Universitat de Lleida

Metabolic engineering and genome editing in rice

Lucía Pérez Álvarez

Universitat de Lleida

TESI DOCTORAL

Metabolic engineering and genome editing in rice

Lucía Pérez Álvarez

Memòria presentada per optar al grau de Doctor per la Universitat de Lleida
Programa de Doctorat en Ciencia y Tecnología Agraria i Alimentaria

Director/a
Paul Christou
Gemma Villorbina

Tutor/a
Paul Christou

December 2018

Table of Contents

INDEX OF FIGURES

Chapter 3

INDEX OF TABLES

ACKNOWLEDGEMENTS

First of all, I would like to express my deep sense of gratitude to my supervisor Paul Christou, for his patient guidance, encouragement and excellent advice throughout my PhD thesis. I would like to thank my co-supervisor Gemma Villorbina, who with his concise helpfull feedback and in depth knowledge was invaluable to the project, and Teresa Capell, who was always ready to help, for her constant feedback and her energetic personality and enthusiasm on science.

I would like to offer my special thanks to Ludovic Bassie for his useful discussions, for his constant help and his kind advice. To the professors, that with their contribution enriched my work professionally, Vicente Medina and Pilar Muñoz.

My PhD was immeasurable aided by the very friendly atmosphere and continuous technical and emotional support during the last three years. It is a pleasure to convey my gratitude to all people who helped me in achieving this important task. I wish you lot of luck and success in your work and personal life.

I would like to thank to my very good friends Gemma Farré and Erika Soto for supporting me in each step of this stage in my life, without you I could not finish my PhD. Thank you for your help in my experiments, for the knowledge in procedures and for your deep emotional support. I would also like to thank Xin Jin for her friendship in the lab and outside. I would like to thank Laura Perez for helping me in the metabolomic analysis.

To all my colleagues and degree ex-students: Marco Castellani, Andrew Montecillo, Raviraj Banakar, Can Baysal, Marlies Hoefkens, Derry Álvarez , Victoria Armario, Amaya Blanco, Ines Samperi, Roser Salvia, Julia Juanos, Nuria Farrus and Julia Saez for the moments we shared in the lab helped me to become the scientific I am right now.

To Riad Nadi, Giobbe Forni, Gemma Masip, Daniela Zanga, Eduard Molinero and Mohamed Herma, we shared many fun moments during our PhDs/MSc together and in the future we will have more lovely time together. I value our friendship very much.

I am thankful to Núria Gabernet for dealing with administrative work and for always being kind and helpful. I would like to thanks Jaume Capell for his practical training in growing plants and sharing a great time in the greenhouse.

It has been a great privilege to work with all of you. Thank you for being patient and always trying my experimental cakes.

My PhD thesis could never be fulfilled without the full support from my family and friends. I am deeply and forever indebted to my parents for their love, support and encouragement throughout my entire life. Finally, I would like to express my special thanks to Gerard Utrilla for his moral support and generous love.

SUMMARY

My research program used rice as an experimental model to address fundamental bottlenecks and mechanisms limiting the transition from metabolic engineering to synthetic biology in plants. I concentrated on an in depth molecular and biochemical characterization of plants I generated in two distinct, yet interrelated sets of research lines.

In the first set of experiments I addressed the hypothesis that by knocking out specific genes in a primary metabolic pathway, starch biosynthesis, resulting mutant plants might exhibit phenotypes conducive to synthetic biology applications in terms of redirecting flux and limiting biosynthetic precursors to specific secondary metabolic pathways. In this context I used CRISPR/Cas9 to create two heterozygous mutants, one with a severely truncated and non-functional cytosolic glucose-1-phosphate adenylyl transferase (AGPase) and the other with a C-terminal structural modification causing a partial loss of activity.

Unexpectedly, both mutants exhibited depletion of starch in the leaves and a corresponding increase in the level of soluble sugars. This reflected the unanticipated expression of both OsAPL2 and OsAPS2b in the leaves, generating a complete ectopic AGPase in the leaf cytosol, and a corresponding decrease in the expression of the plastidial small subunit OsAPS2a that was only partially complemented by an increase in the expression of OsAPS1.

In a subsequent set of experiments along similar lines, I investigated the broader effects of mutations in an additional starch biosynthetic gene, granule bound starch synthase (GBBS, waxy). I used CRISPR/Cas9 to introduce a range of mutations with different effects in this specific gene.

All mutations I recovered reduced but did not abolish GBSS activity in seeds due to partial compensation caused by the ectopic upregulation of GBSSII. The GBSS activity in the mutants was 61–71% of wild-type levels, but the amylose content nevertheless declined to 8–12% in heterozygous seeds and to as low as 5% in homozygous seeds, accompanied by abnormal cellular organization in the aleurone layer and amorphous starch grain structures. Almost every starch pathway gene was impacted at different degrees in leaves and seeds. These gene expression changes resulted in changes in AGPase and sucrose synthase activity that explained the corresponding levels of starch and soluble sugars.

The second line of my program focused on the engineering of an ectopic MVA pathway in rice plastids in order to investigate the hypothesis that by reconstituting such an ectopic pathway the strict regulation of the native MVA pathway might be relieved to a certain degree in turns increasing the pool of essential terpenoid precursors. Results indicated a profound increase in the levels of fatty acids, lutein and tocopherol, a decrease in squalene levels and similar levels of sterols.

My results set the stage for further experiments to ascertain whether germplasm I created and characterized, can serve as a basis for more complex metabolic engineering and synthetic biology interventions.

RESUMEN

Mi programa de investigación se ha basado en la utilización del arroz como modelo experimental para estudiar mecanismos y cuellos de botella que limitan la transición de la ingeniería metabólica a la biología sintética en las plantas. Me concentré en la caracterización a nivel molecular y bioquímica de plantas que generé en dos conjuntos de líneas de investigación distintas pero interrelacionadas. En el primer conjunto de experimentos abordé la hipótesis de que al eliminar genes específicos en una ruta metabólica primaria, concretamente la biosíntesis de almidón, las plantas mutantes resultantes podrían mostrar fenotipos propicios para aplicaciones de biología sintética en términos de redireccionar el flujo y limitar los precursores biosintéticos a vías metabólicas secundarias específicas. En este contexto, utilicé CRISPR/Cas9 para crear dos mutantes heterocigotos, uno con glucosa-1-fosfato adenil transferasa (AGPasa) citosólica severamente truncada y no funcional y el otro con una modificación estructural en el C-terminal causando una pérdida parcial de actividad. Inesperadamente, observamos una reducción del almidón en las hojas de ambos mutantes y un aumento concomitante en el nivel de azúcares solubles. Esto reflejó la expresión no prevista de OsAPL2 y OsAPS2b en las hojas, generando una AGPasa ectópica completa en el citosol de la hoja, y una disminución en la expresión de la subunidad pequeña plastidial OsAPS2a que se complementó solo parcialmente con un aumento en la expresión de OsAPS1

En un conjunto posterior de experimentos, con similar base, investigué los efectos más amplios de las mutaciones en un gen de la biosíntesis de almidón, la almidón sintasa unida a gránulos (GBBS, waxy). Utilicé CRISPR/Cas9 para introducir un rango de mutaciones con diferentes efectos en este gen específico. Encontré que las mutaciones producidas redujeron, pero no abolieron la actividad de GBSS en las semillas, debido a una compensación parcial causada por la regulación ectópica de GBSSII. La actividad de GBSS en los mutantes fue de 61 a 71% de los niveles de los controles, pero el contenido de amilosa, sin embargo, disminuyó a 8 a 12% en semillas heterocigotas y fue tan bajo como 5% en semillas homocigotas, acompañado por una organización celular anormal en la capa de aleurona y con estructuras del grano de almidón amorfas. Casi todos los genes de la vía del almidón se vieron afectados a diferentes niveles en las hojas y semillas. Estos cambios en la expresión génica dieron como resultado cambios en la actividad de la AGPasa y de la sucrosa sintasa que coincidían con las alteraciones en los niveles de almidón y azúcares solubles.

La segunda línea de mi programa de investigación se centró en la ingeniería de una vía ectópica de MVA en plastidios de arroz para investigar la hipótesis de que al reconstituir la vía ectópica, la regulación estricta de la vía de MVA nativa podría relajarse en cierto grado a medida que aumentase el conjunto de precursores de terpenoides esenciales. Los resultados indicaron un aumento en los niveles de ácidos grasos, luteína y tocoferol, una reducción en los niveles de escualeno y niveles similares de esteroles.

Mis resultados son el fundamento para futuros experimentos encaminados a determinar si el germoplasma que he creado y caracterizado puede servir de base para intervenciones de ingeniería metabólica y biología sintética más complejas.

RESUM

El meu programa d'investigació s'ha basat en la utilització de l'arròs com a model experimental per a estudiar mecanismes i colls d'ampolla que limiten la transició de l'enginyeria metabòlica a la biologia sintètica en les plantes. Em vaig concentrar en la caracterització a nivell molecular i bioquímica de plantes que vaig generar en dos conjunts de línies d'investigació diferents però interrelacionades. En el primer conjunt d'experiments vaig abordar la hipòtesi que en eliminar gens específics en una ruta metabòlica primària, concretament la biosíntesi de midó, les plantes mutants resultants podrien mostrar fenotips propicis per a aplicacions de biologia sintètica en termes de redireccionar el flux i limitar els precursors biosintètics a vies metabòliques secundàries específiques. En aquest context, vaig utilitzar CRISPR/Cas9 per crear dos mutants heterozigots, un amb una glucosa-1-fosfat adenil transferasa (AGPasa) citosòlica severament truncada i no funcional i l'altre amb una modificació estructural C-terminal causant una pèrdua parcial d'activitat. Inesperadament, vaig observar una reducció del midó en les fulles de tots dos mutants i un augment concomitant en el nivell de sucres solubles. Això es va reflectir en l'expressió no prevista d'OsAPL2 i OsAPS2b en les fulles, generant una AGPasa ectòpica completa en el citosol de la fulla, i una disminució en l'expressió de la subunitat petita plastidial OsAPS2a que es va complementar només parcialment amb un augment en l'expressió de OsAPS1

En un conjunt posterior d'experiments amb base similar, vaig investigar els efectes més amplis de les mutacions en un gen de la biosíntesi de midó, la midó sintasa unida a grànuls (GBBS, waxy). Vaig utilitzar CRISPR / Cas9 per introduir un rang de mutacions amb diferents efectes en aquest gen específic. Vaig trobar que les mutacions produïdes van reduir, però no van abolir totalment, l'activitat de GBSS en les llavors a causa d'una compensació parcial causada per la regulació ectòpica de GBSSII. L'activitat de GBSS en els mutants va ser de 61 a 71% dels nivells dels controls, però el contingut d'amilosa, va disminuir del 8 al 12% en llavors heterozigotes i va ser tan baix com nomes un 5% en llavors homozigotes, acompanyat per una organització cel·lular anormal en la capa d'aleurona i amb estructures del gra de midó amorfes. Gairebé tots els gens de la via del midó es van veure afectats a diferents nivells en les fulles i llavors. Aquests canvis en l'expressió gènica van donar com a resultat canvis en l'activitat de la AGPasa i de la sacarosa sintasa que coincidien amb els nivells corresponents de midó i sucres solubles.

La segona línia del meu programa d'investigació es va centrar en l'enginyeria d'una via ectòpica de MVA en plastidis d'arròs per investigar la hipòtesi que al reconstituir la via ectòpica, la regulació estricta de la via de MVA nativa podria relaxar-se en cert grau a mesura que augmentés el conjunt de precursors terpenoides essencials. Els resultats van indicar un augment molt elevat dels nivells d'àcids grassos, luteïna i tocoferol, i una reducció en els nivells d'esqualè i d'esterols.

Els meus resultats són el fonament per a futurs experiments encaminats a determinar si el germoplasma que he creat i caracteritzat pot servir de base per a intervencions d'enginyeria metabòlica i biologia sintètica més complexes.

LIST OF ABREVIATIONS

3-PGA: 3-phosphoglycerate

AACT: Acyl-coenzyme A-cholesterol acyltransferase

ADP: adenosine diphosphate

AGPase: ADP-glucose pyrophosphorylase

AGPlar: Glucose-1-phosphate adenylyltransferase

APL1: AGPase large subunit 1

APL2: AGPase large subunit 2

APL3: AGPase large subunit 3

APL4: AGPase large subunit 4

APS1: AGPase small subunit 1

APS2: AGPase small subunit 2

APS2a: AGPase small subunit 2a

APS2b: AGPase small subunit 2b

Arg: Arginine

ATP: adenosine triphosphate

BEI: Branching enzyme 1

BEIIa: Branching enzyme 2a

BEIIb: Branching enzyme 2b

C3 intermediates:

Cas9: CRISPR associated protein 9

Cas9D10A: Cas9 mutant that generate single strain break

Cas9WT: CRISPR associated protein 9 that generate double strain break

cDNA: complementary DNA

CDP-ME: 4-(cytidine-5'-diphospho)-2-C-methyl-d-Erythritol

CDP-MEP: 4-Diphosphocytidyl-2C-methyl-d-erythritol 4-phosphate

CMK: 4-(cytidine-5'-diphospho)-2-C-methyl-d-Erythritol kinase

CMS: 4-Diphosphocytidyl-2C-methyl-d-erythritol 4-phosphate synthase

CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats

C-terminal: Carboxi-terminal

DBE: Debranching enzymes

DMAPP: Dimethylallyl diphosphate

DNA: deoxyribonucleic acid

DPE1: 4-alpha-glucanotransferase 1 or disproportionating enzyme 1

DPE2: 4-alpha-glucanotransferase 2 or disproportionating enzyme 2

DSB: double strand break

DXR: 1-deoxy-D-xylulose-5-phosphate reductoisomerase

DXS: 1-deoxy-D-xylulose-5-phosphate synthase

F-6-P: Fructose-6-phosphate

fosm: fosmidomicin

FPP: Farnesyl diphosphate

FPS: farnesyl-diphosphate synthase

FW: Fresh weight

G-1-P: Glucose-1-phosphate

G3P: glyceraldehyde 3 phosphate

GBSS: granule-bound starch synthase

GBSSI: granule-bound starch synthase 1

GBSSII: granule-bound starch synthase 2

GGPP: Geranylgeranyl diphosphate

GGPS: Glucosylglycerol-phosphate synthase

Glc-6-P: Glucose-6-phosphate

Gln: Glutamine

Glu: Glutamic acid

GPS: Geranyl diphosphate synthase

HDS: 4-Hydroxy-3-methylbut2-en-yl-diphosphate synthase

HEPES: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HMBPP: (E)-4-Hydroxy-3-methyl-but-2-enyl pyrophosphate

HMG-CoA: 3-hydroxy-3-methylglutaryl-CoA

HMGR: 3-hydroxy-3-methylglutaryl-CoA reductase

HMGS: 3-hydroxy-3-methylglutaryl-CoA synthase

IDI: Isopentenyl diphosphate isomerase)

IDS: 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate reductase

IPP: Isopentenyl pyrophosphate

ISA1: Isoamylase 1

ISA2: Isoamylase 2

ISA3: Isoamylase 3

IU: International unit

KUR: GBSS mutant strain

Lys: Lysine

MCS: 2C-methyl-d-erythritol 2,4-cyclodiphosphate synthase

ME-cPP: 2C-methyl-d-erythritol 2,4-cyclodiphosphate

MEP: 2C-methylerythritol 4-phosphate

mev: mevalonate

MgCl2: Magnesium Chloride

mRNA: Messenger RNA

MS: Murashige and Skoog medium

Musa: Musashimochi GBSS mutant strain

MVA: mevalonic acid

MVK: mevalonate kinase

MVP: mevalonate-3-phosphate

MVPP: mevalonate- 3,5-diphosphate

NAT6: member of the N-acetyltransferase family

NHEJ: Non-homologous end joining

N-terminal: Amine-terminal

Os: Oryza sativa

PAM: Protospacer adjacent motif

PCR: polymerase chain reaction

PHO: starch phosphorylase

PHO1/L: Plastidial starch phosphorylase

PHO2/H: Cytosolic starch phosphorylase

Pi: inorganic phosphate

PMD: mevalonate diphosphate decarboxylase

PMK: 5-phosphomevalonate decarboxylase

PPi: inorganic diphosphate

PTST: protein targeting to starch

PUL: Pullulanase

RNA: Ribonucleic acid

RNase: Ribonuclease

RT-PCR: real time PCR

SBE: Starch branching enzymes

SBEI: Starch branching enzyme 1

SBEIIb: Starch branching enzyme 2b

SDs: Standard deviations

SEM: Scanning Electron Microscope

sgRNA: single guide RNA

SS: Starch synthase

SSI: Starch synthase 1

SSIIa/b/c: Starch synthase 2 a, b or c

SSIIIa/b: Starch synthase 3 a or b

SSIVa/b: Starch synthase 4 a or b

SuSy: Sucrose synthase

T-DNA: transfer DNA

TEM: Transmission Electron Microscopy

Thr: Threonine

Tyr: Tyrosine

UBQ5: Ubiquitin 5

UDP: Uridine diphosphate

UDP-Glc: UDP-glucose

Waxy: GBSSI

WR1: WRINKLED 1

WT: Wild-type

Wx: Waxy

ZIP58: basic leucine zipper transcription facto

GENERAL INTRODUCTION

1. Food insecurity and human health

Food insecurity is defined as the inhability to consistently access or afford adequate and nutritionally balanced food to maintain a healthy life. Food security is one of the bigger global challenges. Food security exists when all people, at all times, have physical and economic access to sufficient, safe and nutritious food that meets their dietary needs and food preferences for an active and healthy life. (World Food Summit, 1996). The international community has committed to reaching zero hunger by 2030. But the estimations predict an increase in the world population to around 9 billion people by 2050 and this will increase the demand for food. The agricultural productivity required by 2050 ranges between 100-110% compared with levels of 2005 (Tilman et al., 2011). Eight hundred and fifteen million people do not get enough to eat (FAO, 2016; Figure 1). Another two billion people eat enough calories but not enough essential nutrients, and this translates to ca: half the world's population suffers from malnutrition (Graham et al., 2001).

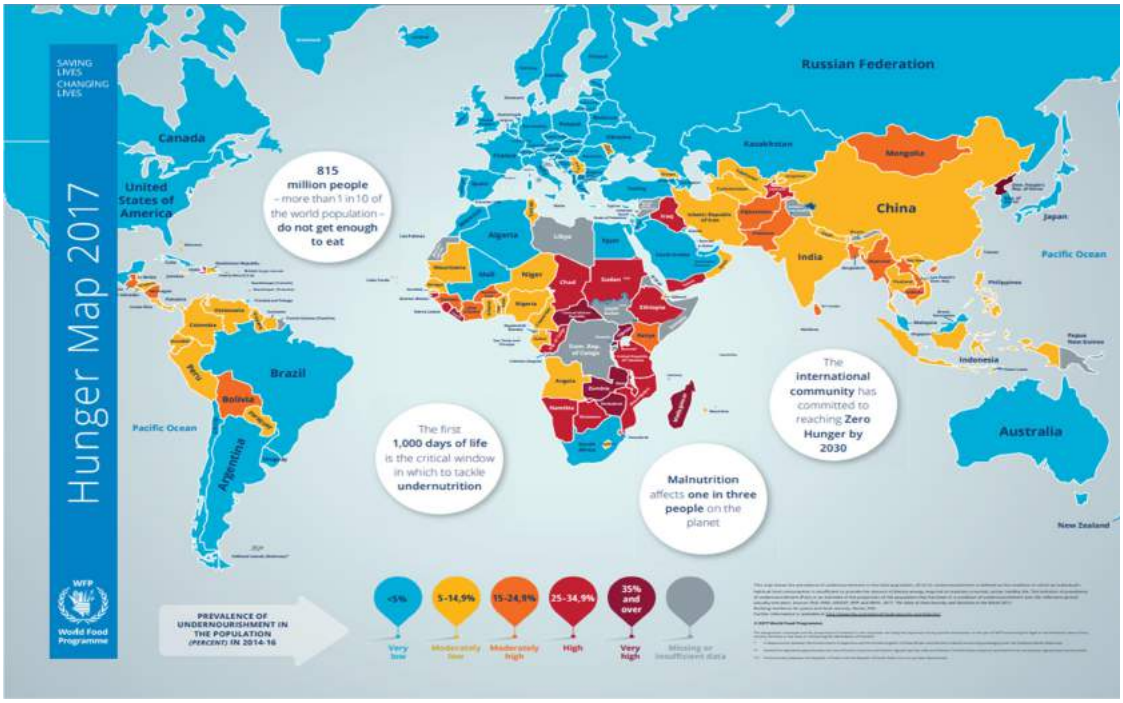


Figure 1. The 2017 hunger map (Source: World Food Programme 2017; https://www.wfp.org/content/2017-hunger-map)

Food deficiency also increases malnutrition, which reflects poverty, poor access to food or the poor distribution of food and agriculture based on individual crops that are deficient in essential nutrients. In developing countries, malnutrition is an endemic problem where the absence of a diverse diet generates an increase in the risk of deficiency diseases (Perez-Massot et al., 2013). Nevertheless, the increase in yield in

major crops such as rice has been very limited over recent years (Zhu et al., 2010). Plant biotechnology provides tools for improving crop productivity and in turn food security (Raines 2011).

