi M. V., Barchiesi J., Martin M., & Gomez-Casati D. F. (2015). Starch metabolism in green algae. *Starch*, 66, 28–40. 10.,

Butardo, V. M., Fitzgerald, M. A., Bird, A. R., Gidley, M. J., Flanagan, B. M., Larroque, O., Resurreccion, A. P., Laidlaw, H. K., Jobling, S. A., Morell, M. K., & Rahman, S. (2011). Impact of down-regulation of starch branching enzyme IIb in rice by artificial microRNA- and hairpin RNA-mediated RNA silencing. *Journal of Experimental Botany*, 62, 4927–4941.

Capell, T., Bassie, L., & Christou, P. (2004). Modulation of the polyamine biosynthetic pathway in transgenic rice confers tolerance to drought stress. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 9909–9914.

Crofts, N., Abe, N., Oitome, N. F., Matsushima, R., Hayashi, M., Tetlow, I. J., Emes, M. J., Nakamura, Y., & Fujita, N. (2015). Amylopectin biosynthetic enzymes from developing rice seed form enzymatically active protein complexes. *Journal of Experimental Botany*, 66, 4469–4482.

Crofts, N., Iizuka, Y., Abe, N., Miura, S., Kikuchi, K., Matsushima, R., & Fujita, N. (2018). Rice Mutants Lacking Starch Synthase I or Branching Enzyme IIb Activity Altered Starch Biosynthetic Protein Complexes. *Frontiers in Plant Science*, 9, 1817.

Decourcelle, M., Perez-Fons, L., Baulande, S., Steiger, S., Couvelard, L., Hem, S., Zhu, C., Capell, T., Christou, P., Fraser, P., & Sandmann, G. (2015). Combined transcript, proteome, and metabolite analysis of transgenic maize seeds engineered for enhanced carotenoid synthesis reveals pleotropic effects in core metabolism. *Journal of Experimental Botany*, 66, 3141–3150.

Drapal, M., Rossel, G., Heider, B., & Fraser, P. D. (2019). Metabolic diversity in sweet potato (*Ipomoea batatas*, Lam.) leaves and storage roots. *Horticulture Research*, 6, 2.

Fraser, P. D., Pinto, M. E., Holloway, D. E., & Bramley, P. M. (2000). Technical advance: application of high-performance liquid chromatography with photodiode array detection to the metabolic profiling of plant isoprenoids. *The Plant Journal*, 24, 551–558.

Hazard, B., Zhang, X., Colasuonno, P., Uauy, C., Beckles, D. M., & Dubcovsky, J. (2012). Induced mutations in the *starch branching enzyme II* (*SBEII*) genes increase amylose and resistant starch content in durum wheat. *Crop Science*, 52, 1754–1766.

Hennen-Bierwagen, T. A., Lin, Q., Grimaud, F., Planchot, V., Keeling, P. L., James, M. G., & Myers, A. M. (2009). Proteins from multiple metabolic pathways associate with starch biosynthetic enzymes in high molecular weight complexes: a model for regulation of carbon allocation in maize amyloplasts. *Plant Physiology*, 149, 1541–1559.

Hong, S., & Preiss, J. (2000). Localization of C-terminal domains required for the maximal activity or for determination of substrate preference of maize branching enzymes. *Archives of Biochemistry and Biophysics*, 378, 349–355.

James, M. G., Denyer, K., & Myers, A. M. (2003). Starch synthesis in the cereal endosperm. *Current Opinion in Plant Biology*, 6, 215–222.

Jeon, J. S., Ryoo, N., Hahn, T. R., Walia, H., & Nakamura, Y. (2010). Starch biosynthesis in cereal endosperm. *Plant Physiology and Biochemistry*, 48, 383–392

Jin, X., Baysal, C., Gao, L., Medina, V., Drapal, M., Ni, X., Sheng, Y., Shi, L., Capell, T., Fraser, P. D., Christou, P., & Zhu, C. (2020). The subcellular localization of two isopentenyl diphosphate isomerases in rice suggests a role for the endoplasmic reticulum in isoprenoid biosynthesis. *Plant Cell Reports*, 39, 119–133.

Jordy M. N. (2004). Seasonal variation of organogenetic activity and reserves allocation in the shoot apex of Pinus pinaster Ait. *Annals of Botany*, 93, 25–37.

Kang, H. G., Park, S., Matsuoka, M., & An, G. (2005). White-core endosperm floury endosperm-4 in rice is generated by knockout mutations in the C-type pyruvate orthophosphate dikinase gene (OsPPDKB). *The Plant Journal*, 42, 901–911.

Kumari, M., & Asthir, B. (2016). Transformation of sucrose to starch and protein in rice leaves and grains under two establishment methods. *Rice Science*, 23, 255–265.

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

Laskowski, W., Górska-Warsewicz, H., Rejman, K., Czeczotko, M., & Zwolińska, J. (2019). How Important are Cereals and Cereal Products in the Average Polish Diet? *Nutrients*, 11, 679.

Liu, F., Makhmoudova, A., Lee, E. A., Wait, R., Emes, M. J., & Tetlow, I. J. (2009). The amylose extender mutant of maize conditions novel protein-protein interactions between starch biosynthetic enzymes in amyloplasts. *Journal of Experimental Botany*, 60, 4423–4440.