2. Staple Food Biofortification

The most effective intervention to address micronutrient malnutrition is a varied diet that includes fresh fruit, vegetables, fish and meat. This can be a solution in developed countries, but is impractical in many developing countries because diverse food is not widely available or affordable (Gomez-Galera et al., 2010).

Fortification involves the addition of essential micronutrients to food (Zhu et al., 2011). Biofortification is another strategy where the nutrients in crops increase at source. There are two principal strategies for biofortification with essential nutrients, conventional breeding and genetic engineering (Farré et al., 2010). Both strategies attempt to create plant lines with additional genes that favour the most efficient biosynthesis and/or accumulation of essential nutrients and other compounds essential for health (Zhu et al., 2011). Conventional breeding improves crops without using recombinant DNA technology, sometimes in combination with mutagenesis or marker assisted selection to introgress genes from distant relatives, which similarly would not occur in nature (Bai et al., 2011). Conventional breeding has limitations such as the relatively long lead time to have an effect and its dependence on a compatible gene pool. The long time frames to obtain nutritionally improved lines are one of the greatest challenges in conventional breeding (Pérez-Massot et al., 2013).

Genetic engineering can generated biofortified crops by transferring genes directly into breeding lines, obtaining transgenic plants with enhanced nutritional traits. Compared to conventional breeding, genetic manipulation has the advantages of speed, direct engineering of breeding lines, simplicity, multiple simultaneous biofortification of different nutrients and unrestricted access to genetic diversity including recombinant genes that do not occur in nature (Zhu et al., 2007).

Even though genetically modified crops provide recognized socioeconomic benefits, it is very difficult to have them on the world market and politicians need to facilitate regulation of the current onerous regulatory bottlenecks. In the near future, nutritionally enhanced crops have the protection to save millions of lives and in the longer term will impact health, wellbeing and economic prosperity in developing countries (Farré at al., 2011).

2.1 Conventional Breeding

Conventional breeding can generate new plant varieties by selecting desirable traits/targets over multiple generations and for that, uses genetic material present within a species or sexually compatible species. Many plants show different nutritional traits due to the presence of major alleles (Farré et al., 2010).

Simple selection is the easier method used by our ancestor to improve crops where a genetically heterogeneous population of plants is screened and plants with the most desired traits are selected for continued propagation

Exposing seeds to mutagenic agents or chemical mutagens to induce random changes in the DNA sequence was a mutation breeding strategy used to generate some changes in the plants and to obtain new properties, normally not beneficial, without introducing genes from other species. Mutations induced in crops are not regulated for food or environmental safety per se (in most countries), and breeders generally do not conduct molecular genetic analyses to characterize the mutations or determine their extent (Broertjes and Van Harten., 2013).

2.2 Genetic Engineering

Genetic engineering is the deliberate modification of the characteristics of an organism by manipulating its genetic material. Over the past 50 years, genetic engineering (GE) has developed rapidly due to the greater understanding the structure and function of DNA. GE allows the transfer of one or more genes with desirable properties directly to an organism. To obtain a crop with a desirable trait, the time is shorter compared to conventional breeding and it permits modification of plants by inserting useful genes or removing or switching off particular genes with undesirable properties (Gepts, P.,2002).

There are diverse techniques to modify a plant genome. Gene transfer by *Agrobacterium tumefaciens*, a pathogenic bacterium that naturally transfers DNA to plants during the infection process. The transferred DNA is stably integrated into the plant genome, and the plant then expresses the transferred genes similarly to endogenous genes. By substituting the DNA of interest for the crown gall disease-causing DNA, scientists derived new strains of Agrobacterium that deliver and stably integrate specific new genetic material into the cells of target plant species (Toki., 1997). Another technique uses naked DNA that is delivered to plant cells by “bombardment” with microscopic metal particles to which DNA had been adhered. It is an effective physical method of DNA delivery, especially in species such as corn, rice, and other cereal grains (Altpeter et al., 2005).

Genome editing using engineered nucleases is an effective genetic engineering method that uses ‘molecular scissors’, or artificially engineered nucleases, to target and digest DNA at specific locations in the genome. The engineered nucleases induce a double-stranded DNA break (DSB) at the target site that is then repaired by the natural processes of homologous recombination (HR) or non-homologous end-joining (NHEJ) (Osakabe and Osakabe., 2014). Sequence modifications then occur at the cleaved sites, which include deletions or insertions that result in gene disruption in the case of NHEJ, or integration of exogenous sequences by HR in the presence of a repair DNA template with sufficient homology. Currently, four types of engineered nucleases are used for genome editing: engineered homing

endonucleases/meganucleases (EMN), zinc finger nucleases (ZFN), transcription activator-like effector nucleases (TALEN) and CRISPR (clustered regularly interspaced short palindromic repeats)/Cas 9 (CRISPR-associated). In particular, CRISPR/Cas9 is now used widely in various organisms (Figure 2; Osakabe and Osakabe., 2014) (Bortesi et al. 2016) (Zhu et al., 2017) .

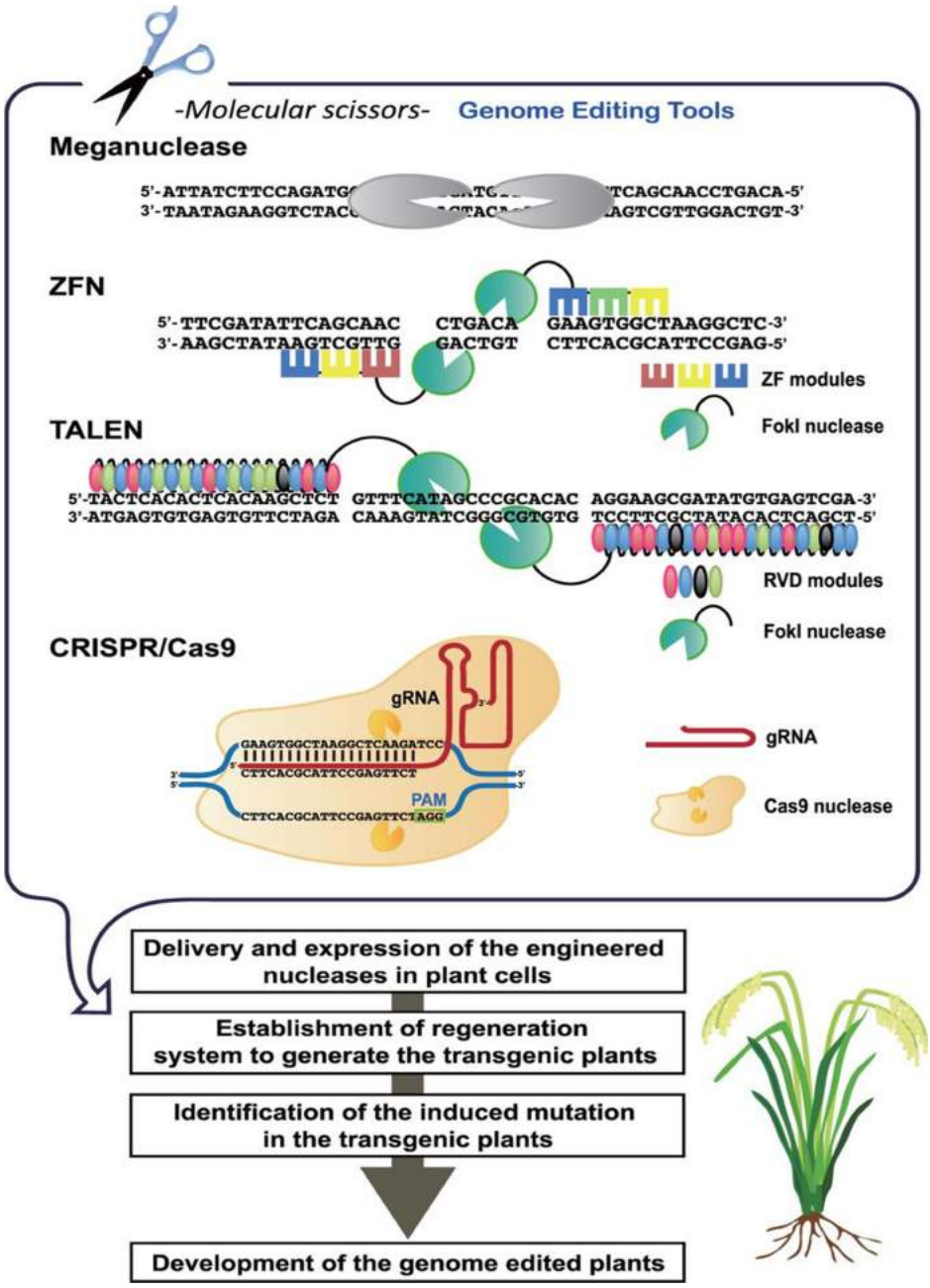


Figure 2. Current methods in Genome Editing EndoNucleases and an overview of the workflow of plant genome editing. Homing endonucleases/meganucleases (EMNs) recognize long (20 bp) DNA sequences. FokI nuclease is used as the DNA cleavage domain in ZFN, which bind the target DNAs via engineered zinc finger (ZF) domains and in TALEN that recognize the targets by the engineered TALEs composed of Repeat Variable Diresidue domains (the DNA binding domain contains a repeated highly conserved 33–34 amino acid sequence with divergent 12th and 13th amino acids. These two positions, referred to as the Repeat Variable Diresidue) derived from the plant pathogen Xanthomonas. The CRISPR/Cas9 system utilizes RNA-guided engineered nucleases (RGNs) that use a short guide RNA (gRNA) to recognize DNA sequences at the target sites (Osakabe and Osakabe., 2014).

3. Isoprenoid biosynthesis, function and relevance

Isoprenoids were discovered in plants. Synthesis of isoprenoids is essential in all living organism. Organisms that do not synthetize isoprenoids are obligate parasites (Boucher and Doolittle, 2000). Isoprenoids constitute the largest family of compounds found in nature with more than 50,000 knowns examples and they have an incalculable value for industry as useful components such as rubber, fuel, flavors, antibiotics, pharmaceuticals and plant hormones. All isoprenoids are derived from isoprene units assembled and modified in many different ways. Isoprene is the key precursor of all isoprenoids (Vranova et al., 2013). Isoprene is obtained mostly from petroleum reserves. Humans normally look for faster, more economical and more sustainable ways to produce chemicals that were originally derived from nature. There have been two different approaches for that objective. First, increasing the production of isoprenoid precursors and metabolites in engineered microbes such as *E.coli* or yeast, and cultivate them in bioreactors. Second, increasing their production in plants or plant cell cultures where they will be accumulated and extracted (Zhou et al., 2016).

There are two pathways (figure 3) involved in the biosynthesis of isoprenoids. The cytosolic mevalonate (MVA) and the plastidial 2-C-methyl-D-erythriol 4-phosphate (MEP) pathways produce the common precursors of isoprenoids, isopentenyl (IPP) and dimethylallyl diphosphate (DMAPP). However, most organisms use one of the pathways, except plants that use both pathways (Vranova et al., 2012). In the majority of organisms, MVA is present, including archaebacterial, some gram-positive bacteria, yeasts and animals. While most gram-negative bacteria, cyanobacteria and green algae use the MEP pathway (Disch et al., 1998; Kovacs et al., 2002).

IPP and DMAPP produced by the MVA pathway are used for the synthesis of isoprenoids in the cytosol and mitochondria; whereas IPP and DMAPP produced by the MEP pathway are used for the synthesis of isoprenoids in plastids. There is a cross-flow of IPP-DMAPP between both pathways (Berthelot et al., 2012).

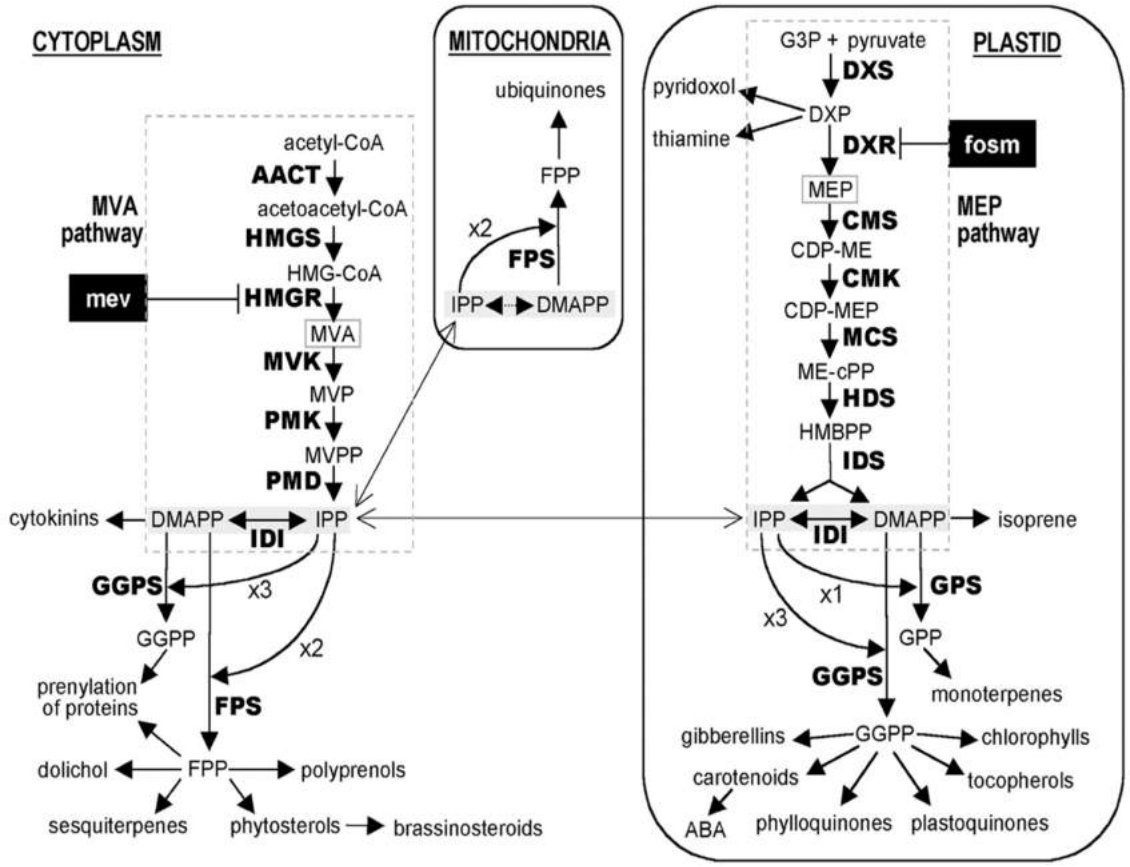


Figure 3. Isoprenoid biosynthesis pathways in the plant cell. HMG-CoA, Hydroxymethylglutaryl CoA; MVP, 5-phosphomevalonate; MVPP, 5-diphosphomevalonate; HBMPP, hydroxymethylbutenyl 4-diphosphate; FPP, farnesyl diphosphate; ABA, abscisic acid. The first intermediate specific to each pathway is boxed. Enzymes are indicated in bold: AACT, acetoacetyl CoA thiolase; HMGS, HMG-CoA synthase; HMGR, HMG-CoA reductase; MVK, MVA kinase; PMK, MVP kinase; PMD, MVPP decarboxylase; IDI, IPP isomerase; GPS, GPP synthase; FPS, FPP synthase; GGPS, GGPP synthase; DXS; DXR, DXP reductoisomerase; CMS; CMK;MCS; HDS; IDS, IPP/DMAPP synthase. The steps specifically inhibited by mevinolin (mev) and fosmidomycin (fosm) are indicated (Rodrıguez-Concepción and Boronat, 2002).

3.1 MVA pathway

Precursors for cytosolic and mitochondrial isoprenoids are synthesized via the MVA pathway. The equilibrium between IPP/DMAPP is controlled by IPP isomerase. Mutations in this pathway cause male sterility and arrest embryo development. The initial reaction of the MVA pathway is catalysed by acetyl-coenzyme A (CoA) *C*-acetyltransferase, also named acetoacetyl-CoA thiolase that condenses two molecules of acetyl-CoA to acetoacetyl-CoA in a reversible reaction. Acetoacetyl-CoA is converted to 3-hydroxy- 3-methylglutaryl-CoA (HMG-CoA) by HMG synthase. HMG-CoA is converted to mevalonic acid (MVA) in two reduction steps, each requiring NADPH as the reducing equivalent. The reaction is catalyzed by the enzyme 3-hydroxy-3-methylglutaryl-CoA reductase. All known plant HMGR proteins bind to the endoplasmic reticulum (ER) with their catalytic site facing the cytosol. In two successive reactions catalyzed by MVA kinase (MK) and phospho-MVA kinase (PMK), MVA is then phosphorylated to MVA 5-diphosphate. The last

step of IPP biosynthesis is an ATP-dependent decarboxylation of MVA 5- diphosphate, which is catalyzed by diphospho- MVA decarboxylase, also named MVA diphosphate decarboxylase (MPDC) (Vranova et al., 2013).

The first steps of isoprenoid biosynthesis are under strict regulation. The main limiting step in the pathway is the negative feedback of mevalonate to HMGR. HMGR is regulated at transcriptional and posttranslational levels (Burg and Espenshade, 2011). Light is one of the major environmental signals for the differential accumulation of metabolites. MVA is mostly expressed during the night (light has a negative effect in transcription) in seedlings and mature plants (in roots, seeds and flowers) (Ghassemian et al., 2006).

3.2 MEP pathway

The MEP pathway produces IPP and dimethylallyl diphosphate (DMAPP) in plastids. The MEP pathway starts with the condensation of thiamin and glyceraldehyde 3-phosphate (GA-3P) to form 1- deoxy-D-xylulose 5-phosphate (DXP). The condensation reaction is catalyzed by 1-deoxy-D-xylulose 5-phosphate synthase (DXS). The intramolecular rearrangement and reduction of DXP to MEP is catalyzed by DXP reductoisomerase (DXR). MEP is then converted to 4-(cytidine 5_-diphospho)-2-*C*-methyl-D-erythritol (CDP-ME) in a CTP-dependent reaction catalyzed by 2-*C*-methyl-D-erythritol 4-phosphate cytidylyltransferase (MCT). CDP-ME is further phosphorylated by 4-(cytidine 5_- diphospho)-2-*C*-methyl-D-erythritol kinase (CMK). The product 2-phospho-4-(cytidine 5_-diphospho)- 2-*C*-methyl-D-erythritol (CDP-ME2P) is subsequently converted to 2-*C*-methyl-Derythritol 2,4-cyclodiphosphate (MEcPP), which is catalyzed by 2-*C*-methyl-D-erythritol 2,4-cyclodiphosphate synthase (MDS). MEcPP is then reduced by 4-hydroxy-3-methylbut-2-enyldiphosphate synthase (HDS) to HMBPP, which is finally converted by HMBPP reductase (HDR) into a mixture of IPP and DMAPP (Vranova et al., 2013).

The MEP pathway is regulated at transcriptional and posttranscriptional levels (Hemmerlin et al., 2012). The genes are principally active in photosynthetic tissues and light increases the accumulation of transcripts (Ghassemian et al., 2006).

Both pathways are regulated by several other environmental signals such as osmotic stress, dehydration, high or low temperature, UV radiation, bacterial pathogens, herbivory, fungal elicitors, wounding and mycorrhiza fungae (Cordoba et al., 2009).

3.3 WRINKLED1 transcription factor

Rice homologue of the Arabidopsis transcription factor WRINKLED1 (WR1) controls the expression of genes for plastidial glycolysis and fatty acid biosynthesis and it is not expressed in photosynthetic tissues (Cernac and Benning; 2004). By controlling expression, availability of Acetyl-CoA increases. Normally, the pool of Acetyl-CoA in plastids is used for fatty acid biosynthesis and expressing an ectopically MVA pathway will limit the Acetyl-CoA in plastid (Kumar et al., 2012). By overexpressing WR1, an increase of precursors for the ectopic MVA pathway has been reported (Cernac and Benning; 2004).

4. Metabolic engineering to enhance nutrient content in plants

Organic molecules such as amino acids, fatty acids, terpenes and vitamins are synthesized by the plant, and increasing the nutritional content therefore requires some form of metabolic engineering with the aim of increasing the amount of desirable compounds. To this end, methods had to be developed for the coordinated expression of multiple genes (Capell & Christou, 2004) to meet objectives, such as: (a) enhance the activity of enzymes at multiple rate-limiting steps in target pathways, e.g. by overexpression or expression of enzymes that are released from feedback inhibition; (b) increase the availability of upstream precursors to enhance the flux through the target pathway; (c) modulate pathway branch points to prevent the loss of flux; and (d) promote the development of sink compartments to store target compounds (Figure 4) (Zorrilla-Lopez et al., 2013).