Liu, F., Ahmed, Z., Lee, E. A., Donner, E., Liu, Q., Ahmed, R., Morell, M. K., Emes, M. J., & Tetlow, I. J. (2012). Allelic variants of the amylose extender mutation of maize demonstrate phenotypic variation in starch structure resulting from modified protein-protein interactions. *Journal of Experimental Botany*, 63, 1167–1183.

Liu, F., Romanova, N., Lee, E. A., Ahmed, R., Evans, M., Gilbert, E. P., Morell, M. K., Emes, M. J., & Tetlow, I. J. (2012). Glucan affinity of starch synthase IIa determines binding of starch synthase I and starch-branching enzyme IIb to starch granules. *The Biochemical Journal*, 448, 373–387.

Long, W., Wang, Y., Zhu, S., Jing, W., Wang, Y., Ren, Y., Tian, Y., Liu, S., Liu, X., Chen, L., Wang, D., Zhong, M., Zhang, Y., Hu, T., Zhu, J., Hao, Y., Zhu, X., Zhang, W., Wang, C., Zhang, W., … Wan, J. (2018). FLOURY SHRUNKEN ENDOSPERM1 Connects Phospholipid Metabolism and Amyloplast Development in Rice. *Plant Physiology*, 177, 698–712.

MacGregor, E. A., Janecek, S., & Svensson, B. (2001). Relationship of sequence and structure to specificity in the alpha-amylase family of enzymes. *Biochimica et Biophysica Acta*, 1546, 1–20.

Maddelein, M. L., Libessart, N., Bellanger, F., Delrue, B., D'Hulst, C., Van den Koornhuyse, N., Fontaine, T., Wieruszeski, J. M., Decq, A., & Ball, S. (1994). Toward an understanding of the biogenesis of the starch granule. Determination of granule-bound and soluble starch synthase functions in amylopectin synthesis. *The Journal of Biological Chemistry*, 269, 25150–25157.

Makhmoudova, A., Williams, D., Brewer, D., Massey, S., Patterson, J., Silva, A., Vassall, K. A., Liu, F., Subedi, S., Harauz, G., Siu, K. W., Tetlow, I. J., & Emes, M. J. (2014). Identification of multiple phosphorylation sites on maize endosperm starch branching enzyme IIb, a key enzyme in amylopectin biosynthesis. *The Journal of Biological Chemistry*, 289, 9233–9246.

McCleary, B. V., McNally, M., & Rossiter, P. (2002). Measurement of resistant starch by enzymatic digestion in starch and selected plant materials: collaborative study. *Journal of AOAC International*, 85, 1103–1111.

Matesanz, S., Gianoli, E., & Valladares, F. (2010). Global change and the evolution of phenotypic plasticity in plants. *Annals of the New York Academy of Sciences*, 1206, 35–55.

Méchin, V., Thévenot, C., Le Guilloux, M., Prioul, J. L., & Damerval, C. (2007). Developmental analysis of maize endosperm proteome suggests a pivotal role for pyruvate orthophosphate dikinase. *Plant Physiology*, 143, 1203–1219.

Mizuno, K., Kobayashi, E., Tachibana, M., Kawasaki, T., Fujimura, T., Funane, K., Kobayashi, M., & Baba, T. (2001). Characterization of an isoform of rice starch branching enzyme, RBE4, in developing seeds. *Plant & Cell Physiology*, 42, 349–357.

Morrison, W.R., Law, R.V., & Snape, C.E. (1993). Evidence for inclusion complexes of lipids with V-amylose in maize, rices and oat starches. *Journal of Cereal Science*, 18, 107–109.

Nakamura Y. (2002). Towards a better understanding of the metabolic system for amylopectin biosynthesis in plants: rice endosperm as a model tissue. *Plant & Cell Physiology*, 43, 718–725.

Nakamura, Y., Utsumi, Y., Sawada, T., Aihara, S., Utsumi, C., Yoshida, M., & Kitamura, S. (2010). Characterization of the reactions of starch branching enzymes from rice endosperm. *Plant & Cell physiology*, 51, 776–794.

Nakamura, Y., Aihara, S., Crofts, N., Sawada, T., & Fujita, N. (2014). In vitro studies of enzymatic properties of starch synthases and interactions between starch synthase I and starch branching enzymes from rice. *Plant Science*, 224, 1–8.

Nakata, M., Miyashita, T., Kimura, R., Nakata, Y., Takagi, H., Kuroda, M., Yamaguchi, T., Umemoto, T., & Yamakawa, H. (2018). MutMapPlus identified novel mutant alleles of a rice starch branching enzyme IIb gene for fine-tuning of cooked rice texture. *Plant Biotechnology Journal*, 16, 111–123.

Naqvi, S., Zhu, C., Farre, G., Ramessar, K., Bassie, L., Breitenbach, J., Perez Conesa, D., Ros, G., Sandmann, G., Capell, T., & Christou, P. (2009). Transgenic multivitamin corn through biofortification of endosperm with three vitamins representing three distinct metabolic pathways. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 7762–7767.

Nishi, A., Nakamura, Y., Tanaka, N., & Satoh, H. (2001). Biochemical and genetic analysis of the effects of amylose-extender mutation in rice endosperm. *Plant Physiology*, 127, 459–472.

Nogueira, M., Mora, L., Enfissi, E. M., Bramley, P. M., & Fraser, P. D. (2013). Subchromoplast sequestration of carotenoids affects regulatory mechanisms in tomato lines expressing different carotenoid gene combinations. *The Plant Cell*, 25, 4560–4579.

Okonechnikov, K., Golosova, O., Fursov, M., & UGENE team (2012). Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics*, 28, 1166–1167.

Ohdan, T., Francisco, P. B., Jr, Sawada, T., Hirose, T., Terao, T., Satoh, H., & Nakamura, Y. (2005). Expression profiling of genes involved in starch synthesis in sink and source organs of rice. *Journal of Experimental Botany*, 56, 3229–3244.

Ordonio, R. L., & Matsuoka, M. (2016). Increasing resistant starch content in rice for better consumer health. *Proceedings of the National Academy of Sciences of the United States of America*, 113, 12616–12618.

Parry, M. A., Madgwick, P. J., Bayon, C., Tearall, K., Hernandez-Lopez, A., Baudo, M., Rakszegi, M., Hamada, W., Al-Yassin, A., Ouabbou, H., Labhilili, M., & Phillips, A. L. (2009). Mutation discovery for crop improvement. *Journal of Experimental Botany*, 60, 2817–2825.

Pérez, L., Soto, E., Farré, G., Juanos, J., Villorbina, G., Bassie, L., Medina, V., Serrato, A. J., Sahrawy, M., Rojas, J. A., Romagosa, I., Muñoz, P., Zhu, C., & Christou, P. (2019). CRISPR/Cas9 mutations in the rice Waxy/GBSSI gene induce allele-specific and zygosity-dependent feedback effects on endosperm starch biosynthesis. *Plant Cell Reports*, 38, 417–433.

Piasecka, A., Jedrzejczak-Rey, N., & Bednarek, P. (2015). Secondary metabolites in plant innate immunity: conserved function of divergent chemicals. *The New Phytologist*, 206, 948–964.

Qu, J., Xu, S., Zhang, Z., Chen, G., Zhong, Y., Liu, L., Zhang, R., Xue, J., & Guo, D. (2018). Evolutionary, structural and expression analysis of core genes involved in starch synthesis. *Scientific reports*, 8, 12736.

Regina, A., Bird, A., Topping, D., Bowden, S., Freeman, J., Barsby, T., Kosar-Hashemi, B., Li, Z., Rahman, S., & Morell, M. (2006). High-amylose wheat generated by RNA interference improves indices of large-bowel health in rats. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 3546–3551.

Satoh, H., Nishi, A., Fujita, N., Kubo, A., Nakamura, Y., Kawasaki, T., & Okita, T. (2003a). Isolation and characterization of starch mutants in rice. *Journal of Applied Glycoscience*, 50, 225–230.

Satoh, H., Nishi, A., Yamashita, K., Takemoto, Y., Tanaka, Y., Hosaka, Y., Sakurai, A., Fujita, N., & Nakamura, Y. (2003b). Starch-branching enzyme I-deficient mutation specifically affects the structure and properties of starch in rice endosperm. *Plant Physiology*, 133, 1111–1121.

Sawada, T., Francisco, P. B., Jr, Aihara, S., Utsumi, Y., Yoshida, M., Oyama, Y., Tsuzuki, M., Satoh, H., & Nakamura, Y. (2009). Chlorella starch branching enzyme II (BEII) can complement the function of BEIIb in rice endosperm. *Plant & Cell Physiology*, 50, 1062–1074.

Sawada, T., Itoh, M., & Nakamura, Y. (2018). Contributions of Three Starch Branching Enzyme Isozymes to the Fine Structure of Amylopectin in Rice Endosperm. *Frontiers in Plant Science*, 9, 1536..

Schmid, K.M., & Ohlrogge, J.B. (2002) Lipid metabolism in plants. In: D.E. Vance and J.E. Vance (eds), *Biochemistry of lipids, lipoproteins and membra*nes. Elsevier Science. pp 93–126.

Shaar-Moshe, L., Hayouka, R., Roessner, U., & Peleg, Z. (2019). Phenotypic and metabolic plasticity shapes life-history strategies under combinations of abiotic stresses. *Plant Direct*, 3, e00113.

Shaul O. (2015). Unique Aspects of Plant Nonsense-Mediated mRNA Decay. *Trends in Plant Science*, 20, 767–779.

Pérez, L., Soto, E., Villorbina, G., Bassie, L., Medina, V., Muñoz, P., Capell, T., Zhu, C., Christou, P., & Farré, G. (2018). CRISPR/Cas9-induced monoallelic mutations in the cytosolic AGPase large subunit gene APL2 induce the ectopic expression of APL2 and the corresponding small subunit gene APS2b in rice leaves. *Transgenic Research*, 27, 423–439.

South, J., Morrison, W., & Nelson, O. (1991). A relationship between the amylose and lipid contents of starches from various mutants for amylose in maize. *Journal of Cereal Science*, 4, 267–268.

Sun, Y., Jiao, G., Liu, Z., Zhang, X., Li, J., Guo, X., Du, W., Du, J., Francis, F., Zhao, Y., & Xia, L. (2017). Generation of High-Amylose Rice through CRISPR/Cas9-Mediated Targeted Mutagenesis of Starch Branching Enzymes. *Frontiers in Plant Science*, 8, 298.

Tang, X. J., Peng, C., Zhang, J., Cai, Y., You, X. M., Kong, F., Yan, H. G., Wang, G. X., Wang, L., Jin, J., Chen, W. W., Chen, X. G., Ma, J., Wang, P., Jiang, L., Zhang, W. W., & Wan, J. M. (2016). ADP-glucose pyrophosphorylase large subunit 2 is essential for storage substance accumulation and subunit interactions in rice endosperm. *Plant Science*, 249, 70–83.