Most agronomic traits in plants are controlled by multiple genes, as is also the case for

the synthesis of complex organic compounds from primary and secondary metabolism, which often represent the outputs of long and convoluted metabolic pathways. Therefore, genetic engineering has seen a progressive change from single-gene intervention to multigene transformation to tackle increasingly ambitious objectives (Halpin, 2005). Examples of metabolic engineering in plants include primary metabolic pathways (carbohydrates, amino acids, and lipids) and secondary metabolic pathways (e.g. alkaloids, terpenoids, flavonoids, lignins, quinones, and other benzoic acid derivatives) (Zhu et al., 2011). These pathways generate a large number of compounds that are useful to humans, including energy-rich foods, vitamins and many different pharmaceuticals. The introduction of multiple genes into plants was initially achieved using iterative processes, such as successive rounds of crosses between transgenic lines (Halpin, 2005) or sequential retransformation (Blancquaert et al., 2013). Both methods are labour intensive in terms of breeding, as the transgenes are unlinked and segregate independently in later generations (Zorrilla-Lopez et al., 2013). The simultaneous introduction of two or more transgenes (cotransformation) via direct DNA transfer allows transgenic plants carrying multiple genes at one locus to be produced in a single

generation, which can be achieved using genes in tandem on the same transformation plasmid or unlinked genes on different transformation plasmids (Naqvi et al., 2010).

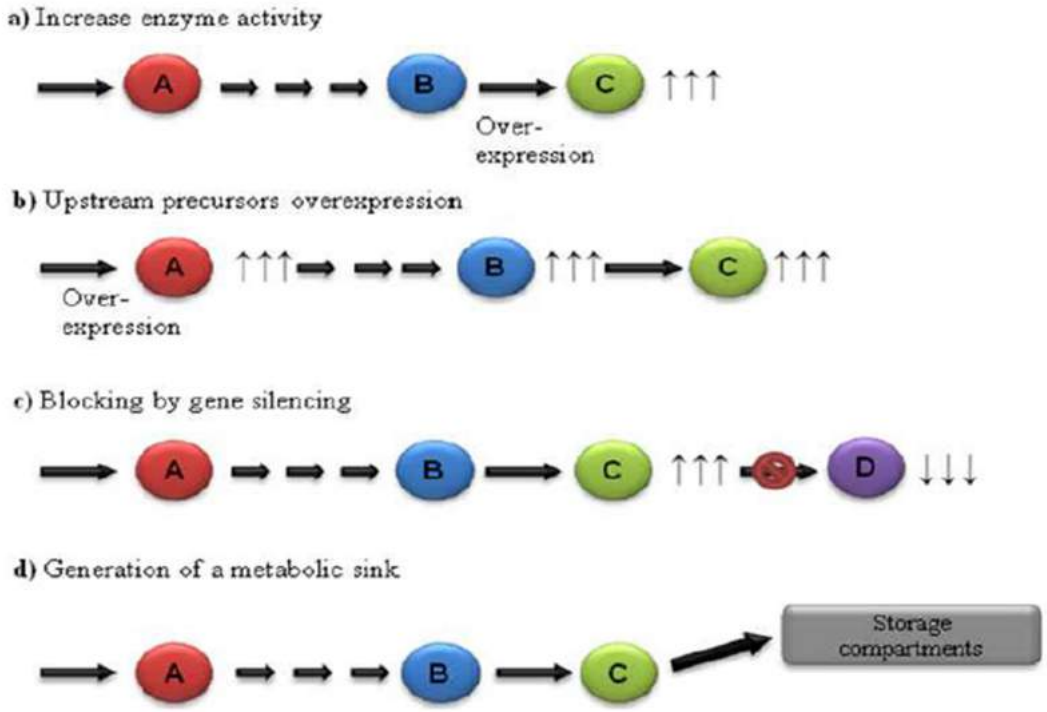


Figure 4. Strategies to modulate organic compound levels in plants. A and B are the precursors of C; C is the target product; D is the result of the target product conversion. (A) Modification of the activity of enzymes implicated in rate-limiting steps in the target pathway by modulation of one or two key enzymes, or multiple enzymes. (B) Upstream precursor enhancement by increasing flux through the pathway by overexpressing the enzyme(s) that catalyze(s) the first committed step of the pathway. (C) Blocked pathway branch points. (D) Enhanced accumulation of target metabolite by increasing sink compartments (Zorrilla-Lopez et al., 2013).

The next level of metabolic engineering is synthetic biology. Synthetic biology involves the *de novo* assembly of genetic systems using pre-validated components (Haseloff & Ajioka, 2009). In the context of metabolic engineering in plants, a synthetic biology approach might utilize specific promoters, genes, and other regulatory elements to create ideal genetic circuits that will facilitate the accumulation of particular metabolites. Synthetic biology combines engineering principles and mathematical models to predict and validate the behavior of the resulting system, which can be considered as the next step in multigene metabolic engineering because it removes any dependence on naturally occurring sequences and allows the design of ideal functional genetic circuits from first principles (Haseloff & Ajioka, 2009). Thus far, most work on synthetic biology has been carried out in microorganisms (Nikel et al., 2014; Cameron et al., 2014). Simple synthetic biology approaches have been described in plants, mostly in the context of signalling

pathways and development, but also in the development of phytodetectors (Zurbriggen et al., 2012) and biofortified crops (Naqvi et al., 2009). The use of synthetic biology in development as well as metabolism is important because it not only controls the metabolic capacity of a cell, but also steps one level up in terms of organization and use of particular promoters and genes that control developmental processes to generate novel tissues, in which the cells have specialized biosynthetic or storage functions to accumulate target products in particular organs. This approach will facilitate the achievement of goals that are unattainable by conventional genetic engineering, such as the development of novel organisms producing recombinant pharmaceuticals such as antibodies and vaccines, the production of biofuels, and the removal of hazardous waste (Purnick & Weiss, 2009).

5. Gene deactivation (knockout) using CRISPR-Cas9 in the starch biosynthetic pathway

A gene knockout is the outcome of a genetic intervention in which one gene is made inoperative. Knockouts are accomplished through a variety of methods. Originally, naturally occurring mutations were identified and then gene loss or inactivation had to be established by DNA sequencing or other methods. Recent studies demonstrated that KO genes with CRISPR-Cas9 technology is more efficient than with traditional breeding methods due to time consuming and laborious work of traditional methods and the advantage of CRISPR-Cas9 not affecting other genes because of target specificity (Perez et al., 2018; Zhang et al., 2018).

Site-directed nucleases such as CRISPR/Cas9 induce double-strand breaks (DSBs) at precise sites in the plant genome, allowing targeted modifications when the breaks are repaired by non-homologous end joining (NHEJ) or homology-dependent repair (HDR). The predominant repair pathway in plants is NHEJ which, in the absence of donor DNA, generates short indels in the form of insertions or deletions. Both events tend to cause gene knockouts. Homology dependent repair (HDR) occurs if a donor DNA template is available carrying the desired mutation albeit at much lower frequencies compared to NHEJ. In the absence of donor DNA, HR involving sister chromatids will restore the locus to its original state (Zhu et al., 2017).

Gene knockouts are primarily used to determine gene function by comparing the knockout organism to a wild-type with a similar genetic background (near isogenic) (Li et al., 2017).

5.1 Precision of Genome Editing Systems

Genome editing was first achieved using natural meganucleases (or hybrid/mutated/ engineered variants) with target sites up to 18 bp (Osakabe and Osakabe, 2015). The DSB is made within the target site and is a staggered cut with overhangs. Target site specificity, therefore, depends on the natural specificity of the

nuclease and any variation that can be introduced by creating hybrid enzymes or selecting variant binding specificities by mutation and screening (Rosen et al. 2006). This limitation has been overcome by designing nucleases with bespoke specificity – the two principal examples are zinc-finger endonucleases (ZFN), comprising the endonuclease domain from a restriction enzyme such as FokI, where the DSB is introduced outside the recognition site, combined with a series of zinc-finger DNA-binding modules each with specificity for a 3-bp sequence. A bespoke nuclease can therefore be generated by arranging several zinc fingers as a tandem array. Target specificity can be designed by selecting appropriate combinations of zinc fingers (Kim et al. 1996). Transcription activator-like effector nucleases (TALEN) are similar in principle to ZFN. A promiscuous endonuclease domain is paired with multiple transcription activator-like effector domains that recognize single base pairs (Christian et al.2010). The fourth and most recent genome editing technology is based on a form of bacterial adaptive immunity that neutralizes previously encountered invasive DNA sequences by expressing clustered regularly interspaced short palindromic repeats (CRISPRs) representing DNA fragments (spacers) captured from invading pathogens. The CRISPR/Cas9 system differs from the other three systems in that the specificity of binding is determined by a synthetic guide RNA (sgRNA). The resulting CRISPR RNAs act as guides for CRISPR-associated (Cas) nucleases that attack the same pathogens if they enter the cell again (Marraffini and Sontheimer, 2010). Genome editing using the CRISPR system is achieved by constructing synthetic guide RNAs (sgRNAs) that direct the Cas nuclease to genomic targets, in contrast to the other three systems which are protein-guided (Jinek et al. (2012). Genome editing is primarily based on the CRISPR/Cas9 system (Bortesi and Fischer, 2015) although alternatives such as the CRISPR/Cpf1 system (with the difference that it introduces a staggered DSB) have been described more recently. A similar outcome can be achieved by assembling two CRISPR/Cas9 complexes on distinct but adjacent targets and inactivating one of the nuclease domains so that each complex introduces a single-strand break (Cas9/D10A) (Zhu et al., 2017).

The key advantages of CRISPR/Cas9 include: no dependence on random recombination or integration events; only two components (Cas9 endonuclease and synthetic guide RNA are recruited); CRISPR/Cas9 specificity relies on 20-nt genomic targets (referred to 'spacers') and it is possible to find unique spacers in most genes, particularly in species that have proven genetically intractable (plant species with unsequenced genomes); the mutation frequency with CRISPR/Cas9 is much higher than with other genome editing technologies; it can be used to insert multiple sgRNAs or designing a promiscuous sequence that can be introduced simultaneously into a cell and generates simultaneous mutations of different genes and the creation of more extensive mutations; CRISPR/Cas9 can also be combined to induce large deletions and other rearrangements; and is simple and efficient, unlike synthetic transcription activator-like effector nucleases (TALEN) where cloning is still laborious and plasmid instabilities can occur (Fauser et al. 2014; Mikami et al. 2016; Ran et al. 2013).

5.2 Starch biosynthetic pathway genes

Starch is the major component in rice grain endosperm and it is a mixture of two components, amylose and amylopectin. It is accumulated in the cereal endosperm as an energy reserve for germination. Amylose is a polysaccharide with linear chains of α (1,4) linked glucose residues while amylopectin contains additional α (1,6) linked branches every 24-30 residues and is highly branched (Martin and Smith 1995). Starch from different plant species varies substantially in its physicochemical properties due to the relative ratio of amylose to amylopectin, and differences in chain length and/or amylopectin branching density (Jobling 2004).

In plants starch synthesis involves the conversion of glucose 1-phosphate to ADP-glucose by ADP-glucose pyrophosphorylase (AGPase), which is then converted to amylose by the co-ordinated action of AGPase and granule-bound starch synthase (GBSS). Amylopectin is synthesized by the coordinated action of ADP-glucose pyrophosphorylase (AGPase), starch synthase (SS), starch branching enzyme (BE), and starch debranching enzyme (DBE). Starch branching enzyme introduces α (1,6) glycosidic bonds to generate amylopectin, a reaction that is reversed by the starch debranching enzyme isoamylase (ISA). Earlier studies suggest that disproportionating enzyme (DPE) and starch phosphorylase (PHO) play a role in initiation and elongation of the starch structure (Satoh et al. 2008)) (Figure 5).

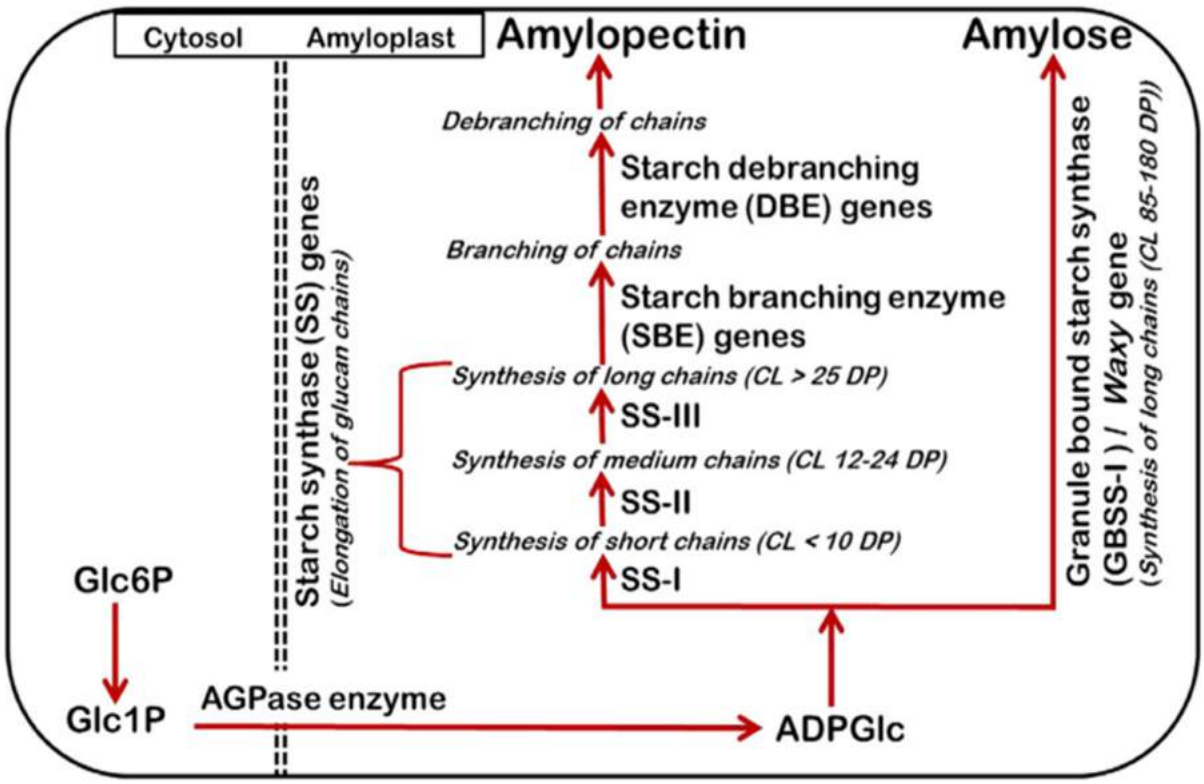


Figure 5. Diagrammatic representation of starch pathway steps to generate the different components of starch (amylose and amylopectin) in rice endosperm (Pandey et al., 2012). Abreviations: Glc6P (glucose-6-phosphate); Glc1P (glucose-1-phosphate); ADPGlc (ADP-Glucose) and AGPase (ADP-glucose pyrophosphorylase).

5.2.1 AGPase large subunit 2 gene

Amylose is synthesized by AGPase and GBSS whereas amylopectin also requires SBE and DBE to introduce and refine its branching structure (Ohdan et al. 2005). The heterotetrameric AGPase catalyzes the first committed step in starch synthesis. In rice there are two genes encoding small catalytic subunits (*APS1* and *APS2*) and four encoding larger regulatory subunits (*APL1*, *APL2*, *APL3* and *APL4*). Furthermore, *OsAPS2* produces two distinct polypeptides through alternative splicing: *APS2a* and *APS2b*, the former including a transit peptide for import into the plastids and the latter lacking this sequence causing it to remain in the cytosol (Tang et al. 2016). AGPases are key enzymes in the starch biosynthesis pathway and are regulated by the ratio 3-PGA/Pi (Preiss et al. 1982). Among the larger regulatory subunits, only APL2 lacks a transit peptide and remains in the cytosol, whereas the other three subunits are imported into the plastid. Thus, only one AGPase assembles in the cytosol, comprising subunits APL2 and APS2b. *OsAPL2* encodes a 518 amino acid polypeptide with a catalytic site spanning residues 88–364. The catalytic site is composed of α-helices and β-sheets that form a substrate binding cleft. A further α-helix and β-sheet outside the catalytic center are required to maintain the correct tertiary conformation and bind with the other subunit, APS2b, to form the tetrameric structure (Zhang et al. 2012).

APS1 and *APL1* are expressed strongly during early endosperm development, whereas expression of *APS2b* and *APL2* begins 3 days after fertilization and remains at high levels thereafter. *APL3* is expressed at low levels throughout seed development, and the *APL4* and *APS2a* transcripts are barely detected (Ohdan et al. 2005). This suggests that the endosperm contains one predominant plastid AGPase, comprising APL1 and APS1, which is important during early development, and one predominant cytosolic AGPase, comprising APL2 and APS2b, which is important during the middle and late stages of development, when starch accumulates (Ohdan et al. 2005). During seed development, starch is accumulated in amyloplasts which serve as a reservoir for the germinating seed. The major tetramer (APL2-APS2b) is in the cytosol and generates stable (i.e. not transitory) starch in this compartment (Rychter AM and Rao IM 2005). Non transitory starch serves as a reservoir for seed development. In leaf, starch accumulation is transitory in chloroplasts to generate photosynthetic precursors. The starch deposited in chloroplasts is degraded during the night and resulting G-1-P is converted to triose-P and exchanged for Pi from the cytosol (Heldt et al., 1977; Stitt et al., 1981; Dennis et al., 1982; Lee SK et al., 2016)). Pi is an activator of photosynthetic enzymes, including Rubisco (Heldt et al., 1978; Bhagwat 1981). Thus, in leaves the main AGPases are plastidial. *APS2a* and *APL3* are strongly expressed in young leaves, whereas *APS1* is expressed at low levels, *APS2b* is not detected at all, and the remaining subunits (*APL1*, *APL2* and *APL4*) are minimally expressed. Later in development, *APS2a* and *APL3* expression declines (although they are still the most abundant subunits) and *APL1* expression increases to parity with *APL3*. This suggests that the major leaf AGPase initially comprises subunits *APS2a* and *APL3*, with *APS1* and *APL3* combining to form a less abundant subunit, but the increase in *APS1* expression may result in a progressive accumulation of the *APS1/APL3*

heterotetramer later in development. Importantly, all the AGPases in the leaf appear to be located in the plastid given that no cytosolic forms are expressed at any developmental stage [*APS2b* is not detected in leaves (Ohdan et al. 2005)].

5.2.2 Granule-bound starch synthase I gene

There are two major groups of starch synthases. The first group is the classical starch synthases (SS) and comprises four isoforms, some represented by more than one paralog: SSI, SSIIa/b/c, SSIIIa/b and SSIVa/b (Nakamura, 2002). These synthesize the linear chains of amylopectin and their distribution between granular and stromal fractions can vary between species, tissues and developmental stages (Ball and Morell, 2003). The second group is the granule-bound starch synthases (GBSS) which are restricted to the granule matrix. There are two isoforms (GBSSI and GBSSII), the first mainly expressed in the endosperm and the second mainly in the leaves (Ohdan et al., 2005). GBSSI catalyzes the extension of long glucan chains in amylose (Maddelein et al., 1994). The GBSSI protein is 609 amino acids in length with a catalytic site, spanning residues 78–609, composed of α-helices and β-sheets that form a substrate binding cleft. A further transit peptide outside the catalytic center is required to maintain the correct location and bind with the substrates (Momma et al., 2014).

In rice, GBSSI is encoded by the *Waxy* gene, so named because of the waxy appearance of the amylose-free grain in *Wx* null mutants (Hirano, 1993). As well as the *Wx* allele, two other natural alleles are common in rice (Sano 1984). The *Wxa* allele is predominant among indica subspecies and has strong GBSSI activity and thus more amylose in the endosperm, whereas *Wxb* has weaker GBSSI activity and the amount of amylose and amylopectin is more variable (Hirano et al., 1991; Umenmoto and Terashima, 2002). Because of the impact on grain quality, *Wx* is a key target in rice quality improvement programs and in studies of starch biosynthesis and metabolism (Tran et al., 2011; Zhang et al., 2012).

References

Altpeter, F., Baisakh, N., Beachy, R., Bock, R., Capell, T., Christou, P., Daniell, H., Datta, K., Datta, S., Dix, P.J. and Fauquet, C. (2005). Particle bombardment and the genetic enhancement of crops: myths and realities. Molecular Breeding, 15: 305-327.

Bai, C., Twyman, R. M., Farré, G., Sanahuja, G., Christou, P., Capell, T., & Zhu, C. (2011). A golden era-pro-vitamin A enhancement in diverse crops. In Vitro Cellular & Developmental Biology-Plant, 47: 205-221.

Ball, SG. and Morell, MK. (2003). From bacterial glycogen to starch: Understanding the biogenesis of the plant starch granule. Annual Review of Plant Biology, 54:207-233.

Berthelot, K., Estevez, Y., Deffieux, A., & Peruch, F. (2012). Isopentenyl diphosphate isomerase: a checkpoint to isoprenoid biosynthesis. Biochimie, 94: 1621-1634.

Bhagwat, AS. (1981). Activation of spinach ribulose 1,5-bisphosphate carboxylase by inorganic phosphate. Plant Sci Lett 23:197-206

Blancquaert, D., Storozhenko, S., Van Daele, J., Stove, C., Visser, R. G. F., Lambert, W., and Van Der Straeten, D. (2013). Enhancing pterin and para-aminobenzoate content is not sufficient to successfully biofortify potato tubers and Arabidopsis thaliana plants with folate. Journal of Experimental Botany, 64:3899–909.

Bortesi, L., Zhu, C., Zischewski, J., Perez, L., Bassié, L., Nadi, R., Forni, G., Lade, S.B., Soto, E., Jin, Xin., Medina, V., Villorbina, G., Muñoz, P., Farré, G., Fischer, R., Twyman, R.M., Capell, T., Christou, P and Schillberg, S. (2016). Patterns of CRISPR/Cas9 activity in plants, animals and microbes. Plant biotechnology journal, 14: 2203-2216.

Boucher, Y., & Doolittle, W. F. (2000). The role of lateral gene transfer in the evolution of isoprenoid biosynthesis pathways. Molecular microbiology, 37: 703-716.

Broertjes, C., & Van Harten, A. M. (2013). Applied mutation breeding for vegetatively propagated crops (Vol. 12). Elsevier, Wageningen, Netherlands.

Burg, J. S., & Espenshade, P. J. (2011). Regulation of HMG-CoA reductase in mammals and yeast. Progress in lipid research, 50: 403-410.

Cameron, D. E., Bashor C. J., and Collins, J. J. (2014). A brief history of synthetic biology. Nature Reviews Microbiology, 12: 381–390.

Capell T, and Christou, P. (2004). Progress in plant metabolic engineering. Current Opinion in Biotechnology, 15:148–154.

Cernac, A., and Benning, C. (2004). WRINKLED1 encodes an AP2/EREB domain protein involved in the control of storage compound biosynthesis in Arabidopsis. The Plant Journal, 40: 575-585.

Christian, M., Cermak, T., Doyle, E. L., Schmidt, C., Zhang, F., Hummel, A., Bogdanove, A.J. and Voytas, D. F. (2010). Targeting DNA double-strand breaks with TAL effector nucleases. Genetics, 186: 757-761.

Constabel, F. (1976). Somatic hybridization in higher plants. In vitro, 12: 743-748.

Cordoba, E., Salmi, M., and León, P. (2009). Unravelling the regulatory mechanisms that modulate the MEP pathway in higher plants. Journal of Experimental Botany, 60: 2933-2943.

Dennis, DT, Miernyk, JA. (1982). Compartmentation of non-photosynthetic carbohydrate metabolism. Ann. Rev. Plant Physiol, 33:27-50

Disch, A., Schwender, J., Müller, C., Lichtenthaler, H. K., and Rohmer, M. (1998). Distribution of the mevalonate and glyceraldehyde phosphate/pyruvate pathways for isoprenoid biosynthesis in unicellular algae and the cyanobacterium Synechocystis PCC 6714. Biochemical Journal, 333: 381-388.

Farré, G., Sanahuja, G., Naqvi, S., Bai, C., Capell, T., Zhu, C., & Christou, P. (2010). Travel advice on the road to carotenoids in plants. Plant Science, 179: 28-48.

Farre, G., Twyman, R. M., Zhu, C., Capell, T., & Christou, P. (2011). Nutritionally enhanced crops and food security: scientific achievements versus political expediency. Current opinion in biotechnology, 22: 245-251.

Fauser, F., Schiml, S, and Puchta, H. (2014). Both CRISPR/Cas-based nucleases and nickases can be used efficiently for genome engineering in Arabidopsis thaliana. Plant J, 79:348-359

Gepts, P. (2002). A comparison between crop domestication, classical plant breeding, and genetic engineering. Crop Science, 42, 1780-1790.

Ghassemian, M., Lutes, J., Tepperman, J. M., Chang, H. S., Zhu, T., Wang, X., Quail, P.H. and Lange, B. M. (2006). Integrative analysis of transcript and metabolite profiling data sets to evaluate the regulation of biochemical pathways during photomorphogenesis. Archives of Biochemistry and Biophysics, 448, 45-59.

Gómez-Galera, S., Rojas, E., Sudhakar, D., Zhu, C., Pelacho, A. M., Capell, T., and Christou, P. (2010). Critical evaluation of strategies for mineral fortification of staple food crops. Transgenic research, 19: 165-180.

Graham, R. D., Welch, R. M., & Bouis, H. E. (2001). Addressing micronutrient malnutrition through enhancing the nutritional quality of staple foods: principles, perspectives and knowledge gaps. Advances in agronomy, 70: 78-142.

Gupta, PK., and Tsuchiya, T., (1991). *Chromosome Engineering in Plants: Genetics, Breeding Evolution* (Vol. 2B) Elsevier, Amsterdam

Halpin, C. (2005). Gene stacking in transgenic plants--the challenge for 21st century plant biotechnology. Plant Biotechnology Journal, 3:141–55.

Haseloff, J., and Ajioka, J. (2009). Synthetic biology: history, challenges and prospects. Journal of the Royal Society Interface, 6(Suppl 4), S389–S391.

Heldt HW, Chon CH, Maronde D, Herold A, Stankovic AZ, Walker DA, Kraminer A, Kirk MR, Heber U (1977) Role of orthophosphate and other factors in the regulation of starch formation in leaves and isolated chloroplasts. Plant Physiol 59:1146-1155

H. W. Heldt, C. J. Chon and G. H. Lorimer. (1978). Phosphate requirement for the Light Activation of Ribulosc-1.5- Bisphosphate Carboxylase in Intact Spinach Chloroplast. FEBS Letters, Vol. 92, 2: 234-240

Hirano, HY. (1993). Genetic variation and gene regulation at the wx locus of rice. Gamma field symposia. 32:19-31.

Hirano, H. Y., and Sano, Y. (1991). Molecular characterization of the waxy locus of rice (Oryza sativa). Plant and cell physiology, 32: 989-997.

Ivics, Z., and Izsvák, Z. (2010). The expanding universe of transposon technologies for gene and cell engineering. Mobile DNA, 1: 25.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. Science, 1225829.

Jobling, S. (2004). Improving starch for food and industrial applications. Curr Opin Plant Biol, 7:210-218

Kim, Y. G., Cha, J., & Chandrasegaran, S. (1996). Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proceedings of the National Academy of Sciences, 93: 1156-1160.

Kovacs, W. J., Olivier, L. M., and Krisans, S. K. (2002). Central role of peroxisomes in isoprenoid biosynthesis. Progress in lipid research, 41: 369-391.

Kozukue, N., Misoo, S., Yamada, T., Kamijima, O., and Friedman, M. (1999). Inheritance of morphological characters and glycoalkaloids in potatoes of somatic hybrids between dihaploid Solanum acaule and tetraploid Solanum tuberosum. Journal of agricultural and food chemistry, 47: 4478-4483.

Kumar, S., Hahn, F. M., Baidoo, E., Kahlon, T. S., Wood, D. F., McMahan, C. M., Cornish, K., Keasling, J.D., Daniell, H. and Whalen, M. C. (2012). Remodeling the isoprenoid pathway in tobacco by expressing the cytoplasmic mevalonate pathway in chloroplasts. Metabolic engineering, 14: 19-28.

Lee, S.K., Hwang, S.K., Han, M., Eom, J.S., Kang, H.G., Han, Y., Choi, S.B. and Jeon, J.S. (2007) Identification of the ADP-glucose pyrophosphorylase isoforms essential for starch synthesis in the leaf and seed endosperm of rice (Oryza sativa L.). Plant Mol Biol, 65:531-546

Li, C., Powell, P. O., and Gilbert, R. G. (2017). Recent progress toward understanding the role of starch biosynthetic enzymes in the cereal endosperm. Amylase, 1: 59-74.

Maddelein ML, Libessart N, Bellanger F, Delrue B, D'Hulst C, Vvan den Koornhuyse, N, Fontaine T,Wieruszeski JM, Decq, A. and 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. Journal of biological chemistry, 269: 25150-25157.

Marraffini, L.A. and Sontheimer, E.J. (2010). CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. Nat. Rev. Genet., 11, 181–190

Martin, C. and Smith, AM. (1995). Starch biosynthesis. Plant Cell, 7: 971-985

Mikami, M., Toki, S. and Endo, M. (2016). Precision targeted mutagenesis via Cas9 paired nickases in rice. Plant Cell Physiol, 57: 1058-1068

Momma, M., and Fujimoto, Z. (2014). Interdomain disulfide bridge in the rice granule bound starch synthase I catalytic domain as elucidated by X-ray structure analysis. Bioscience, biotechnology, and biochemistry, 76: 1591-1595

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

Naqvi, S., Farre, G., Sanahuja, G., Capell, T., Zhu, C., and Christou, P. (2010). When more is better: multigene engineering in plants. Trends in Plant Science, 15: 48–56.

Neuhaus, G., and Spangenberg, G. (1990). Plant transformation by microinjection techniques. Physiologia Plantarum, 79: 213-217.

Niedz, R. P., McKendree, W. L., and Shatters, R. C. (2003). Electroporation of embryogenic protoplasts of sweet orange (Citrus sinensis (L.) Osbeck) and regeneration of transformed plants. In Vitro Cellular & Developmental Biology-Plant, 39: 586.

Nikel, P.I., Martinez-Garcia, E., and de Lorenzo, V. (2014). Biotechnological domestication of pseudomonads using synthetic biology. Nature Reviews Microbiology, 12: 368–379.

Ohdan, T., Francisco, PB Jr., Sawada, T., Hirose, T., Terao, T., Satoh, H. and Nakamura, Y. (2005) Expression profiling of genes involved in starch synthesis in sink and source organs of rice. J. Exp. Bot. 56: 3229-3244

Osakabe, Y., and Osakabe, K. (2014). Genome editing with engineered nucleases in plants. Plant and Cell Physiology, 56: 389-400.

Pandey, MK., Rani, NSM., Sheshu, M., Sundaram, RM., Varaprasad, GS., Sivaranjani,, AKP., Bohra, A., Kumar, GR., and Kumar, A. (2012). Different isoforms of starch-synthesizing enzymes controlling amylose and amylopectin content in rice (Oryza sativa L.). Biotechnolgy Advances 30: 1697-1706.

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, 1-17.

Perez-Massot, E., Banakar, R., Gomez-Galera, S., Zorrilla-Lopez, U., Sanahuja, G., Arjo, G., Miralpeix, B., Vamvaka, E., Farre, G., Rivera, S.M., Dashevskaya, S., Berman, J., Sabalza, M., Yuan, D., Bai, C., Bassie, L., Twyman, R.M., Capell, T., Christou, P., and Zhu, C. (2013). The contribution of transgenic plants to better health through improved nutrition: Opportunities and constraints. Genes and Nutrition, 8: 29-41.

Preiss, J. (1982). Regulation of the biosynthesis and degradation of starch. Ann Rev Plant Physiol 33: 431-454

Purnick, P. E. M., and Weiss, R. (2009). The second wave of synthetic biology: from modules to systems. Nature Reviews. Molecular Cell Biology, 10: 410–22.

Raines, C. A. (2010). Increasing photosynthetic carbon assimilation in C3 plants to improve crop yield: current and future strategies. Plant physiology, 110.

Ran, F.A., Hsu, P.D., Lin, C.Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y. and Zhang, F. (2013) Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell 154: 1380-1389

Rodrıguez-Concepción, M., and Boronat, A. (2002). Elucidation of the methylerythritol phosphate pathway for isoprenoid biosynthesis in bacteria and plastids. A metabolic milestone achieved through genomics. Plant physiology, 130: 1079-1089.

Rosen, L. E., Morrison, H. A., Masri, S., Brown, M. J., Springstubb, B., Sussman, D., Stoddard, B.L. and Seligman, L. M. (2006). Homing endonuclease I-CreI derivatives with novel DNA target specificities. Nucleic acids research, 34: 4791-4800

Rowland, GG., McHughen, AG., Hormis, YA. and Rashid, KY. (2002). CDC Normandy flax. *Can J Plant Sci,* 82: 425–426.

Rychter, AM. and Rao, IM. (2005). Role of phosphorus in photosynthetic carbon metabolism. In: Pessarakli M. (ed) Handbook of photosynthesis, 2nd edn. Taylor y Francis group, Tucson, pp 123-148

Sano, Y. (1984). Theoret. Appl. Genetics 68: 467.

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. and Nakamura, Y. (2008). Mutation of the plastidial α-glucan phosphorylase gene in rice affects the synthesis and structure of starch in the endosperm. The Plant Cell, 20(7), 1833-1849.

Shen, X., Gmitter, F. G., and Grosser, J. W. (2011). Immature embryo rescue and culture. In Plant Embryo Culture (pp. 75-92). Humana Press.

Stitt, M. and Heldt, HW. (1981). Physiological rates of starch breakdown in isolated intact spinach chloroplasts. Plant Physiol, 68:755-761

Tang, X.J., Peng, C., Zhang, J., Cai, Y., You, X.M., Kong F., Zhang, W.W. and Wan, J.M. (2016). ADP-glucose pyrophosphorylase large subunit 2 is essential for storage substance accumulation and subunit interactions in rice endosperm. Plant Sci, 249:70-83

Tilman, D., Balzer, C., Hill, J., and Befort, B. L. (2011). Global food demand and the sustainable intensification of agriculture. Proceedings of the National Academy of Sciences, 108: 20260-20264.

Toki, S. (1997). Rapid and efficient Agrobacterium-mediated transformation in rice. Plant Molecular Biology Reporter, 15: 16-21.

Tran, N. A., Daygon, V. D., Resurreccion, A. P., Cuevas, R. P., Corpuz, H. M., and Fitzgerald, M. A. (2011). A single nucleotide polymorphism in the Waxy gene explains a significant component of gel consistency. Theoretical and Applied Genetics, 123: 519-525.

Umemoto, T. and Terashima, K. (2002). Activity of granule-bound starch synthase is an important determinant of amylose content in rice endosperm. Functional plant biology 29:1121-1124.

Vranová, E., Coman, D., and Gruissem, W. (2012). Structure and dynamics of the isoprenoid pathway network. Molecular plant, 5: 318-333.

Vranová, E., Coman, D., and Gruissem, W. (2013). Network analysis of the MVA and MEP pathways for isoprenoid synthesis. Annual review of plant biology, 64: 665-700.

Zhang, D., Wu, J., Zhang, Y. and Shi, C. (2012). Phenotypic and candidate gene analysis of a new floury endosperm mutant (osagpl2-3) in rice. Plant Mol Biol Report 30: 1303-1312

Zhang, J., Zhang, H., Botella, J. R., and Zhu, J. K. (2018). Generation of new glutinous rice by CRISPR/Cas9-targeted mutagenesis of the Waxy gene in elite rice varieties. Journal of integrative plant biology, 60: 369-375.

Zhou, K., Edgar, S., & Stephanopoulos, G. (2016). Engineering Microbes to Synthesize Plant Isoprenoids. In Methods in enzymology (Vol. 575, pp. 225-245). Academic Press.

Zhu, C., Bortesi, L., Baysal, C., Twyman, R. M., Fischer, R., Capell, T., Schillberg, S. and Christou, P. (2017). Characteristics of genome editing mutations in cereal crops. Trends in plant science, 22: 38-52.

Zhu, C., Farre, G., Gomez-Galera, S., Shaista, N., Chao, B., Sanahuja, G., Yuan, D., Zorrilla, U., Tutusaus Codony, L., Rojas, E., Fibla, M., Twynman, R.M., Capell, T., and Christou, P. (2011). Transgenic Crops with Enhanced Nutritional Traits. In Encyclopedia of Sustainability Science and Technology. Meyers, R. A., Eds.: Springer Science Business Media LLC: NY, US, 599:293–326.

Zhu, C., Naqvi, S., Gomez-Galera, S., Pelacho, A.M, Capell, T., and Christou, P. (2007). Transgenic strategies for the nutritional enhancement of plants. Trends in Plant Science 12: 548–555.

Zhu, X. G., Long, S. P., and Ort, D. R. (2010). Improving photosynthetic efficiency for greater yield. Annual review of plant biology, 61: 235-261.

Zorrilla-Lopez, U., Masip, G., Arjo, G., Bai, C., Banakar, R., Bassie, L., Berman, J., Farre, G., Miralpeix, B., Perez-Massot, E., Sabalza, M., Sanahuja, G., Vamvaka, E., Twyman, R.M., Christou, P., Zhu, C., and Capell, T. (2013). Engineering metabolic pathways in plants by multigene transformation. International Journal of Developmental Biology, 57:565–576.

Zurbriggen, M. D., Moor, A., and Weber, W. (2012). Plant and bacterial systems biology as platform for plant synthetic biotechnology. Journal of Biotechnology, 160:80–90.

AIMS & OBJECTIVES

The overarching aim of my dissertation was to modulate the metabolism of rice by two different methods. The first aim was to knockout genes involved in the starch metabolic pathway (AGPase large subunit 2 and granule-bound starch synthase) to modify starch content, starch structure and to elucidate the regulation of the starch pathway. The second aim was to introduce the MVA pathway ectopically in plastids, in order to modify terpenoids content and composition.

My specific objectives were to:

1. Target an exon downstream of the active site of APL2 to maintain some gene expression activity, while perturbing the tertiary structure.
2. Target first exon of APL2 aiming to truncate the enzyme substantially and abolish its activity entirely.
3. Achieve a targeted knock-out of the *Waxy* gene usingCRISPR-Cas9D10A and two gRNAs.
4. To increase the flux in the MVA pathway by over-expression of a truncated and deregulated HMGR in plastids
5. Investigate the influence of introducing an ectopic tHMGR on fluxes between the MVA and MEP pathways.
6. To introduce required to duplicate an ectopic MVA pathway in rice plastids
7. To confirm the stable expression and functioning of the ectopic MVA pathway and ascertain interactions or cross talk with the native MVA and MEP pathways
8. To overexpress the transcription factor WRINKLED1 to increase the supply of plastidial acetyl-CoA

CHAPTER 1

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

CHAPTER 1: 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

1.1 Abstract

The first committed step in the endosperm starch biosynthetic pathway is catalyzed by the cytosolic glucose-1-phosphate adenylyl transferase (AGPase) comprising large and small subunits encoded by the *OsAPL2* and *OsAPS2b* genes, respectively. *OsAPL2* is expressed solely in the endosperm so we hypothesized that mutating this gene would block starch biosynthesis in the endosperm without affecting the leaves. We used CRISPR/Cas9 to create two heterozygous mutants, one with a severely truncated and nonfunctional AGPase and the other with a C-terminal structural modification causing a partial loss of activity. Unexpectedly, we observed starch depletion in the leaves of both mutants and a corresponding increase in the level of soluble sugars. This reflected the unanticipated expression of both *OsAPL2* and *OsAPS2b* in the leaves, generating a complete ectopic AGPase in the leaf cytosol, and a corresponding decrease in the expression of the plastidial small subunit *OsAPS2a* that was only partially complemented by an increase in the expression of *OsAPS1*. The new cytosolic AGPase was not sufficient to compensate for the loss of plastidial AGPase, most likely because there is no wider starch biosynthesis pathway in the leaf cytosol and because pathway intermediates are not shuttled between the two compartments.

1.2 Introduction

Rice (*Oryza sativa* L.) is one of the most important food crops in the world, accounting for 21% of the calories and 15% of the protein consumed by humans globally, and more than 70% of calories consumed by developing country populations in Asia (Yuan et al. 2011; Zhu et al. 2013). The major energy-providing component of rice grains is starch, a mixture of the two glucose-derived polysaccharides amylose and amylopectin. Amylose predominantly comprises linear chains of α (1,4) linked glucose residues, whereas amylopectin contains additional α (1,6) linked branches every 24-30 residues (Martin and Smith 1995). Starch from different plant species varies significantly in its physicochemical properties due to the relative proportions of amylose and amylopectin, and differences in chain length and/or amylopectin branching density (Jobling 2004).

Starch synthesis in plants involves the conversion of glucose 1-phosphate to ADP-glucose by the ATP-dependent enzyme glucose-1-phosphate adenylyltransferase (AGPase), and then the polymerization of ADP-glucose units by starch synthase to form amylose (**Figure 1**). Starch branching enzyme (SBE) introduces α (1,6) linked glycosidic bonds to generate amylopectin, a reaction that is reversed by the starch debranching enzyme isoamylase (DBE). In an alternative route, sucrose synthase in the endosperm cytosol converts

sucrose and ADP directly into fructose and ADP-glucose, and the latter is imported into the amyloplasts for starch synthesis (Li et al. 2013). Different forms of starch can be generated by mutating the genes encoding starch biosynthetic enzymes, but the outcome is complicated by the existence of multiple tissue-specific isoenzymes in many plants and the presence of multiple subunits per enzyme.