Tanaka, N., Fujita, N., Nishi, A., Satoh, H., Hosaka, Y., Ugaki, M., Kawasaki, S., & Nakamura, Y. (2004). The structure of starch can be manipulated by changing the expression levels of starch branching enzyme IIb in rice endosperm. *Plant Biotechnology Journal*, 2, 507–516.

Tanaka, Y., Takahashi, K., Kato, J. I., Sawazaki, A., Akasaka, T., Fujita, N., Kumamaru, T., Saito, Y., Shirouchi, B., & Sato, M. (2018). Starch synthase IIIa and starch branching enzyme IIb-deficient mutant rice line ameliorates pancreatic insulin secretion in rats: screening and evaluating mutant rice lines with antidiabetic functionalities. *The British Journal of Nutrition*, 119, 970–980.

Tetlow, I. J., Wait, R., Lu, Z., Akkasaeng, R., Bowsher, C. G., Esposito, S., Kosar-Hashemi, B., Morell, M. K., & Emes, M. J. (2004). Protein phosphorylation in amyloplasts regulates starch branching enzyme activity and protein-protein interactions. *The Plant Cell*, 16, 694–708.

Tetlow, I. J., Beisel, K. G., Cameron, S., Makhmoudova, A., Liu, F., Bresolin, N. S., Wait, R., Morell, M. K., & Emes, M. J. (2008). Analysis of protein complexes in wheat amyloplasts reveals functional interactions among starch biosynthetic enzymes. *Plant Physiology*, 146, 1878–1891.

Tetlow I.J. (2010). Starch biosynthesis in developing seeds. *Seed Science Research*, 21, 5–32.

Tetlow, I. J., & Emes, M. J. (2014). A review of starch-branching enzymes and their role in amylopectin biosynthesis. *IUBMB Life*, 66, 546–558.

Tian, Z., Qian, Q., Liu, Q., Yan, M., Liu, X., Yan, C., Liu, G., Gao, Z., Tang, S., Zeng, D., Wang, Y., Yu, J., Gu, M., & Li, J. (2009). Allelic diversities in rice starch biosynthesis lead to a diverse array of rice eating and cooking qualities. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 21760–21765.

Vamvaka, E., Farré, G., Molinos-Albert, L. M., Evans, A., Canela-Xandri, A., Twyman, R. M., Carrillo, J., Ordóñez, R. A., Shattock, R. J., O'Keefe, B. R., Clotet, B., Blanco, J., Khush, G. S., Christou, P., & Capell, T. (2018). Unexpected synergistic HIV neutralization by a triple microbicide produced in rice endosperm. *Proceedings of the National Academy of Sciences of the United States of America*, 115, E7854–E7862.

Villwock, V.K., Eliasson, A.C., Silverio, J., & BeMiller, J.N. (1999). Starch-lipid interactions in common, waxy, ae du, and ae su2 maize starches examined by differential scanning calorimetry 1. *Cereal Chemistry*, 76, 292–298.

Wang, W., Li, Y., Dang, P., Zhao, S., Lai, D., & Zhou, L. (2018). Rice Secondary Metabolites: Structures, Roles, Biosynthesis, and Metabolic Regulation. *Molecules*, 23, 3098.

Wu, J. L., Wu, C., Lei, C., Baraoidan, M., Bordeos, A., Madamba, M. R., Ramos-Pamplona, M., Mauleon, R., Portugal, A., Ulat, V. J., Bruskiewich, R., Wang, G., Leach, J., Khush, G., & Leung, H. (2005). Chemical- and irradiation-induced mutants of indica rice IR64 for forward and reverse genetics. *Plant Molecular Biology*, 59, 85–97.

Xia, J., & Wishart, D. S. (2016). Using MetaboAnalyst 3.0 for Comprehensive Metabolomics Data Analysis. *Current Protocols in Bioinformatics*, 55, 14.10.1–14.10.91.

Xu, S. B., Li, T., Deng, Z. Y., Chong, K., Xue, Y., & Wang, T. (2008). Dynamic proteomic analysis reveals a switch between central carbon metabolism and alcoholic fermentation in rice filling grains. *Plant physiology*, 148, 908–925.

Xu, X., Vanhercke, T., Shrestha, P., Luo, J., Akbar, S., Konik-Rose, C., Venugoban, L., Hussain, D., Tian, L., Singh, S., Li, Z., Sharp, P. J., & Liu, Q. (2019). Upregulated lipid biosynthesis at the expense of starch production in potato (*Solanum tuberosum*) vegetative tissues *via* simultaneous downregulation of ADP-glucose pyrophosphorylase and Sugar dependent1 expressions. *Frontiers in Plant Science*, 10, 1444.

Yang, R., Sun, C., Bai, J., Luo, Z., Shi, B., Zhang, J., Yan, W., & Piao, Z. (2012). A putative gene sbe3-rs for resistant starch mutated from SBE3 for starch branching enzyme in rice (*Oryza sativa* L.). *PLOS ONE*, 7, e43026.

Yang, R., Bai, J., Fang, J., Wang, Y., Lee, G., & Piao, Z. (2016). A single amino acid mutation of *OsSBEIIb* contributes to resistant starch accumulation in rice. *Breeding Science*, 66, 481–489.