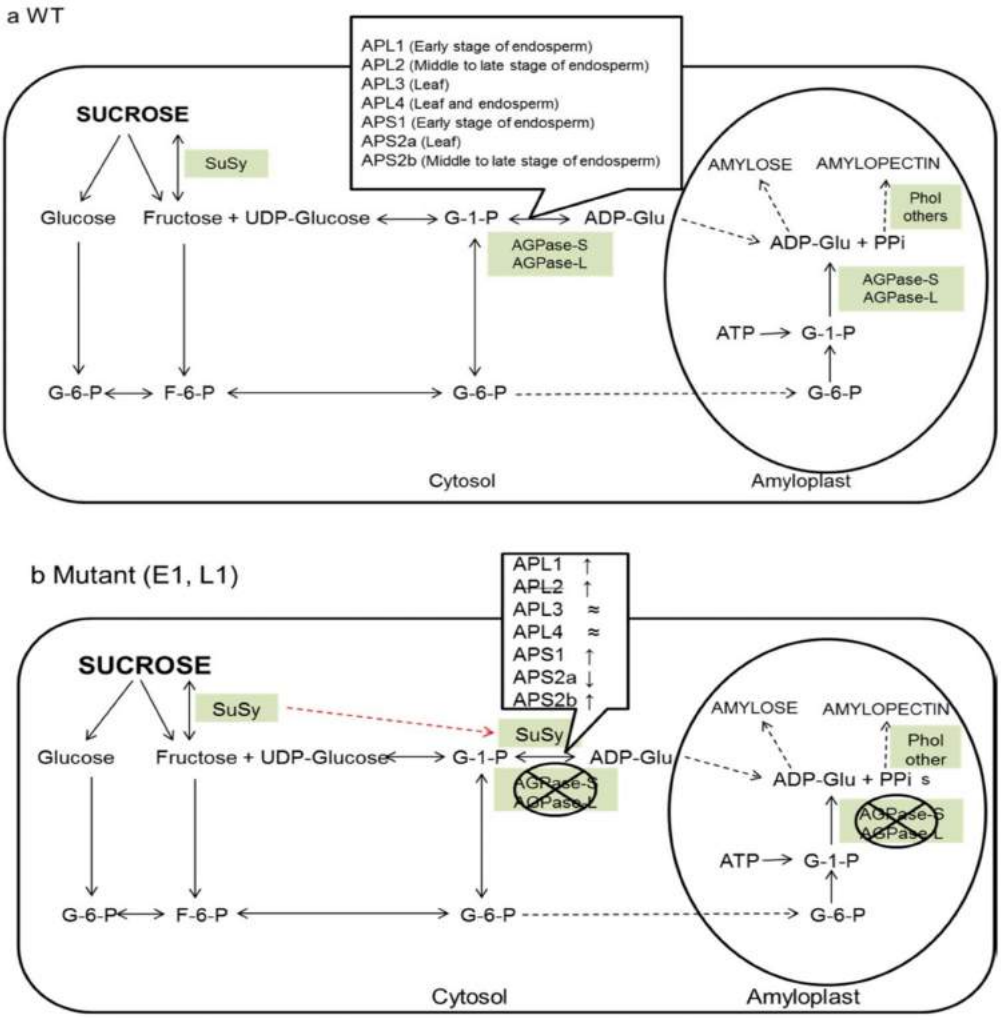


Figure 1 The coordination of different starch biosynthetic genes in rice (modified from Pandey et al. 2012). (a) Expression pattern of AGPase subunits in wild type. (b) Expression patterns of AGPase subunits in leaves in mutants L1 and E1 (↑ upregulated; ↓ downregulated; ≈ similar expression as wild type). The red dotted line represents the alternative pathway and the crossed out circles represent the loss of function of the enzyme. Abbreviations: SuSy (sucrose synthase), G-1-P (glucose-1-phosphate), G-6-P (glucose-6-phosphate), ADP-Glu (ADP-Glucose), PPi (inorganic diphosphate); F-6-P (fructose-6-phosphate); ATP (adenosine triphosphate); ADP (adenosine diphosphate), UDP (uridine diphosphate), PhoI (plastidial α-glucan phosphorylase); AGPase (ADP-glucose pyrophosphorylase); APS1 (AGPase small subunit 1); APS2a (AGPase small subunit 2a); APS2b (AGPase small subunit 2b); APL1 (AGPase large subunit 1); APL2 (AGPase large subunit 2); APL3 (AGPase large subunit 3); APL4 (AGPase large subunit 4).

Conventional mutagenesis such as irradiation, chemical mutagenesis and T-DNA/transposon insertional mutagenesis generate random lesions in DNA sequences and require the screening of large populations to isolate useful mutants. These techniques have been largely supplanted by targeted

mutagenesis using designer nucleases, particularly CRISPR/Cas9 (reviewed by Bortesi et al. 2016; Zhu et al. 2017). CRISPR/Cas9 mutagenesis is based on a bacterial defense system that targets invasive DNA by collecting DNA sequences as clustered regularly interspaced short palindromic repeats (CRISPRs) (Doudna et al. 2012). The transcription of these repeats into CRISPR RNAs, which pair with a CRISPR-associated (Cas) nuclease such as Cas9, allows the same invading DNA to be targeted and destroyed if it is encountered again (Lee et al. 2016). This mechanism can be harnessed for targeted mutagenesis if a synthetic guide RNA (sgRNA) is designed to match a genomic target instead of an invasive DNA sequence. The delivery of sgRNA and Cas9 to plant cells results in a double strand break (DSB) at the target site, which is generally repaired by the error-prone non-homologous end joining (NHEJ) pathway, resulting in small insertions or deletions at the site of the DSB that disrupt gene function by causing a frameshift mutation. The wild-type Cas9 generates a blunt DSB at the target site, which is specified by a 20-nt spacer sequence in the sgRNA. An alternative approach is to mutate one of the two endonuclease domains in Cas9 so that the enzyme only cleaves one DNA strand. A DSB therefore requires two such Cas9 nickases, and if these are recruited by different sgRNAs annealing a few base pairs apart, a staggered break is introduced within a 40-nt target sequence, significantly increasing the specificity of targeting and all but eliminating off-target cleavage activity (Fauser et al. 2014; Mikami et al. 2016; Ran et al. 2013).

High-amylose rice mutants have been produced by targeting the genes encoding SBEI and SBEIIb using CRISPR/Cas9 (Sun et al. 2017). Bi-allelic T0 mutants with insertions and deletions at the target sites were generated and the mutations were stably transmitted to progeny. The *OsBEIIb* mutants accumulate higher proportion of amylose and debranched amylopectin in the seeds than normal (Sun et al. 2017). *OsBEIIb* has also been targeted using two different sgRNAs with different activity scores and different degrees of conservation with the paralogous gene *OsBEIIa* to confirm the absence of off-target mutations (Baysal et al. 2016). Another study using wild-type Cas9 targeted three different sites in *OsWaxy* encoding granule-bound starch synthase (GBSS). Only one or two of the sites were mutated in the resulting primary transformant, but the amylose content in T1 seeds was reduced from 14.6% to 2.6% (Ma et al. 2015).

Although the CRISPR/Cas9 system has been used to target GBSS and SBE, it has not yet been used to target AGPase, which catalyzes the first step in the starch biosynthesis pathway (Tang et al. 2016; Lee et al. 2016). Rather than altering the ratio of amylose and amylopectin, blocking AGPase would therefore prevent starch synthesis all together. AGPase comprises two large and two small subunits that together form a hetero-tetrameric complex (Tuncel et al. 2014; Ballicora 2004). In rice, there are two small subunit genes (*OsAPS1* and *OsAPS2*, the latter producing the two mRNA variants *OsAPS2a* and *OsAPS2b* by alternative splicing) and four large subunit genes (*OsAPL1*, *OsAPL2, OsAPL3 and OsAPL4*). AGPase is the key enzyme for starch synthesis. Its regulation is controlled by 3-phosphoglycerate (3-PGA) which upregulates its activity and by inorganic phosphate (Pi) which downregulates AGPase (Preiss 1982). All transcripts are tissue-specific and the corresponding proteins are localized differentially (Ohdan et al. 2005). *OsAPS2a* is mainly expressed in the leaves and the protein is localized in plastids, whereas *OsAPS2b* is only expressed in the endosperm and

OsAPL2 is mainly expressed in the endosperm and the proteins are localized in the cytosol. *OsAPL1* and *OsAPS1* are mostly expressed in early endosperm plastids, whereas *OsAPL3* is expressed in leaf plastids. *OsAPL4* is expressed at high levels in leaf plastids but at low levels in endosperm plastids (Ohdan et al. 2005; Lee et al. 2007). Because *OsAPL2* and *OsAPS2b* are the only cytosolic subunits and *OsAPS2b* is expressed exclusively in endosperm there is no cytosolic AGPase in the leaves. Mutations in *OsAPL2* and *OsAPS2b* cause a marked reduction in starch levels (Lee et al. 2007; Ohdan et al. 2005; Tsai and Nelson 1966; Tester et al. 1993; Johnson 2003; Muller et al. 1992; Giroux 1994, Tang et al. 2016). In the rice *OsAPL1* mutant, the starch content in the leaves was reduced to ~5% of wild-type levels, but growth and development were normal (Rosti et al. 2007). Homozygous *OsAPL3* mutants displayed ~23% of wild-type AGPase activity and accumulated much less starch than normal in the culm (Cook et al. 2012). The shrunken mutant has a nonfunctional *OsAPS2* and therefore lacks both the *OsAPS2a* and *OsAPS2b* transcripts, and exhibits ~20% of wild-type AGPase activity (Kawagoe et al. 2005; Tuncel et al. 2014). Mutation of *OsAPS2b* caused *OSAPS1* and *OsAPL1* transcript levels to increase in leaves whereas *OsAPL3* transcript levels remained unaffected (Ohdan et al. 2005).

The absence of a cytosolic AGPase in leaves suggests that manipulation of *OsAPL2* and *OsAPS2b* might offer a strategy to modulate starch production in the endosperm without affecting starch metabolism in vegetative tissues. Our experiments focused on *OsAPL2* because that is the only AGPase subunit gene that is largely expressed in the endosperm. *OsAPL2* encodes a 518 amino acid polypeptide with a catalytic site spanning residues 88–364. The catalytic site is composed of α-helices and β-sheets that form a substrate binding cleft. A further α-helix and β-sheet outside the catalytic center are required to maintain the correct tertiary conformation and bind with the other subunit, APS2b, to form the tetrameric structure (**Figure 2** and **Figure 3**) (Zhang et al. 2012). Experiments with null and missense *OsAPL2* mutants indicate that this subunit plays an important role in both the catalytic and regulatory properties of AGPase (Tuncel et al. 2014). We hypothesized that knocking out *OsAPL2* would have no effect on starch levels in leaves because its expression is very low in this tissue. We mutated the *OsAPL2* using two different strategies. We targeted the first exon aiming to truncate the enzyme substantially and abolish its activity entirely. For this purpose, we used Cas9 nickase with two closely-spaced targets in order to generate a deletion. In separate experiments, we targeted an exon downstream of the active site, to maintain some activity but perturb the tertiary structure. In this case we used the wild-type Cas9 with a single target in order to generate indels. We anticipated a reduction in endosperm starch levels in both lines and a corresponding increase in the abundance of soluble sugars, but we anticipated no impact on starch metabolism in leaves.

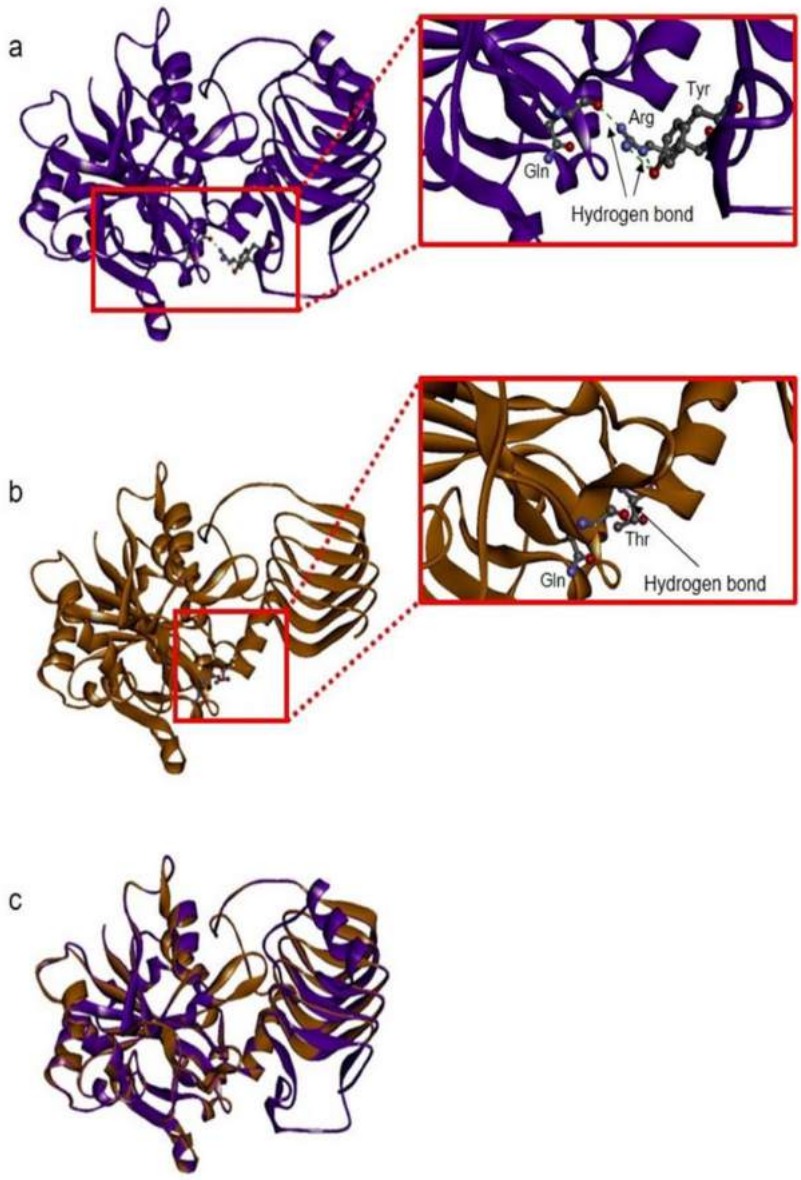


Figure 2 (a) 3-D model of wild-type APL2 and zoom of the homology modeled structure where hydrogen bonding between Gln 98 and Arg 500 and Tyr 498 is shown as a dashed green line. (b) 3-D model of the mutated ADP-glucose pyrophosphorylase (AGPase) in L1 and zoom of the homology modeled structure in the L1 mutant where the hydrogen bond between Gln 98 and Thr 351 is shown. (c) Superimposed wild-type (in purple) and mutated protein in L1 (in orange). The mutated *APL2* sequences were translated into polypeptides (http://web.expasy.org/translate/) and automated homology modeling was carried out using SWISS-MODEL (https://swissmodel.expasy.org/interactive/wrGgpx/models) with the potato tuber AGPase SS (Protein Databank: 1yp4) as the template. The model of the mutant protein was superimposed on the wild-type version using DS Visualizer (http:/accelrys.com/products/collaborative-science/biovia-discovery-studio/visualization.html).

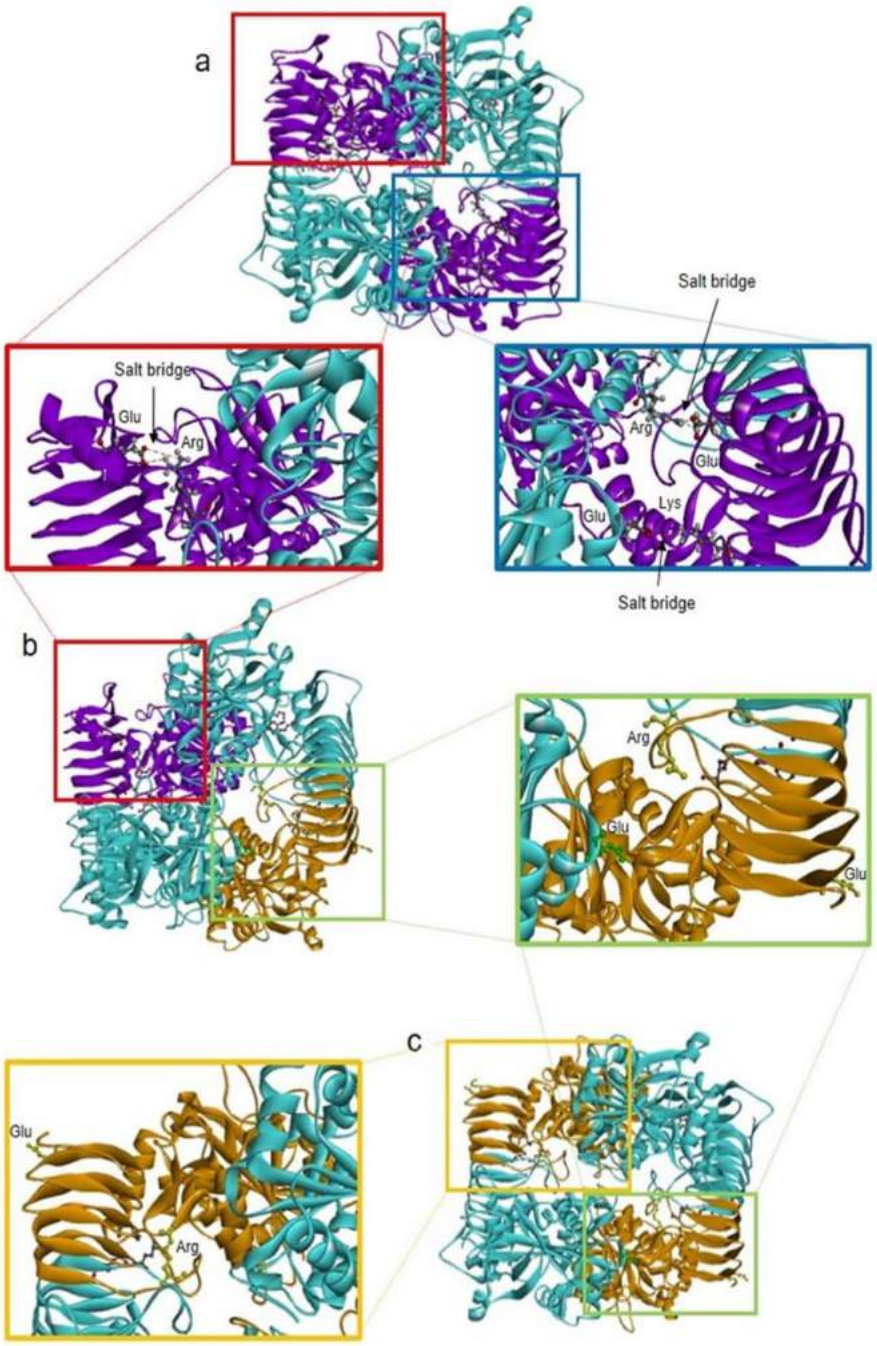


Figure 3 (a) Heterotetrameric structure of wild type AGPase that consists of two large subunits (APL2, in purple) and two small subunits (APS2b, in blue). Salt bridge interaction (between oppositely charged residues that are sufficiently close to each other for electrostatic attraction) (shown in blue square) between Lys 508 (Chain C) and Glu 161(Chain B) and between Glu 444 (Chain C) and Arg 335 (Chain D). Salt bridge interaction (shown in red squared) between Glu 444 (Chain A) and Arg 335 (Chain B). (b) Heterotetrameric structure of AGPase that consists of a wild type large subunit (APL2, in purple), a mutated large subunit (L1, in orange) and two small subunits (APS2b, in blue). Salt bridge interaction (shown in red squared) between Glu 444 (Chain A) and Arg 335 (Chain B) and (shown in green squared) between Arg 335 (Chain D, in yellow), Glu 444 (Chain C, in yellow) and Glu 161 (Chain B, in green). (c) Heterotetrameric structure of AGPase consisting of two mutated large subunits (L1, in orange) and two small subunits (APS2b, in blue). Salt bridge interaction (shown in orange squared) between Arg 335 (Chain B, in yellow) and Glu 444 (Chain A, in yellow) and (shown in green squared) between Arg 335 (Chain D, in yellow), Glu 444 (Chain C, in yellow) and the Glu 161 (Chain B, in green).

Materials and methods

Target sites and sgRNA design

Target sites for wild-type Cas9 (single sgRNA) and Cas9D10A (two sgRNAs targeting adjacent sites) were selected within the *OsAPL2* coding sequence (GenBank AK071497.1) using E-CRISP (Heigwer et al. 2014) with the following parameters: only NGG PAM, only G as 5' base, off-target tolerates many mismatches, non-seed region ignored, introns ignored. The sgRNAs (**Figure 4a**) catalytic efficiencies were designed to minimize off-targets. The catalytic efficiency of the sgRNAs was predicted using gRNA scorer (Chari et al. 2015).