Yang, X., Darko, K. O., Huang, Y., He, C., Yang, H., He, S., Li, J., Li, J., Hocher, B., & Yin, Y. (2017). Resistant Starch Regulates Gut Microbiota: Structure, Biochemistry and Cell Signalling. *Cellular Physiology and Biochemistry*, 42, 306–318.

Yu, L., Fan, J., Yan, C., & Xu, C. (2018). Starch Deficiency Enhances Lipid Biosynthesis and Turnover in Leaves. *Plant Physiology*, 178, 118–129.

Zale, J., Jung, J. H., Kim, J. Y., Pathak, B., Karan, R., Liu, H., Chen, X., Wu, H., Candreva, J., Zhai, Z., Shanklin, J., & Altpeter, F. (2016). Metabolic engineering of sugarcane to accumulate energy-dense triacylglycerols in vegetative biomass. *Plant Biotechnology Journal*, 14, 661–669.

Zeeman, S. C., Kossmann, J., & Smith, A. M. (2010). Starch: its metabolism, evolution, and biotechnological modification in plants. *Annual Review of Plant Biology*, 61, 209–234.

Zhu, L., Gu, M., Meng, X., Cheung, S. C., Yu, H., Huang, J., Sun, Y., Shi, Y., & Liu, Q. (2012). High-amylose rice improves indices of animal health in normal and diabetic rats. *Plant Biotechnology Journal*, 10, 353–362.

Zhu, C., Naqvi, S., Breitenbach, J., Sandmann, G., Christou, P., & Capell, T. (2008). Combinatorial genetic transformation generates a library of metabolic phenotypes for the carotenoid pathway in maize. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 18232–18237.

Chapter 5

General Discussion

5.1. General discussion

Rice is one of the most important sources of nutritional calories, as indicated by the increasing portion of the total arable land occupied by the crop. However, with accelerating loss of productive land to rising sea levels and salinity, rice supply cannot meet with demand (Berman *et al*., 2013). If we are to provide food for the predicted world population of 10 million people by 2050, we need to achieve substantial increases in the yield of rice and other staple food security crops (Albajes *et al*., 2013; Berman *et al*., 2013). In addition, use of nitrogen (N) fertilizer has been increasing continuously worldwide in the last few decades, especially cereal crops, require N for storage proteins in the grain, an important quality attribute. Therefore, it is inavitable to explore strategies to reduce the dependence on nitrogen fertilizers or to improve nutritional quality of cereal crops.

Plants can thrive in a changing environment due to their phenotypic plasticity, which involves the modulation of gene expression to optimize fitness (Matesanz *et al*., 2010). One facet of phenotypic plasticity is adaptive metabolism, in which plants regulate their metabolic processes in response to stress (Shaar-Moshe *et al*., 2019). Adaptive metabolism also influences the outcome of metabolic engineering because plants adjust their metabolism in response to targeted exogenous interventions such as the introduction of one or more heterologous genes (Capell *et al*., 2004; Zhu *et al*., 2008; Naqvi *et al*., 2009; Vamvaka *et al*., 2018) or the modification of endogenous genes by targeted (Pérez *et al*., 2019; Baysal *et al*., 2016) or random mutagenesis (Wu *et al*., 2005; Parry *et al*., 2009).

The integration of nitrogen fixation traits into cereal crops is a long-term goal of agricultural development. To achieve this goal conventional breeding has to be aided by additional techologies, including molecular genetics and biotechnology. Direct transfer of the N_2 fixation genes to the host plant genome to make the plant build its own N_2-fixing machinery without need of bacterial interactions is one of the most important strategies to achieve nitrogen-fixing plants (Curatti *et al*., 2014; Burén *et al*., 2017). Fixation of one molecule of N_2 by nitrogenase requires (at least) eight electrons for its reduction and the hydrolysis of sixteen ATP molecules (Seefeldt, Hoffman and Dean 2009, Burén *et al*., 2018). Thus, nitrogenase can only function in cellular compartments rich in reducing power and energy (Burén *et al*., 2020). Chloroplasts and mitochondria

(two plant organelles of endosymbiont origin) were thought to be candidate compartments for nitrogenase assembly and its function (Beatty and Good 2011; Curatti and Rubio 2014; Burén *et al*., 2018). However, targeting recombinant proteins to subcellular organelles such as mitochondria and plastids, has been a difficult task, especially in species where direct organel genome transformation remains a challenge (Scharff and Bock, 2014, Eseverri *et al*., 2020). Mitochondrial targeting could be useful when the aim is to modulate a mitochondrial function such as energy generation, iron–sulfur cluster assembly, developmental signals, or responses to biotic and abiotic stress (Pierrel *et al*. 2007; Atkin and Macherel, 2009). Furthermore, the low-oxygen environment of mitochondria is ideal for metabolic engineering of oxygen sensitive enzymes such as nitrogense components (Curatti and Rubio, 2014; Burén *et al.*, 2017) and also the control of enzyme metalation, given the abundance of copper, iron, manganese and zinc in the mitochondrial matrix (Pierrel *et al*., 2007; Pérez-González *et al*., 2017).

The extreme O_2 sensitive NifH (dinitrogenase reductase) is the key protein for the engineering of nitrogen fixation in plants. The engineered enzyme needs to be stable and soluble in the targetted organelles (e.g., mitochondria or plastids) and it also has to be compatible with the dinitrogenase NifDK (Jiang *et al.*, 2021). The most known and well studied diazotroph *A. vinelandii* MoFe-protein component NifDK is used in all *in vitro* activity assays. Mitochondria-expressed *S. cerevisia*e NifH (*Sc*NifH^{Ht}) and *N. benthamiana* NifH (*Nb*NifH^{Ht}) supported nitrogenase activity when combined with NifDK^{Av} *in vitro* (Jiang *et al*., 2021). I demonstrated the stable expression and production of the enzymatically active nitrogenase Fe protein (*Os*NifH^{Ht}) in any cereal (in this case rice). Moreover, the isolated NifH carried out the fundamental roles of NifH protein required to engineer nitrogen fixation, including electron transfer NifDK^{Av} and FeMo-co biosynthesis (functional interspecies interactions of rice NifH protein with bacterial NifB, NifEN, and NifDK proteins constituting the core of diazotrophy).

My results confirmed that it should be possible to produce an active nitrogenase enzyme in cereals and enable plants to synthesize their own nitrogen. The plant lines I generated will serve as a model to develop a better understanding of remaining limitations for the production of nitrogenase and its components in plants and point to optimization approaches needed for the production of functional nitrogenase able to support plant

metabolism in the absence of externally supplied nitrogen. Previous reports in plants demonstrated transient expression and accumulation of Nif proteins in different subcellular locatizations. Only the study of Ivleva *et al*. (2016) confirmed the activity of Fe protein stably in tobacco chloroplasts. However, the enzyme was only active when the plants were grown under a low O_2 atmosphere. Even under low O_2 conditions the enzymatic activitiy was just above control levels. The absence and/or availibility of metal clusters in subcellular organelles also has been one of the most important barriers in engineering nitrogen fixation in plants. Recently, Eseverri *et al.* (2020) and Jiang *et al.* (2021) transiently expressed Fe protein in *N. benthamiana* chloroplasts and mitochondria, respectively. Mitochondria-expressed *Nb*NifH^{Ht} was active as isolated. In contrast plastid-targeted *Nb*NifH^{Av} in *N. benthamiana* was not active without co-expressing with NifS^{Av} and NifU^{Av} proteins, suggesting the absence of [Fe-S] clusters in *N. benthamiana* chloroplasts. In the case of mitochondria-targeted *Nb*NifH^{Ht} enzymatic activity was lower compared to reconstituted Fe protein by the addition of [Fe-S] cluster (Jiang *et al*., 2021).

Mitochondria-targeted Fe protein in rice (*Os*NifH^{Ht}) exhibited similar enzymatic activity like mitochondria-targeted Fe protein (*Nb*NifH^{Ht}) in *N. benthamiana.* These results strongly suggest that mitochondria can provide the necessary [Fe-S] clusters, at least to some extend. Importantly, mitochondria appear to be an appropriate subcellular organelle to engineer nitrogen fixation.

The rice plants described in my thesis provide an invaluable resource for further studies in biological nitrogen fixation and the engineering of the nitrogenase complex in plants. They may also serve as a starting point for further engineering of additional components of the nitrogenease machinery as they become available.

In addition to the direct transfer of genes and the expression of the corresponding recombinant proteins to engineer and/or modify metabolic pathways in plants, genome editing technologies such as “CRISPR/Cas” is an alternative tool to investigate and improve the quality of important agronomic traits such as resistance to herbicides or more complex traits related to grain yield and nitrogen use efficieny (Sun *et al*., 2016; Xu *et al*., 2016; Li *et al*., 2018). One of the pre-requisites for effective genome editing is the ability to introduce DSBs at unique and predefined sites in the plant genome (Baysal *et al*., 2016). I found that CRISPR/Cas9 mutation of *OsSBEIIb* in rice caused the complete

inactivation of the corresponding enzyme OsSBEIIb. In turn, this resulted in the accumulation of amylose-rich resistant starch and triggered the broad transcriptional reprogramming of starch metabolism. Starch metabolism has been studied in detail because starch is the major storage carbohydrate in higher plants and is therefore the main source of carbon for all primary and secondary metabolites in storage tissues (Busi *et al*., 2015). Interest in starch metabolism also reflects its importance in human nutrition and as a feedstock for industrial applications (Zeeman *et al*., 2010). Starch accounts for up to 75% of the mass of cereal grains (Laskowski *et al*., 2019), and rice in particular is a major source of both dietary and industrial starch (Ordonio *et al*., 2016). The metabolic engineering of starch metabolism is partly driven by the health benefits of resistant starch, which has a high content of amylose and is more difficult to digest (Zhu *et al*., 2012). Resistant starch inhibits insulin release and thus lowers the risk of diabetes, obesity, and heart disease (Ordonio *et al*., 2016, Regina *et al*., 2006), and also acts as a prebiotic to support a healthy colon microbiome (Yang *et al.,* 2017). Normal rice starch, consists of ~20% amylose and ~80% amylopectin (Tetlow, 2010; Crofts *et al.*, 2015) and the content of resistant starch is typically <2% (Yang *et al*., 2012). The CRISPR/Cas9 generated homozygous mutant rice plants, in which OsSBEIIb was completely inactivated by abolishing the catalytic center and C-terminal regulatory domain, produced opaque seeds with depleted starch reserves. Amylose content in the mutant increased from 19.6 to 27.4% and resistant starch (RS) content increased from 0.2 to 17.2%. In addition, the first time I demonstrated the effects of this specific mutation on the expression of entire starch biosynthesis genes in the endosperm and leaf. Furthermore, the mutation also had a wide effect on general primary and secondary metabolism in the endosperm, causing the accumulation of sugars, fatty acids, amino acids, and phytosterols which will make it easier to predict the effect of metabolic engineering in cereals for nutritional improvement or the production of valuable metabolites.