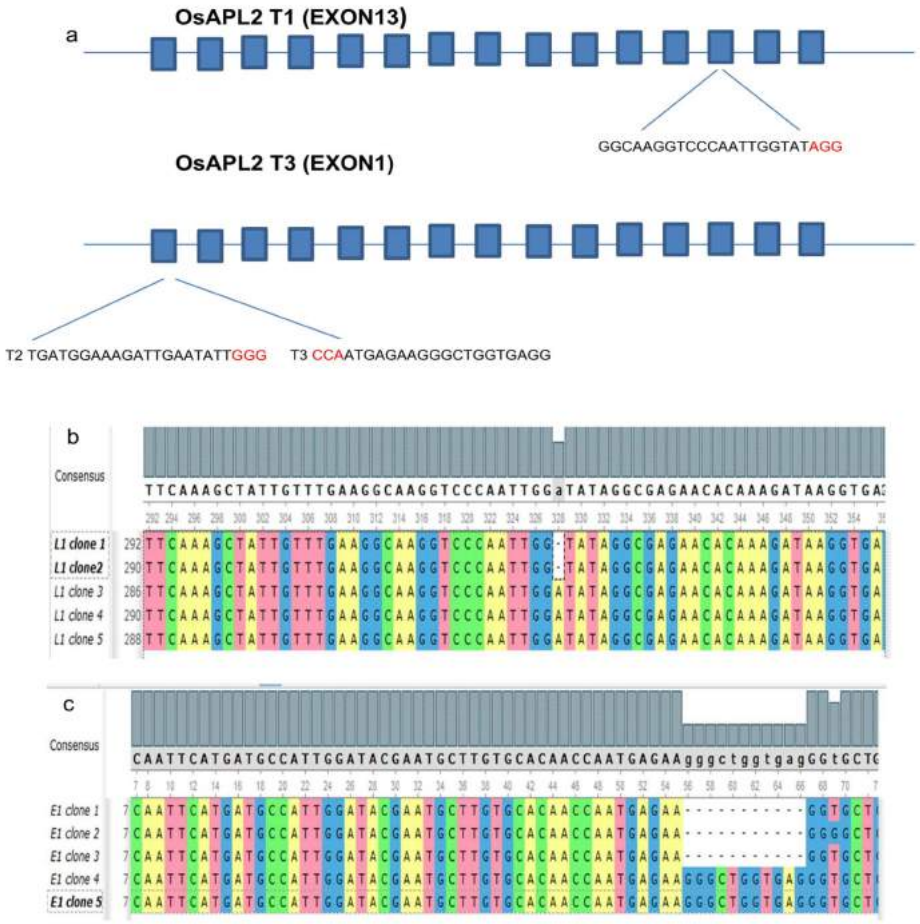


Figure 4 gRNA sites and sequencing results in the mutants. (A) Schematic representation of the gRNA target sites of the genomic region of the *OsAPL2* gene. Introns are represented by lines and exons by blue boxes. Expanded region corresponds to the target sites. PAM sequence is shown in red. (B) Sequencing results of the *OsAPL2* mutant induced by wild-type Cas9 (L1). (C) *OsAPL2* mutant induced by Cas9D10A (E1). The target site of the wild-type protein is shown as a dotted line. Clones in bold represent the wt sequence

Vector construction

The wild-type Cas9 vector pJIT163-2NSCas9 and the sgRNA vector pU3-gRNA were obtained from Dr. C. Gao, Chinese Academy of Sciences, Beijing, China (Shan et al. 2013). The nickase vector pJIT163-2NSCas9D10A was constructed in-house by mutating the *cas9* gene in vector pJIT163-2NSCas9 to produce Cas9D10A and combining this with the maize *ubiquitin*-1 promoter and Cauliflower mosaic virus 35*S* terminator. The three sgRNAs described above were prepared as synthetic double-stranded oligonucleotides and introduced separately into pU3-gRNA at the AarI restriction site; thus all genomic sites with the form 5'-(20)-NGG-3' can be targeted. The *hpt* selectable marker gene was provided on a separate vector as previously described (Christou et al. 1991).

Rice transformation and recovery of transgenic plants

Seven-day-old mature zygotic embryos (*Oryza sativa* cv. EYI) were transferred to osmotic medium (MS medium supplemented with 0.3 g/L casein hydrolysate, 0.5 g/L proline, 72.8 g/L mannitol and 30 g/L sucrose) 4 h before bombardment with 10 mg gold particles coated with the transformation vectors. The Cas9 vector (wild type or nickase), the corresponding sgRNA vector(s) and the selectable marker *hpt* were introduced at a 3:3:1 ratio for wild-type Cas9 and a 3:3:3:1 ratio for the nickase with two sgRNAs (Christou et al. 1991; Sudhakar al. 1998; Valdez et al. 1998). The embryos were returned to osmotic medium for 12 h before selection on MS medium (MS medium supplemented with 0.3 g/L casein, 0.5 g/L proline and 30 g/L sucrose) with 50 mg/L hygromycin and 2.5 mg/l 2,4-dichlorophenoxyacetic acid in the dark for 2–3 weeks. Callus was maintained on selective medium for 6 weeks with sub culturing every 2 weeks as described (Farré et al. 2012). Transgenic plantlets were regenerated and hardened off in soil. Negative controls were regenerated plants from the same experiment which were no transformed (i.e. they did not contain Cas9WT/D10A, hpt and sgRNA)., Negative controls behaved exactly like wild type plants.

Confirmation of the presence of Cas9 and gRNA

Genomic DNA was isolated from callus . leaves and panicles of regenerated plants by phenol extraction and ethanol precipitation (Bassie et al., 2008; Kang and Yang, 2004). The presence of the wild-type Cas9 sequence was confirmed by PCR using primers 5'-GTC CGA TAA TGT GCC CAG CGA-3' and 5'-GAA ATC CCT CCC CTT GTC CCA-3'; the presence of the Cas9D10A sequence was determined using primers 5'-GCA AAG AAC TTT CGA TAA CGG CAG CAT CCC TCA CC-3' and 5'-CCT TCA CTT CCC GGA TCA GCT TGT CAT TCT CAT CGT-3'; and the presence of the pU3-gRNA vectors was confirmed using the conserved primers 5'-TTG GGT AAC GCC AGG GTT TT-3' and 5'-TGT GGA TAG CCG AGG TGG TA-3'.

Analysis of induced mutations

The *OsAPL2* mutation induced by the wild-type Cas9 was detected by PCR using primers 5′-CGT TAG CAT CGG GTG TGA ACT-3′ and 5′-GGA CCC CCT ATC ATA CGC AGT-3′. The *OsAPL2* mutation induced by Cas9D10A was similarly detected using primers 5′-CTT GTT GTT CAG GAT GGA TGC-3′ and 5′-GTG CAT TGT GCC TGT GGA A-3′. The PCR products were sequenced using an ABI 3730xl DNA analyzer by Stabvida (http://www.stabvida.com/es/). To confirm the mutations, PCR products generated using the primers listed above were purified using the Geneclean® II Kit (MP Biomedicals), transferred to the pGEM-T Easy vector (Promega) and introduced into competent *Escherichia coli* cells. From *E. coli*, PCR fragments of ~400 bp were purified, cloned and sequenced using an ABI 3730xl DNA analyzer by Stabvida. At least five clones per PCR product were sequenced using primerM13Fwd (**Table 1**).

Table 1. Primer sequences for RT-qPCR and sequencing analysis

RT-qPCR for expression pattern analysis	APS1-RT-Forward	5′-GTGCCACTTAAAGGCACCATT-3′
	APS1-RT-Reverse	5′-CCCACATTTCAGACACGGTTT-3′
	APS2a-RT- Forward	5′-ACTCCAAGAGCTCGCAGACC-3′
	APS2a-RT- Reverse	5′-GCCTGTAGTTGGCACCCAGA-3′
	APS2b-RT- Forward	5′-AACAATCGAAGCGCGAGAAA-3′
	APS2b-RT- Reverse	5′-GCCTGTAGTTGGCACCCAGA-3′
	APL1-RT- Forward	5′-GGAAGACGGATGATCGAGAAAG-3′
	APL1-RT- Reverse	5′-CACATGAGATGCACCAACGA-3′
	APL2-RT- Forward	5′-AGTTCGATTCAAGACGGATAGC-3′
	APL2-RT- Reverse	5′-CGACTTCCACAGGCAGCTTATT-3′
	APL3-RT- Forward	5′-AAGCCAGCCATGACCATTTG-3′
	APL3-RT- Reverse	5′-CACACGGTAGATTCACGAGACAA-3′
	APL4-RT- Forward	5′-TCAACGTCGATGCAGCAAAT-3′
	APL4-RT- Reverse	5′-ATCCCTCAGTTCCTAGCCTCATT-3′
	Pho1-RT- Forward	5′-TTGGCAGGAAGGTTTCGCT-3′
	Pho1-RT- Reverse	5′-CGAAGCCTGAAGTGAACTTGCT-3′
	Ubi- RT- Forward	5′-ACCACTTCGACCGCCACTACT-3′
	Ubi –RT- Reverse	5′-ACGCCTAAGCCTGCTGGTT-3′
	APL2-end-RT-Forward	5’-GTGATCATTGCAAACACTCAGG-3’
	APL2-end-RT-Reverse	5’-GGATCACCACAATTCCAGACC-3’
Sequencing analysis	M13 Fwd	5’-TGTAAAACGACGGCCAGT-3’

Protein structural modelling

The mutated *OsAPL2* sequences were translated into polypeptides (http://web.expasy.org/translate/) and automated homology modeling was carried out using SWISS-MODEL (https://swissmodel.expasy.org/interactive/wrGgpx/models) with the potato tuber AGPase SS (Protein Databank: 1yp4) as the template. The model of the mutant protein was superimposed on the wild-type version using DS Visualizer (http:/accelrys.com/products/collaborative-science/biovia-discovery-studio/visualization.html). Heterotetrameric AGPase structures were energy minimized using DS Visualizer.

Enzymatic activity and carbohydrate levels

Leaf extracts were prepared as previously reported (Tang et al. 2016) for the measurement of AGPase activity (ADP-glucose pyrophosphorylase, E.C. 2.7.7.27) in the forward direction (Nishi et al. 2001), and sucrose synthase activity (UDP-Glc:D-fructose 2-glucosyltransferase, EC 2.4.1.13) as previously described (Doehlert et al. 1988). Flag leaf samples harvested at 19:00 hours were homogenized under liquid nitrogen and extracted in perchloric acid to measure the starch content, or in ethanol to measure the soluble sugar content. The quantity of each carbohydrate was determined by spectrophotometry (Yoshida et al. 1976).

RNA extraction and real-time qRT-PCR analysis

Total leaf RNA was isolated using the RNeasy Plant Mini Kit (Qiagen) and DNA was digested with DNase I (RNase-free DNase Set, Qiagen). Total RNA was quantified using a Nanodrop 1000 spectrophotometer (Thermo Fisher Scientific) and 2 µg total RNA was used as template for first strand cDNA synthesis with Quantitech® reverse transcriptase (Qiagen) in a 20-µl total reaction volume, following the manufacturer's recommendations. Real-time qRT-PCR was performed on a BioRad CFX96™ system using 20-µl mixtures containing 5 ng synthesized cDNA, 1x iQ SYBR green supermix and 0.5 µM forward and reverse primers. The *OsAPL1*, *OsAPL3*, *OsAPL4*, *OsAPS1*, *OsAPS2a/b, OsAPL2* and *OsPho1* transcripts were amplified using the primers listed in **Table 1**, as described by Tang et al. Primers at the end of *OsAPL2* were designed to amplify the non-common region between the WT and mutants E1 and L1 (**Table 1**). (2016). Serial dilutions of cDNA (80–0.0256 ng) were used to generate standard curves for each gene. PCR was performed in triplicate using 96-well optical reaction plates. Values represent the mean of three biological replicates ± SE. Amplification efficiencies were compared by plotting the ΔCt values of different primer combinations of serial dilutions against the log of starting template concentrations using the CFX96™ software. The rice housekeeping *OsUBQ5* (LOC_Os01g22490) was used as an internal control.

1.3 Results

Design of a CRISPR/Cas9 mutation strategy

We designed three sgRNAs targeting the *OsAPL2* and selected the target sequences using E-CRISP to minimize the likelihood of off-targets. For the wild-type Cas9, we selected a single target site in exon 13 (hereafter named T1), whereas for Cas9D10A nickase we selected two adjacent targets in the first exon (hereafter named T2 and T3). The locations of each target are shown in **Figure 4a**. The sgRNA cassettes were separately transferred to the pU3-gRNA vector and introduced into rice embryos along with the rice codon-optimized wild-type Cas9 or Cas9D10A sequence, and the selectable marker *hpt* conferring hygromycin resistance.

Recovery and analysis of mutant lines

We regenerated transgenic plants representing each transformation strategy. Sequencing *E. coli* colonies revealed that the mutation generated by wild-type Cas9 in an early transformant (mutant L1) was an insertion of a single nucleotide at site T1 (**Figure 4b**) whereas a mutation generated by the Cas9D10A nickase (mutant E1) in another transformant was a deletion of 11 nucleotides at sites T2/T3 (**Figure 4c**). We focused on heterozygous mutations in order to determine whether the expression of the corresponding wild-type allele was affected in regenerated plants (see below). As well as testing for on-target mutations, E-CRISP identified potential off-target cleavage sites at three loci based on the number of mismatches allowed in the target sequence and 2 bp upstream of the DSB. A single potential off-target was identified for mutant L1 in the *OsAGPlar* (EU267957.1) whereas two potential off-targets were identified for mutant E1 in the *OsXPO7* (LOC4334606) and the *OsNAT6* (LOC4330689). Sequencing these loci revealed no evidence of off-target mutations.

Structural comparisons

In order to investigate potential changes at the protein level, we translated the L1 and E1 mutant *OsAPL2* sequences and generated three-dimensional models using the SWISS-MODEL program. Compared to the wild-type sequence (**Figure 5a**), the E1 mutation resulted in an early stop codon such that the residual product was only 21 amino acids in length (**Figure 5c**). In contrast, the insertion of a single nucleotide in mutant L1 generated a change in the protein sequence downstream of the catalytic site (**Figure 5b**). Clearly, the E1 mutation generated a non-functional product. However, much of the L1 protein sequence was preserved and by superimposing the mutant sequence over that of the wild-type protein we found that the L1 mutant also

contained a variant loop structure in the same vicinity. The resulting changes to the active site are shown in **Figure 2**. The key difference is the change in orientation of the Gln 98 side chain relative to the active site, which affects the topology of the substrate-binding cleft and influences substrate accommodation (Tuncel et al., (2014) and Tang et al., (2016)) (**Figure 2b**). The change occurs because hydrogen bonding between residue Arg 500 and Tyr 498 is eliminated (**Figure 2a**) and a new hydrogen bond is formed with Thr 351 (**Figure 2b**). The heterotetrameric structure of the AGPase comprises two APS2b and two APL2 subunits (**Figure 3a**). The four subunits of the wild type heterotetramer are stabilized by three salt bridges between Lys 508 (Chain C)-Glu161 (Chain B), Glu 444 (Chain C)-Arg 335 (Chain D) and Glu 444 (Chain A)-Arg 335(Chain B), that are eliminated in the mutated forms.

a WT protein sequence

MQFMMPLDTNACAQPMRRAGEGAGTERLMERLNIGGMTQEKALRKRCFGDGVTGTARCV
FTSDADRDTPHLRTQSSRKNYADASHVSAVILGGGTGVQLFPLTSTRATPAVPVGGCYRLIDI
PMSNCFNSGINKIFVMTQFNSASLNRHIHHTYLGGGINFTDGSVQVLAATQMPDEPAGWFQ
GTADAIRKFMWILEDHYNQNNIEHVVILCGDQLYRMNYMELVQKHVDDNADITISCAPIDGSR
ASDYGLVKFDDSGRVIQFLEKPEGADLESMKVDTSFLSYAIDDKQKYPYIASMGIYVLKKDVL
LDILKSKYAHLQDFGSEILPRAVLEHNVKACVFTEYWEDIGTIKSFFDANLALTEQPPKFEFYD
PKTPFFTSPRYLPPARLEKCKIKDAIISDGCSFSECTIEHSVIGISSRVSIGCELKDTMMMGAD
QYETEEETSKLLFEGKVPIG**IGENTKIRNCIIDMNARIGRNVIIANTQGVQESDHPEEGYYIRSG**
IVVILKNATIKDGTVI

b Mutated L1 protein sequence

MQFMMPLDTNACAQPMRRAGEGAGTERLMERLNIGGMTQEKALRKRCFGDGVTGTARCV
FTSDADRDTPHLRTQSSRKNYADASHVSAVILGGGTGVQLFPLTSTRATPAVPVGGCYRLIDI
PMSNCFNSGINKIFVMTQFNSASLNRHIHHTYLGGGINFTDGSVQVLAATQMPDEPAGWFQ
GTADAIRKFMWILEDHYNQNNIEHVVILCGDQLYRMNYMELVQKHVDDNADITISCAPIDGSR
ASDYGLVKFDDSGRVIQFLEKPEGADLESMKVDTSFLSYAIDDKQKYPYIASMGIYVLKKDVL
LDILKSKYAHLQDFGSEILPRAVLEHNVKACVFTEYWEDIGTIKSFFDANLALTEQPPKFEFYD
PKTPFFTSPRYLPPARLEKCKIKDAIISDGCSFSECTIEHSVIGISSRVSIGCELKDTMMMGAD
QYETEEETSKLLFEGKVPIG**YRREHKDKELYHRHEC**

c Mutated E1 protein sequence

MQFMMPLDTNACAQPMRRCWD**Stop**

Figure 5 Expected protein sequences using ExPASy. (A) Wild type OsAPL2 protein sequence; (B) OsAPL2 sequence in L1 mutant, with the changed amino acids in bold; (C) OsAPL2 sequence in E1 mutant.

Analysis of AGPase and sucrose synthase activity

We determined the direct impact of each mutation on AGPase activity by comparing enzyme activity in the flag leaves of the mutant and wild-type plants as described by Nishi et al. (2001), using three biological replicates and two technical replicates for each biological sample (**Figure 6a**). No analysis was possible in seed as both mutants were infertile. AGPase activity in mutant E1 was low (10.6 ± 1.8 mmol NADPH/min/mg FW) whereas activity in mutant L1 was much higher (30.7 ± 1.4 nmol NADPH/min/mg FW) than wild-type plants (15.5 ± 1.8 nmol NADPH/min/mg FW). AGPase activity in the corresponding negative controls was similar to

wild-type levels, as expected. We also measured the sucrose synthase activity in each mutant because this provides an alternative route for the synthesis of ADP-glucose, which might be induced when the primary pathway is blocked (**Figure 6b**). Sucrose synthase activity in the E1 mutant was higher (3.5 ± 0.27 nmol sucrose/min/mg FW) whereas the L1 mutant had similar activity to the wild-type plants and negative controls (1.4–2.1 nmol sucrose /min/mg FW).

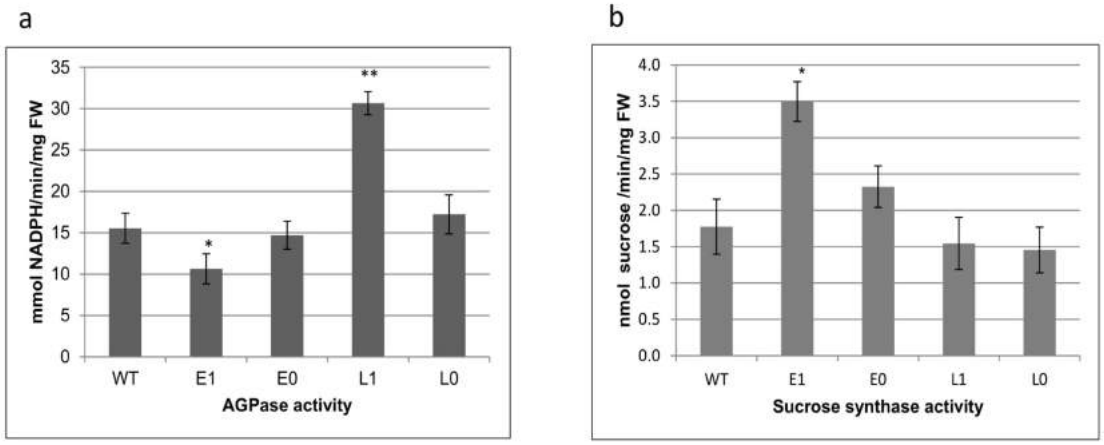


Figure 6 AGPase activity and (b) sucrose synthase activity in the flag leaves of wild-type (WT), mutants, and negative controls. E0 is the negative control (negative transformants regenerated at the same time as the mutated lines) for E1; L0 is the corresponding negative control for L1. Values are means ± SDs (n = 3 biological replicates, 2 technical replicates for each biological replicate). The asterisk indicates a statistically significant difference between WT and mutant, as determined by a Student's t-test (*P < 0.05, **P<0.001).

Analysis of AGPase family gene expression

Next, we measured the expression of Os*APL2* and transcripts of other starch biosynthetic genes (*OsAPL1*, *OsAPL3*, *OsAPL4*, *OsAPS1*, *OsAPS2a*, *OsAPS2b* and *OsPhoI*) in the flag leaves to determine whether mutating the *OsAPL2* gene had an indirect regulatory impact on genes encoding other subunits. Our results indicated that the relative expression levels of *OsAPS2b* and *OsAPL2* increased significantly in both mutants compared to wild-type plants. This was surprising given that *OsAPS2b* expresses mostly in the endosperm, *OsAPL2* is expressed at low levels in leaves and the tetramer between APL2-APS2b is not usually expressed at significant levels in WT leaves. Mutating *OsAPL2* therefore appears to promote the ectopic expression of a cytosolic AGPase which is not typically present in leaves. In contrast, *OsAPS2a* expression levels declined significantly in both mutants compared to wild-type plants, whereas the expression of *OsAPL1* and *OsPhoI* was significantly higher than wild-type levels in both mutants (**Figure 7**). The expression of *OsAPS1* was 9 times higher than wild-type levels in mutant E1, and 2 times higher in mutant L1 (**Figure 7b**). *OsAPL3* and *OsAPL4* expression remained similar to wild-type levels in both mutants.

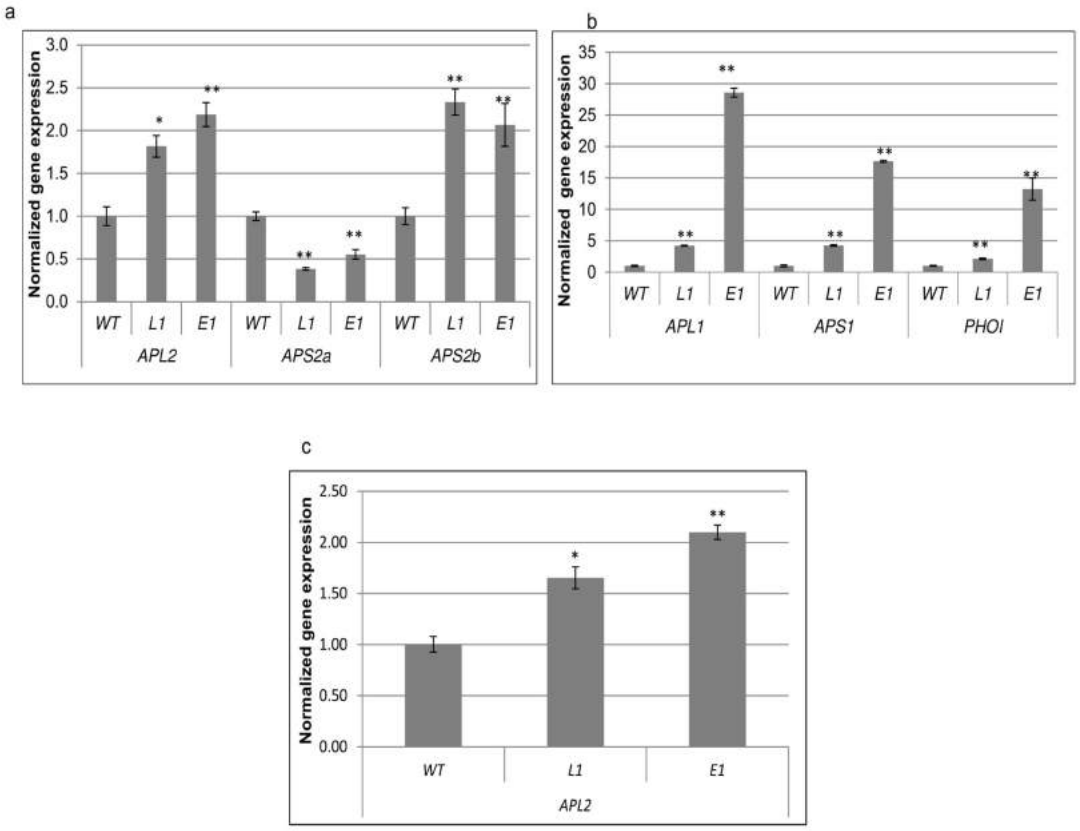


Figure 7. Relative expression levels of rice AGPase family genes in the flag leaves of wild-type and mutants. (a) *OsAPL2*, *OsAPS2a* and *OsAPS2b*. (b) *OsAPL1*, *OsAPS1* and *OsPHOI*. (c) *OsAPL2* amplifying C-terminal end of the protein (non-common between WT and mutants). Values are means ± SDs (n = 3 technical replicates). The asterisks indicate a statistically significant difference between wild-type (WT) and mutants, as determined by a Student's t-test (*P < 0.05; **P < 0.01).

OsAPL2 has three alleles in the genome and the E1 and L1 mutants are heterozygous (at least one of the alleles is mutated while the remaining is/are WT). To determine which allele (WT or mutated) is responsible for the increase in *OsAPL2* expression in leaves, we designed primers at the end of the gene in the non-common region between the WT and mutants E1 and L1 to only amplify WT allele(s). We measured expression levels in the flag leaves. *APL2* expression in E1 remained constant irrespective of using non-common or common region primers, suggesting that the WT allele was responsible for the increased expression of the gene. L1 retained 90% of the expression of *OsAPL2* using non- common region primers suggesting that the increase in *OsAPL2* expression was also due to the WT allele (**Figure. 7c**).

Analysis of starch and sugar levels

A significant decline in starch content was observed in the leaves of both mutants. E1 showed a remarkable decrease in starch content to 2% by weight, ~15% of the normal levels found in wild-type leaves. L1 accumulated 7% starch by weight, approximately half the normal level. The negative controls had similar levels of starch as the wild-type plants (**Figure 8a**). In contrast, the soluble sugar content of both mutants was

higher than wild-type levels: ~40% higher in E1 and ~20% higher in L1. The negative controls contained similar amounts of soluble sugars to the wild-type plants (**Figure 8b**). These results are surprising given that the *OsAPL2* mutation appeared to promote the ectopic expression of a cytosolic AGPase in leaves that is normally expressed at very low levels in wild-type plants, indicating that starch synthesis in leaves remains dependent on the plastidial AGPase even when cytosolic large and small subunits are expressed.

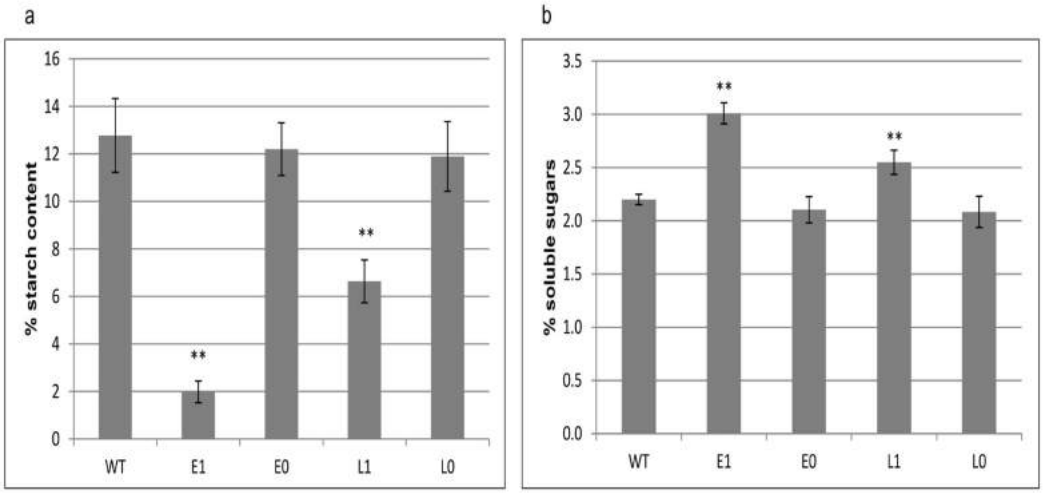


Figure 8. Starch and soluble sugar content in flag leaves. (a) Starch and (b) soluble sugar content in mutants (E1 and L1), corresponding negative controls (E0 and L0) and wild type. Values are expressed as means ± SDs (n = 3 biological replicates, 2 technical replicates for each biological). The asterisks indicate a statistically significant difference between wild-type (WT) and mutant, as determined by a Student's t-test (*P < 0.05; **P < 0.01).

1.4 Discussion

Amylose is synthesized by AGPase and GBSS whereas amylopectin also requires SBE and DBE to introduce and refine its branching structure (Ohdan et al. 2005). The heterotetrameric AGPase catalyzes the first committed step in starch synthesis. In rice there are two genes encoding small catalytic subunits (*APS1* and *APS2*) and four encoding larger regulatory subunits (*APL1*, *APL2*, *APL3* and *APL4*). Furthermore, *OsAPS2* produces two distinct polypeptides through alternative splicing: *APS2a* and *APS2b*, the former including a transit peptide for import into the plastids and the latter lacking this sequence causing it to remain in the cytosol (Tang et al. 2016). AGPases are key enzymes in the starch biosynthesis pathway and are regulated by the ratio 3-PGA/Pi (Preiss et al. 1982). Among the larger regulatory subunits, only APL2 lacks a transit peptide and remains in the cytosol, whereas the other three subunits are imported into the plastid. Thus, only one AGPase assembles in the cytosol, comprising subunits APL2 and APS2b.

APS1 and *APL1* are expressed strongly in the early endosperm, whereas expression of *APS2b* and *APL2* begins 3 days after fertilization and remains at high levels thereafter. *APL3* is expressed at low levels throughout seed development, and the *APL4* and *APS2a* transcripts are barely detected at all (Ohdan et al. 2005). This suggests that the endosperm contains one predominant plastid AGPase, comprising APL1 and APS1, which is important during early development, and one predominant cytosolic AGPase, comprising APL2 and APS2b, which is important during the middle and late stages of development, when starch accumulates (Ohdan et al. 2005). During seed development, starch is accumulated in amyloplasts which serve as a reservoir for the germinating seed. The major tetramer (APL2-APS2b) is in the cytosol and acts to generate stable (i.e. not transitory) starch in this compartment (Rychter AM and Rao IM 2005). In leaf, starch accumulation is transitory in chloroplasts to generate photosynthetic precursors. The starch deposited in chloroplasts is degraded during the night and resulting G-1-P is converted to triose-P and exchanged for Pi from the cytosol (Heldt et al., 1977; Stitt et al., 1981; Dennis et al., 1982; Lee SK et al., 2016)). Pi is an activator of photosynthetic enzymes, including Rubisco (Heldt et al., 1978; Bhagwat 1981). Thus, in leaves the main AGPases are plastidial. *APS2a* and *APL3* are strongly expressed in young leaves, whereas *APS1* is expressed at low levels, *APS2b* is not detected at all, and the remaining subunits (*APL1*, *APL2* and *APL4*) are minimally expressed. Later in development, *APS2a* and *APL3* expression declines (although they are still the most abundant subunits) and *APL1* expression increases to parity with *APL3*. This suggests that the major leaf AGPase initially comprises subunits *APS2a* and *APL3*, with *APS1* and *APL3* combining to form a less abundant subunit, but the increase in *APS1* expression may result in a progressive accumulation of the *APS1/APL3* heterotetramer later in development. Importantly, all the AGPases in the leaf appear to be located in the plastid given that no cytosolic forms are expressed at any developmental stage [*APS2b* is not detected in leaves (Ohdan et al. 2005)].

In addition to the enzymes discussed above, phosphorylase is involved in starch degradation, but earlier studies suggest it may also play a role in starch biosynthesis (Satoh et al. 2008; Steup 1990; Yu et al. 2001). There are two types of phosphorylase that differ in terms of structure, kinetics, expression patterns and subcellular location (Satoh et al. 2008). PhoI is located in plastids and appears to facilitate starch biosynthesis in storage tissues, given that it is expressed at a higher level in the endosperm than in leaves (Colleoni et al. 1999; Ball and Morell 2003; Tetlow et al. 2004; Ohdan et al. 2005) and is essential for starch synthesis and accumulation (Shimomura et al. 1982; Steup 1990; Satoh et al. 2008). In contrast, PhoH is located in the cytosol and does not play a role in starch biosynthesis. Sucrose synthase (SuSy) catalyzes the conversion of UDP and sucrose into fructose and UDP-glucose in the cytosol (Li et al. 2013). The carbon atoms in starch are derived from fructose-6-phosphate in photosynthetic tissues (**Figure 1**), but from cytosolic sucrose converted to UDP-glucose in other tissues, followed by translocation to the amyloplasts (Nakamura 2015).

Based on the expression profiles of the rice AGPase subunits, we hypothesized that mutating *OsAPL2* (and thus removing the large regulatory subunit that is mainly expressed in the endosperm) should have no

impact on starch biosynthesis in leaves. Interestingly, the rice *shrunken* mutant is deficient in AGPase due to the loss of subunit APS2b (the endosperm cytosolic small subunit that assembles with APL2), resulting in the loss of ~80% of wild-type AGPase activity in endosperm. Remarkably, the loss of *APS2b* enhances the expression of several alternative subunits in the endosperm and, in some cases, also in the leaves even though *OsAPS2b* itself is not expressed in vegetative tissues. Specifically, the expression of *OsAPL2* is strongly induced in the endosperm and at very low levels in the leaves. The plastid subunits *APS1*, *APS2a*, *APL1* and *APL3* accumulate to higher levels in the endosperm, and *APL1* and *APS1* also accumulate to higher levels in the leaves (Ohdan et al. 2005).

Other AGPase mutants have been described, but the reports focused mostly on the biochemical phenotype rather than the impact on other subunits. The rice *sugary* mutant is deficient in AGPase activity due to the loss of *OsAPS2* expression, resulting in the absence of both *APS2a* and *APS2b*. This mutant not only accumulates much less starch than wild-type plants in both the endosperm and leaves, but also features larger amyloplasts lacking visible starch granules (Kawagoe et al. 2005). The rice *apl1* mutant lacks an active *APL1* subunit due to the deletion of an essential conserved domain, and although there was no change in AGPase activity in the endosperm, no AGPase activity was detected in the leaves and the leaves contained <5% of normal starch levels but normal levels of soluble sugars (Rosti et al. 2007). These data suggest that the APL1 subunit is necessary for starch synthesis in leaves but not in non-photosynthetic organs. The rice *apl3* mutant lacks an active *APL3* subunit, and AGPase activity was reduced by 67% in the embryos (resulting in 35% less starch than normal) and by 23% in the endosperm, suggesting that *APL3* makes a major contribution in the embryo rather than to the endosperm (Cook et al. 2012).

The rice *phoI* mutant is characterized by the substantial loss of starch and altered amylopectin structure. There were no significant differences in the activity of AGPase, DBE, SBE or GBSS compared to wild-type plants, suggesting that PhoI influences two steps during starch biosynthesis, i.e. starch initiation and starch elongation (Satoh et al. 2008). The overexpression of *SuSy* in potato increased ADP-glucose and starch levels compared to wild-type plants, and the mutant tubers contained 55–85% more starch (Baroja-Fernández et al. 2009). In maize overexpressing *SuSy*, there was likewise an increase in ADP-glucose levels and a 10–15% more starch in the endosperm (Li et al. 2013). However, SuSy is not the major determinant of ADP-glucose production in cereals. An alternative explanation is that SuSy is responsible for the accumulation of UDP-glucose, resulting in higher levels of downstream metabolites such as glucose-1-phosphate, ADP-glucose and starch (Howard et al. 2012).

We knocked out one allele of the *OsAPL2* in two separate experiments to obtain two different mutants. We targeted an upstream exon and produced a truncated protein with no catalytic activity (E1). In a second set of experiments we targeted an exon downstream of the active site, to maintain some activity but perturb the tertiary structure (L1). We hypothesized that knocking out *OsAPL2* would have no effect on leaves. We specifically investigated heterozygous mutations in order to determine whether the mutation caused any feedback effects on the wild-type allele to restore normal starch biosynthesis. Remarkably, we

found that both mutations caused a reduction in starch levels and higher levels of soluble sugars in the leaves, even though the corresponding subunit *APS2b* is not expressed in photosynthetic tissues. Mutant E1 contained 85% less starch and 36% more sugar in the leaves than wild-type plants, whereas mutant L1 contained 50% less starch and 18% more sugar. Mutants were infertile because the lower starch levels resulted in male sterility (Lee Sk et al. 2016). Recently, Tang et al. (2016) reported similar results for the starch *w24* mutant which contained 21% less starch and 13%-43% more sugar. The *w24* mutant carries a homozygous single nucleotide substitution in exon 4 of the *OsAPL2*. In the *w24* mutant Glc-6-P is converted to Glc-1-P which is then converted by AGPases to ADP-Glc in amyloplasts in pollen.

AGPase activity in E1 was reduced by 31% correlating with the mutation of one allele, whereas AGPase activity in L1 increased by 2-fold because the mutated *APL2* allele retained some residual activity, simultaneously with an increase in the activity of the other subunits (please see below). We also measured the activity of SuSy because this enzyme is thought to provide an alternative route for starch biosynthesis, by converting sucrose and ADP directly to ADP-glucose, with fructose as a by-product (Li et al. 2013). We found that SuSy activity increased in E1 possibly as a direct compensation for the loss of AGPase activity, whereas SuSy activity in L1 was similar to wild-type plants, because L1 maintained activity in all alleles (WT and mutated). These data support a hypothesis in which Susy uses ADP which accumulates due to the lack of AGPase activity (Li et al. 2013).

Translation in the E1 mutant was terminated within exon 1. In contrast, the L1 mutant had an insertion of an adenosine residue within exon 13, causing a frameshift that generated a shorter protein with a different C-terminal domain. Sequence has shown that the C-terminal of the protein was conserved within the AGPase large subunit family and forms a loop structure located near the putative substrate and effector binding sites (Tuncel et al. 2014; Tang et al. 2016). Several *OsAPL2* point mutants in this region have been described, including substitutions T139I, A171V and L155P which replaced the original residues with bulkier side chains that altered the structure and modified the topology of the substrate and/or effector binding sites (Tuncel et al. 2014; Tang et al. 2016). The L1 mutant potentially has a similar impact by abolishing the hydrogen bond between residues Gln 98 with Arg 500 and Tyr 498 in the loop structure and forming a new hydrogen bond with Thr 351. Furthermore, 19 amino acids at the C-terminus of the large and small subunits of potato AGPase are essential for correct folding and/or assembly into multimeric complexes (Laughlin et al. 1998). In agreement with these results, the C-termini of *OsAPL2* and *OsAPS2b* were more important for the efficient multimerization of the AGPase subunits than the corresponding N-terminal regions (Tang et al. 2016). Our tetrameric modelling simulations show clearly that the salt bridge that stabilizes the binding between two APL2 and two APS2b subunits changes but the tetramer structure can still form due to a new salt bridge. L1 lacks the C-terminal β-sheets required for efficient subunit interactions. This explains the increase in subunit mRNA abundance without a corresponding increase in starch levels.

APS2b and APL2 are the only cytosolic AGPase subunits. They are preferentially expressed in the endosperm, so mutating *OsAPL2* is not expected to impact in leaves, where neither subunit is present at

significant levels and only plastidial AGPase activity is relevant. We compared the biochemical and structural parameters of AGPase in wild-type and mutant leaves. Surprisingly we observed changes in starch and soluble sugar levels, AGPase activity and SuSy activity. Expression analysis in the mutants showed that the *OsAPL2* transcript was more abundant than in wild-type plants, as was the transcript for the small subunit counterpart *OsAPS2b*. It therefore appears that in both mutants the *OsAPL2* locus is induced to compensate for the mutation of one allele, resulting in a counter-intuitive increase in expression from the wild-type allele, and a trans-acting effect on the expression of *OsAPS2b*. The expression levels of *OsAPL1* and *OsAPS1* also increased compared to wild-type plants. The expression level of *OsAPS2a* was reduced to half the level observed in wild-type plants whereas the levels of *OsAPL3* and *OsAPL4* mRNA remained similar to wild-type plants.

Increases in the expression of functional alleles to compensate for loss of expression in one allele at the same locus is not uncommon and has been reported earlier. Examples include dosage compensation to reflect the differential distribution of sex-chromosomes in animals and in a small number of plants, such as *Silene latifolia* (Muyle et al. 2012); compensation in the case of aneuploidy or polyploidy (Makarevitch and Harris 2010); compensation for monoallelic silencing as observed in paramutation and imprinting (Guidi et al. 2003); and buffering gene expression levels in the case of autosomal gene or allele copy number variation (Trieu et al. 2003). The compensatory increase in *OsAPL2* expression in L1 and E1 is caused by a mutation that changes the number of functional alleles, consistently with buffering. Little is known about inter-allelic buffering at the mRNA level in plants, but studies in yeast and insects have indicated that this may be a common mechanism that acts to stabilize the transcriptome (Lundberg et al. 2012; Bader et al. 2015). Such mechanisms may also partly explain the stability of heterozygous populations and hybrid vigor in plants (Verta et al. 2016).

Transcriptional compensation has also been observed among functionally redundant paralogous genes when one copy is mutated, e.g. among the nicotianamine synthase genes (Klatte et al. 2009). This may help to explain why the feedback mechanism that induced *OsAPL2* expression when one allele was mutated also induced the functionally related *OsAPS2*, although only the *OsAPS2b* mRNA was expressed at high levels. The major small subunit in leaves, APS2a, was significantly downregulated in E1 and L1 mutants, suggesting that the level of *OsAPS2a* mRNA may be independently suppressed at the post-transcriptional level. In turn, *OsAPS1* may be enhanced to compensate for the suppression of *OsAPS2a* in plastids. *OsAPL3* (the major large subunit isoform in leaves) and *OsAPL4* expression were similar to the wild type. The suppression of *OsAPS2a* explains why the increase in AGPase expression did not normalize starch levels in the two mutants. *OsAPS2a* is located in the plastids and the main impact of the two mutations (E1 and L1) was to induce the expression of an ectopic cytosolic form of AGPase, which is decoupled from the rest of the pathway because the latter is located within the plastids. The suppression of the major cytosolic small subunit may therefore be sufficient to limit starch synthesis. Ectopic cytosolic AGPase expression, even at high levels is unable to compensate for the loss of activity due to the absence of metabolic shuttling between the plastids and cytosol. Similar expression patterns were reported in other studies when *OsAPS2b*

expression was blocked, i.e. an increase in *OsAPL2* expression, dramatically increases *OsAPL1* and *OsAPS1* expression, but no changes in *OsAPL3* and *OsAPL4* (Ohdan et al. 2005). *OsPhoI* was also induced in L1 and E1 perhaps reflecting its key role in the initiation of starch biosynthesis (Satoh et al. 2008).