5.2. References

Albajes, R., Cantero-Martínez, C., Capell, T., Christou, P., Farre, A., Galceran, J., López-Gatius, F., Marin, S., Martín-Belloso, O., Motilva, Ma-J., Nogareda, C., Peman, J., Puy, J., Recasens, J., Romagosa, I., Romero, Ma-P., Sanchis, V., Savin, R., Slafer, G.A. Soliva-Fortuny, R., Viñas, I. & Voltas, J. (2013). Building bridges: an integrated strategy for sustainable food production throughout the value chain. *Molecular Breeding* 32, 743–770.

Allen, R. S., Tilbrook, K., Warden, A. C., Campbell, P. C., Rolland, V., Singh, S. P., & Wood, C. C. (2017). Expression of 16 Nitrogenase Proteins within the Plant Mitochondrial Matrix. *Frontiers in Plant Science*, 8, 287.

Baysal, C., Bortesi, L., Zhu, C., Farre, G., Schillberg, S., & Christou, P. (2016). CRISPR/Cas9 activity in the rice OsBEIIb gene does not induce off-target effects in the closely related paralog OsBEIIa. *Molecular Breeding*, 36,108.

Berman, J., Zhu, C., Pérez-Massot, E., Arjó, G., Zorrilla-López, U., Masip, G., Banakar, R., Sanahuja, G., Farré, G., Miralpeix, B., Bai, C., Vamvaka, E., Sabalza, M., Twyman, R. M., Bassié, L., Capell, T., & Christou, P. (2013). Can the world afford to ignore biotechnology solutions that address food insecurity? *Plant Molecular Biology*, 83, 5–19.

Busi M. V., Barchiesi J., Martin M., & Gomez-Casati D. F. (2015). Starch metabolism in green algae. *Starch*, 66, 28–40. 10.,

Capell, T., Bassie, L., & Christou, P. (2004). Modulation of the polyamine biosynthetic pathway in transgenic rice confers tolerance to drought stress. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 9909–9914.

Crofts, N., Abe, N., Oitome, N. F., Matsushima, R., Hayashi, M., Tetlow, I. J., Emes, M. J., Nakamura, Y., & Fujita, N. (2015). Amylopectin biosynthetic enzymes from developing rice seed form enzymatically active protein complexes. *Journal of Experimental Botany*, 66, 4469–4482.

Dolezal, P., Likic, V., Tachezy, J., & Lithgow, T. (2006). Evolution of the molecular machines for protein import into mitochondria. *Science*, 313, 314–318.

Eseverri, Á., López-Torrejón, G., Jiang, X., Burén, S., Rubio, L. M., & Caro, E. (2020). Use of synthetic biology tools to optimize the production of active nitrogenase Fe protein in chloroplasts of tobacco leaf cells. *Plant Biotechnology Journal*, 18, 1882–1896.

Ivleva, N. B., Groat, J., Staub, J. M., & Stephens, M. (2016). Expression of Active Subunit of Nitrogenase via Integration into Plant Organelle Genome. *PLOS ONE*, 11, e0160951.

Jiang, X., Payá-Tormo, L., Coroian, D., García-Rubio, I., Castellanos-Rueda, R., Eseverri, Á., López-Torrejón, G., Burén, S., & Rubio, L. M. (2021). Exploiting genetic diversity and gene synthesis to identify superior nitrogenase NifH protein variants to engineer N2-fixation in plants. *Communications Biology*, 4, 4.

Jordy M. N. (2004). Seasonal variation of organogenetic activity and reserves allocation in the shoot apex of Pinus pinaster Ait. *Annals of Botany*, 93, 25–37.

Laskowski, W., Górska-Warsewicz, H., Rejman, K., Czeczotko, M., & Zwolińska, J. (2019). How Important are Cereals and Cereal Products in the Average Polish Diet? *Nutrients*, 11, 679.

López-Torrejón, G., Jiménez-Vicente, E., Buesa, J. M., Hernandez, J. A., Verma, H. K., & Rubio, L. M. (2016). Expression of a functional oxygen-labile nitrogenase component in the mitochondrial matrix of aerobically grown yeast. *Nature Communications*, 7, 11426.

Matesanz, S., Gianoli, E., & Valladares, F. (2010). Global change and the evolution of phenotypic plasticity in plants. *Annals of the New York Academy of Sciences*, 1206, 35–55.

Naqvi, S., Zhu, C., Farre, G., Ramessar, K., Bassie, L., Breitenbach, J., Perez Conesa, D., Ros, G., Sandmann, G., Capell, T., & Christou, P. (2009). Transgenic multivitamin corn through biofortification of endosperm with three vitamins representing three distinct metabolic pathways. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 7762–7767.

Ordonio, R. L., & Matsuoka, M. (2016). Increasing resistant starch content in rice for better consumer health. *Proceedings of the National Academy of Sciences of the United States of America*, 113, 12616–12618.

Parry, M. A., Madgwick, P. J., Bayon, C., Tearall, K., Hernandez-Lopez, A., Baudo, M., Rakszegi, M., Hamada, W., Al-Yassin, A., Ouabbou, H., Labhilili, M., & Phillips, A. L. (2009). Mutation discovery for crop improvement. *Journal of Experimental Botany*, 60, 2817–2825.

Pérez, L., Soto, E., Farre, G., Juanos, J., Villorbina, G., Bassie, L., & Christou, P. (2019). CRISPR/Cas9 mutations in the rice Waxy/GBSSI gene induce allele-specific and zygosity-dependent feedback effects on endosperm starch biosynthesis. *Plant Cell Reports*, 38, 417-433.

Pfanner, N., & Geissler, A. (2001). Versatility of the mitochondrial protein import machinery. *Nature Reviews. Molecular Cell Biology*, 2, 339–349.

Regina, A., Bird, A., Topping, D., Bowden, S., Freeman, J., Barsby, T., Kosar-Hashemi, B., Li, Z., Rahman, S., & Morell, M. (2006). High-amylose wheat generated by RNA interference improves indices of large-bowel health in rats. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 3546–3551.

Shaar-Moshe, L., Hayouka, R., Roessner, U., & Peleg, Z. (2019). Phenotypic and metabolic plasticity shapes life-history strategies under combinations of abiotic stresses. *Plant Direct*, 3, e00113.

Taylor, R. D., & Pfanner, N. (2004). The protein import and assembly machinery of the mitochondrial outer membrane. *Biochimica et Biophysica Acta*, 1658, 37–43.

Tetlow I.J. (2010). Starch biosynthesis in developing seeds. *Seed Science Research*, 21, 5–32.

Wu, J. L., Wu, C., Lei, C., Baraoidan, M., Bordeos, A., Madamba, M. R., Ramos-Pamplona, M., Mauleon, R., Portugal, A., Ulat, V. J., Bruskiewich, R., Wang, G., Leach, J., Khush, G., & Leung, H. (2005). Chemical- and irradiation-induced mutants of indica rice IR64 for forward and reverse genetics. *Plant Molecular Biology*, 59, 85–97.

Yang, R., Sun, C., Bai, J., Luo, Z., Shi, B., Zhang, J., Yan, W., & Piao, Z. (2012). A putative gene sbe3-rs for resistant starch mutated from SBE3 for starch branching enzyme in rice (*Oryza sativa* L.). *PLOS ONE*, 7, e43026.

Zeeman, S. C., Kossmann, J., & Smith, A. M. (2010). Starch: its metabolism, evolution, and biotechnological modification in plants. *Annual Review of Plant Biology*, 61, 209–234.

Zhu, L., Gu, M., Meng, X., Cheung, S. C., Yu, H., Huang, J., Sun, Y., Shi, Y., & Liu, Q. (2012). High-amylose rice improves indices of animal health in normal and diabetic rats. *Plant Biotechnology Journal*, 10, 353–362.

Chapter 6

General Conclusions

GENERAL CONCLUSIONS

- Particular functional motifs in targeting peptides are more important than phylogenetic origin for the effective import of recombinant proteins into rice mitochondria.
- Mitochondrial import of nuclear-encoded eGFP in rice requires only a single N-terminal hexamer motif whereas the Tom20-binding motifs are not strictly required.
- The efficiency of mitochondrial import is difficult to quantify but can be determined by combining stable gene expression with subcellular localization analysis by immunogold labeling.
- Rice mitochondria are capable of accumulating active Fe protein (NifH) which is one of the two structural components of nitrogenase, despite its extreme O_2 sensitivity.
- The endogenous mitochondrial Fe–S cluster assembly machinery is able to provide metal clusters to NifH and the activity of NifH as isolated enzyme confirmed its ability to incorporate endogenous rice mitochondrial [Fe-S] clusters.
- Reconstitution of NifH by the *in vitro* transfer of [4Fe-4S] from NifU increased NifH activity nine-fold compared to the as-isolated NifH protein. This is the first demonstration of the stable expression and activity of a nitrogenase component in any cereal.
- NifH carried out the fundamental roles of Fe protein required to engineer nitrogen fixation, including electron transfer to NifDK for N_2 reduction and FeMo-co biosynthesis.
- FeMo-co biosynthesis, confirmed that rice Fe protein also formed functional interspecies interactions with NifB, NifEN, and NifDK proteins constituting the core of diazotrophy.
- CRISPR/Cas9 mutation of *OsSBEIIb* in rice caused the complete inactivation of the corresponding enzyme, resulted in the accumulation of amylose-rich resistant starch and triggered broad transcriptional reprogramming of starch metabolism.
- The alteration of rice carbohydrate metabolism in the *OsSBEIIb* mutant resulted in the profound redirection of flux through core metabolism boosting the synthesis of fatty acids and amino acids, as well as secondary products of the MEP pathway (phytosterols) and additional products such as gluconic acid, glycerol–galactose, and triacontanol.

FINAL CONSIDERATIONS

This thesis provides enabling technology and tangible materials and resources instrumental to meet the grand challenge of engineering N_2-fixing cereals by overcoming the bottlenecks for the production of active nitrogenase components. The impact of knocking out a key starch biosynthetic enzyme (OsSBEIIb) on the broader starch metabolism in rice endosperm and its consequences on general primary and secondary metabolism at the whole plant level will contribute towards predictive metabolic engineering in plants. This in turn will make it easier to develop and implement more refined and targeted startegies for nutritional improvement and the production of valuable metabolites in plants.