An alternative explanation for the lack of compensation of starch levels even when transcription of other subunits increases may be changes in Pi levels. In E1 the increase in SuSy activity responsible for the production of Pi in the cytosol promotes the exchange of Triose-P from the plastids reducing C3 intermediates required for starch synthesis (Preiss et al. 1994). In L1 even though SuSy activity did not increase, the concentration of soluble sugars were elevated and this may prevent Pi from being recycled for photophosphorylation and thereby very likely reducing C3 intermediates required for starch biosynthesis (Flugge et al. 1999).

1.5 Conclusion

In summary, mutating one allele of *OsAPL2*, encoding the major endosperm large subunit of AGPase and the only subunit expressed in the cytosol, resulted in the unexpected expression of both *OsAPL2* and *OsAPS2b* in the leaves, the latter encoding the only cytosolic small subunit. However, the formation of a complete ectopic AGPase in the leaf cytosol did not increase overall starch synthesis. Instead, the leaves contained less starch than wild-type plants most likely reflecting the lower levels of plastidial *OsAPS2a*, increased SuSy activity (at least in E1), the increased in soluble sugars and/or the inability of *OsAPS1* to replace *OsAPS2a* function completely. Our findings indicate that the new cytosolic AGPase was not sufficient to compensate for the loss of plastidial AGPase, probably because there is no wider starch biosynthesis pathway in the leaf cytosol and thus no pathway intermediates are shuttled between the two compartments. The principal differences between mutants E1 and L1 reflect the impact of changes in AGPase activity: in L1 overall AGPase activity has been changed whereas the alternative SuSy pathway was activated in E1.

1.7 References

Bader DM, Wilkening S, Lin G, Tekkedil MM, Dietrich K, Steinmetz LM, Gagneur J (2015) Negative feedback buffers effects of regulatory variants. Mol Syst Biol 11:785. http://doi.org/10.15252/msb.20145844

Ballicora MA, Iglesias AA Preiss, J (2003) ADP-glucose pyrophosphorylase: a regulatory enzyme for plant starch synthesis. ADP-Glucose Pyrophosphorylase: A Regulatory Enzyme for Plant Starch Synthesis. Microbiol Mol Biol Rev 67: 213-225

Ball, SG, Morell MK (2003) From bacterial glycogen to starch: Understanding the biogenesis of the plant starch granule. Annu Rev Plant Bio 54:207-233

Baroja-Fernandez E, Muñoz FJ, Montero M et al (2009) Enhancing sucrose synthase activity in transgenic potato (*Solanum tuberosum L*.) tubers results in increased levels of starch, ADPglucose and UDPglucose and total yield. Plant Cell Physiol 50:1651-1662

Bassie L, Zhu C, Romagosa I, Christou P, Capell T (2008) Transgenic wheat plants expressing an oat arginine decarboxylase cDNA exhibit increases in polyamine content in vegetative tissue and seeds. Mol Breed 22:39-50

Baysal C, Bortesi L, Zhu C, Farré 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. Mol Breed 36:108. https://doi.org/10.1007/s11032-016-0533-4

Bhagwat AS (1981) Activation of spinach ribulose 1,5-bisphosphate carboxylase by inorganic phosphate. Plant Sci Lett 23:197-206

Belhaj K, Chaparro-Garcia A, Kamoun S and Nekrasov V (2013) Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR/Cas system. Plant Methods 9:39. https://doi.org/10.1186/1746-4811-9-39

Bortesi L, Zhu C, Zischewski J, et al (2016) Patterns of CRISPR/Cas9 activity in plants, animals and microbes. Plant Biotechnol J 14(12):2203-2216

Chari R, Mali P, Moosburner M, Church GM (2015) Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach. Nat Methods 12:823-826

Chen K Gao C (2014) Targeted genome modification technologies and their applications in crop improvements. Plant Cell Rep 33:575-583

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 immature zygotic embryos. Nat Biotechnol 9:957-962

Cook FR, Fahy B, Trafford K (2012) A rice mutant lacking a large subunit of ADP-glucose pyrophosphorylase has drastically reduced starch content in the culm but normal plant morphology and yield. Funct Plant Biol 39:1068-1078

Colleoni C, Dauvillèe D, Moulle G et al (1999) Genetic and biochemical evidence for the involvement of α-1,4 glucanotransferases in amylopectin synthesis. Plant Physiol 120:993-1003

Dennis DT, Miernyk JA (1982). Compartmentation of non-photosynthetic carbohydrate metabolism. Ann. Rev. Plant Physiol 33:27-50

Doehlert DC, Kuo TM, Felker FC (1988) Enzymes of sucrose and hexose metabolism in developing kernels of two inbreds of maize. Plant Physiol 86:1013-1019

Doudna JA, Charpentier E (2014) The new frontier of genome engineering with CRISPR-Cas9. Science 346:1258096

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 Res 21:1093-1097

Fauser F, Schiml S, Puchta H (2014) Both CRISPR/Cas-based nucleases and nickases can be used efficiently for genome engineering in Arabidopsis thaliana. Plant J 79:348-359

Flugge UI (1999) Phosphate translocators in plastids. Ann Rev Plant Physiol Plant Mol Biol 50:27-45

Giroux MJ, Hannah L (1994) ADP-glucose pyrophosphorylase in shrunken-2 and brittle-2 mutants of maize. Mol Gen Genet 243:400-408

Guidi CJ, Veal TM, Jones SN, Imbalzano AN (2004) Transcriptional Compensation for Loss of an Allele of the Ini1 Tumor Suppressor. J Biol Chem 279:4180-4185

Heldt HW, Chon CH, Maronde D, Herold A, Stankovic AZ, Walker DA, Kraminer A, Kirk MR, Heber U (1977) Role of orthophosphate and other factors in the regulation of starch formation in leaves and isolated chloroplasts. Plant Physiol 59:1146-1155

Heldt HW, Chon CJ, Lorimer H (1978) Phosphate requirement for the light activation of ribulose-1,5-biphosphate carboxylase in intact spinach chloroplasts. FEBS Lett 92:234-240

Heigwer F, Kerr G, Boutros M (2014) E-CRISP: fast CRISPR target site identification. Nat Methods 11:122-123

Howard TP, Fahy B, Craggs A. et al (2012) Barley mutants with low rates of endosperm starch synthesis have low grain dormancy and high susceptibility to preharvest sprouting. New Phytol 194:158-167

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science 17:816-821

Jiang W, Brueggeman AJ, Horken KM, Plucinak TM, Weeks DP (2014) Successful transient expression of cas9 and single guide RNA genes in Chlamydomonas reinhardtii. Eukaryot Cell 13:1465-1469

Jobling, S (2004) Improving starch for food and industrial applications. Curr Opin Plant Biol 7:210-218

Johnson PE, Patron NJ, Bottrill AR, et al (2003) A low-starch barley mutant, risø 16, lacking the cytosolic small subunit of ADP-glucose pyrophosphorylase, reveals the importance of the cytosolic isoform and the identity of the plastidial small subunit. Plant Physiol 131:684-696

Kawagoe Y, Kubo A, Satoh H, TakaiwaF, Nakamura Y (2005) Roles of isoamylase and ADP-glucose pyrophosphorylase in starch granule synthesis in rice endosperm. Plant J 42:164-174

Kang TJ, Yang MS (2004) Rapid and reliable extraction of genomic DNA from various wild-type and transgenic plants. BMC Biotechnol 4:20. https://doi.org/10.1186/1472-6750-4-20

Klatte M, Schuler M, Wirtz M, Fink-Straube C, Hell R, Bauer P (2009) The analysis of Arabidopsis nicotianamine synthase mutants reveals functions for nicotianamine in seed iron loading and iron deficiency responses. Plant Physiol 150:257-271

Laughlin MJ, Chantler SE, Okita TW (1998) N- and C- terminal peptide sequences are essential for enzyme assembly allosteric, and/or catalytic properties of ADP-glucose pyrophosphorylase. Plant J 14:159-168

Lee SK, Hwang SK, Han M, Eom JS, Kang HG, Han Y, et al (2007) Identification of the ADP-glucose pyrophosphorylase isoforms essential for starch synthesis in the leaf and seed endosperm of rice (*Oryza sativa L*.). Plant Mol Biol 65:531-546

Lee J, Chung JH, Kim HM, Kim, DW, Kim H (2016) Designed nucleases for targeted genome editing. Plant Biotechnol J 14:448-462

Lee SK, Eom JS, Hwang SK, Shin D, An G, Okita TW, Jeon JS (2016) Plastidic phosphoglucomutase and ADP-glucose pyrophosphorylase mutants impair starch synthesis in rice pollen grains and cause male sterility. J Exp Bot 67:5557-5569

Li J, Baroja-Fernández E, Bahaji A, Muñoz FJ, Ovecka M, Montero M, et al (2013) Enhancing sucrose synthase activity results in an increased levels of starch and ADP-Glucose in maize (*Zea mays* L.) seed endosperms. Plant Cell Physiol 54:282-294

Lundberg LE, Figueiredo ML, Stenberg P, Larsson J (2012) Buffering and proteolysis are induced by segmental monosomy in Drosophila melanogaster. Nucleic Acids Res 40:5926-5937

Ma X, Zhang Q, Zhu Q, Liu W, Chen Y, Qiu R, et al (2015) A robust CRISPR/Cas9 system for convenient, high-efficiency multiplex genome editing in monocot and dicot plants. Mol Plant 8:1274-1284. doi: 10.1016/j.molp.2015.04.007.

Makarevitch I, Harris C (2010) Aneuploidy causes tissue-specific qualitative changes in global gene expression patterns in maize. Plant Physiol 152:927-938

Maresca M, Lin VG, Guo N, Yang Y (2013) Obligate ligation-gated recombination (ObLiGaRe): custom-designed nuclease-mediated targeted integration through nonhomologous end joining. Genome Res 23:539-546

Martin C, Smith AM (1995) Starch biosynthesis. Plant Cell 7:971-985

Mikami M, Toki S, Endo M (2016) Precision targeted mutagenesis via Cas9 paired nickases in rice. Plant Cell Physiol 57:1058-1068

Müller-Röber B, Sonnewald U, Willmitzer L (1992) Inhibition of the ADP-glucose pyrophosphorylase in transgenic potatoes leads to sugar-storing tubers and influences tuber formation and expression of tuber storage protein genes. EMBO J 11:1229-1238

Muyle A, Zemp N, Deschamps C, Mousset S, Widmer A, Marais GA (2012) Rapid de novo evolution of x chromosome dosage compensation in *Silene latifolia*, a plant with young sex chromosomes. PLoS Biol 10:e1001308

Nakamura Y, Francisco PB Jr, Hosaka Y, Sato A, Sawada T, Kubo A, Fujita N (2005) Essential amino acids of starch synthase IIa differentiate amylopectin structure and starch quality between *japonica* and *indica* rice varieties. Plant Mol Biol 58:213-227

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

Ohdan T, Francisco PB 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. J Exp Bot 56:3229-3244

Osakabe Y, Osakabe K (2015) Genome editing with engineered nucleases in plants. Plant Cell Physiol 56:389-400

Pandey MK, Rani NS, Madhav MS, Sundaram RM, Varaprasad GS, Sivaranjani AK, et al (2012) Different isoforms of starch-synthesizing enzymes controlling amylose and amylopectin content in rice (*Oryza sativa L*.). Biotechnol Adv 30:1697-706

Preiss J (1982) Regulation of the biosynthesis and degradation of starch. Ann Rev Plant Physiol 33:431-454

Preiss J (1994) Regulation of the C_3 reductive cycle and carbohydrate synthesis. In: Tolbert NE, Preiss J (eds) Regulation of atmospheric CO_2 and O_2 by photosynthetic carbon metabolism, 1st edn. Oxford university press, New York, pp 93-102

Quetier F (2016) The CRISPR-Cas9 technology: closer to the ultimate toolkit for targeted genome editing. Plant Science 242:65-76

Ran FA, Hsu PD, Lin CY, Gootenberg JS, Konermann S, Trevino AE, et al (2013) Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell 154:1380-1389

Rychter AM, Rao IM (2005) Role of phosphorus in photosynthetic carbon metabolism. In: Pessarakli M. (ed) Handbook of photosynthesis, 2nd edn. Taylor y Francis group, Tucson, pp 123-148

Rösti S, Fahy B, Denyer K (2007) A mutant of rice lacking the leaf large subunit of ADP-glucose pyrophosphorylase has drastically reduced leaf starch content but grows normally. Funct Plant Biol 34:480-489

Satoh H, Shibahara K, Tokunaga T, Nishi A, Tasaki M, Hwang SK, et al (2008) Mutation of the plastidial alpha-glucan phosphorylase gene in rice affects the synthesis and structure of starch in the endosperm. Plant Cell 20:1833-1849

Shan Q, Wang Y, Li J, Gao C (2014) Genome editing in rice and wheat using the CRISPR/Cas system. Nat Protoc 9:2395-2410

Shimomura S, Nagai M, Fukui T (1982) Comparative glucan specificities of two types of spinach leaf phosphorylase. J Biochem 91:703-717

Steup, M (1990) Starch degrading enzymes. In: Dey PM, Harborne JB (eds) Methods in Plant Biochemistry. Academic Press London, pp. 103-128

Stitt M, Heldt HW (1981) Physiological rates of starch breakdown in isolated intact spinach chloroplasts. Plant Physiol 68:755-761

Sudhakar D, Duc L.T, Bong BB, Tinjuangjun P, Maqbool SB, Valdez M, et al (1998) An efficient rice transformation system utilizing mature seed-derived explants and a portable, inexpensive particle bombardment device. Transgenic Res 7:289-294

Sun Y, Jiao G, Liu Z et al (2017) Generation of high-amylose rice through CRISPR/Cas9-mediated targeted mutagenesis of starch branching enzymes. Front Plant Sci 8:298. https://doi.org/10.3389/fpls.2017.00298

Tang XJ, Peng C, Zhang J, Cai Y, You XM, Kong F et al (2016) ADP-glucose pyrophosphorylase large subunit 2 is essential for storage substance accumulation and subunit interactions in rice endosperm. Plant Sci 249:70-83

Tester RF, Morrison, WR, Schulman AH (1993) Swelling and gelatinization of cereal starches. V. Risø mutants of bomi and carlsberg II barley cultivars. J Cereal Sci 17:1-9

Tetlow IJ, Wait R, Lu Z, Akkasaeng R, Bowsher CG, Esposito S et al (2004) Protein phosphorylation in amyloplasts regulates starch branching enzyme activity and protein-protein interactions. Plant Cell 16:694-708

Trieu M, Ma A, Eng SR, Fedtsova N, Turner EE (2003) Direct autoregulation and gene dosage compensation by POU-domain transcription factor Brn3a. Development 130:111-121

Tsai CY, Nelson OE (1966) Starch-deficient maize mutant lacking adenosine diphosphate glucose pyrophosphorylase activity. Science 151:341-343

Tuncel A, Kawaguchi J, Ihara Y, Matsusaka H, Nishi A, Nakamura T et al (2014) The rice endosperm ADP-Glucose pyrophosphorylase large subunits essential for optimal catalysis and allosteric regulation of the heterotetrameric enzyme. Plant Cell Physiol 55:1169-1183

Valdez M, Cabrera-Ponce JL, Sudhakar D, Herrera-Estrella L, Christou P (1998) Transgenic central american, west african and asian elite rice varieties resulting from particle bombardment of foreign DNA into mature seed-derived explants utilizing three different bombardment devices. Ann Bot 82:795-801

Verta JP, Landry CR, MacKay J (2016) Dissection of expression-quantitative trait locus and allele specificity using a haploid/diploid plant system-insights into compensatory evolution of transcriptional regulation within populations. New Phytol 211:159-171

Villand P, Olsen OA, Kleczkowski LA (1993) Molecular characterization of multiple cDNA clones for ADP-glucose pyrophosphorylase from *Arabidopsis thaliana*. Plant Mol Biol 23:1279-1284

Yoshida S, Forno DA, Cock JH, Gomez, KA (1976) Determination of sugar and starch in plant tissue, 3erd edition, In: Laboratory Manual for Physiological Studies of Rice, The international rice research institute, Laguna Philippines pp. 46-49.

Yu Y, Mu HH, Wasserman BP, Carman GM (2001) Identification of the maize amyloplast stromal 112-kD protein as a plastidic starch phosphorylase. Plant Physiol 125:351-359

Yuan D, Bassie L, Sabalza M, Miralpeix B, Dashevskaya S, Farre G et al (2011) The potential impact of plant biotechnology on the Millennium Development Goals. Plant Cell Rep 30:249-265

Zhang D, Wu J, Zhang Y, Shi C (2012) Phenotypic and candidate gene analysis of a new floury endosperm mutant (*osagpl2-3*) in rice. Plant Mol Biol Report 30:1303-1312

Zhu C, Sanahuja G, Yuan D, Farré G, Arjó G, Berman J, et al (2013) Biofortification of plants with altered antioxidant content and composition: genetic engineering strategies. Plant Biotechnol J 11:129-141

Zhu C, Bortesi L, Baysal C, Twyman RM, Fischer R, Capell T et al (2017) Characteristics of genome editing mutations in cereal crops. Trends Plant Sci 22:38-52

CHAPTER 2

CRISPR/Cas9 mutations in the rice *Waxy*/GBSSI gene induce allele-specific and zygosity-dependent feedback effects on endosperm starch biosynthesis

2.1 Abstract

The mutation of genes in the starch biosynthesis pathway has a profound effect on starch quality and quantity and is an important target for plant breeders. Mutations in endosperm starch biosynthetic genes may impact starch metabolism in vegetative tissues such as leaves in unexpected ways due to the complex feedback mechanisms regulating the pathway. Surprisingly this aspect of global starch metabolism has received little attention. We used CRISPR/Cas9 to introduce mutations affecting the Waxy (Wx) locus encoding granule bound starch synthase I (GBSSI) in rice endosperm. Our specific objective was to develop a mechanistic understanding of how the endogenous starch biosynthetic machinery might be affected at the transcriptional level following the targeted knock out of GBSSI in the endosperm. We found that the mutations reduced but did not abolish GBSS activity in seeds due to partial compensation caused by the ectopic upregulation of GBSSII. The GBSS activity in the mutants was 61–71% of wild-type levels, similarly to two irradiation mutants, but the amylose content declined to 8–12% in heterozygous seeds and to as low as 5% in homozygous seeds, accompanied by abnormal cellular organization in the aleurone layer and amorphous starch grain structures. Expression of many other starch biosynthetic genes was modulated in seeds and leaves. This modulation of gene expression resulted in changes in AGPase and sucrose synthase activity that explained the corresponding levels of starch and soluble sugars.

2.2 Introduction

Starch is the major component of rice endosperm and comprises a mixture of the polysaccharides amylose and amylopectin. Amylose is a linear chain of α(1,4)-linked glucose monomers whereas amylopectin contains additional α(1,6)-linked branches every 24–30 residues (Martin and Smith, 1995). Starch from different plant species varies in its physicochemical properties due to the ratio of amylose and amylopectin and differences in chain length and/or amylopectin branching density (Jobling, 2004). These differences reflect the activity of various enzymes in the starch biosynthesis pathway.

Starch synthesis in plants begins with the conversion of glucose 1-phosphate to ADP-glucose, catalyzed by ADP-glucose pyrophosphorylase (AGPase). This is followed by the polymerization of ADP-glucose to form amylose and amylopectin via the coordinated activity of AGPase and starch synthases, which form the α(1,4)-linked glycosidic bonds of both molecules, and starch branching enzymes (SBEs), which form the α(1,6)-linked glycosidic bonds in amylopectin. Starch synthases catalyze the transfer of the glucosyl moiety of the soluble precursor ADP-glucose to the reducing end of a pre-existing α(1,4)-linked glucan primer (Tetlow et al., 2004) whereas SBEs cleave internal amylose α(1,4) glycosidic bonds and transfer the

reducing ends to C6 hydroxyls to form α(1,6) linkages (Jiang et al., 2013). The latter can be removed by starch debranching enzyme isoamylase, hence the amylopectin content of starch is sensitive to the balance between branching and debranching activities. The structure of amylopectin is also influenced by the two different SBE isoforms, with SBEI showing higher affinity towards amylose than amylopectin and a preference for longer glucan chains and SBEII showing the opposite properties (Tetlow et al., 2004). Finally, disproportionating enzyme (DPE) and starch phosphorylase (PHO) also play a role in the initiation and elongation of starch polymers (Satoh et al., 2008). The pathway is summarized in **Figure 1**.

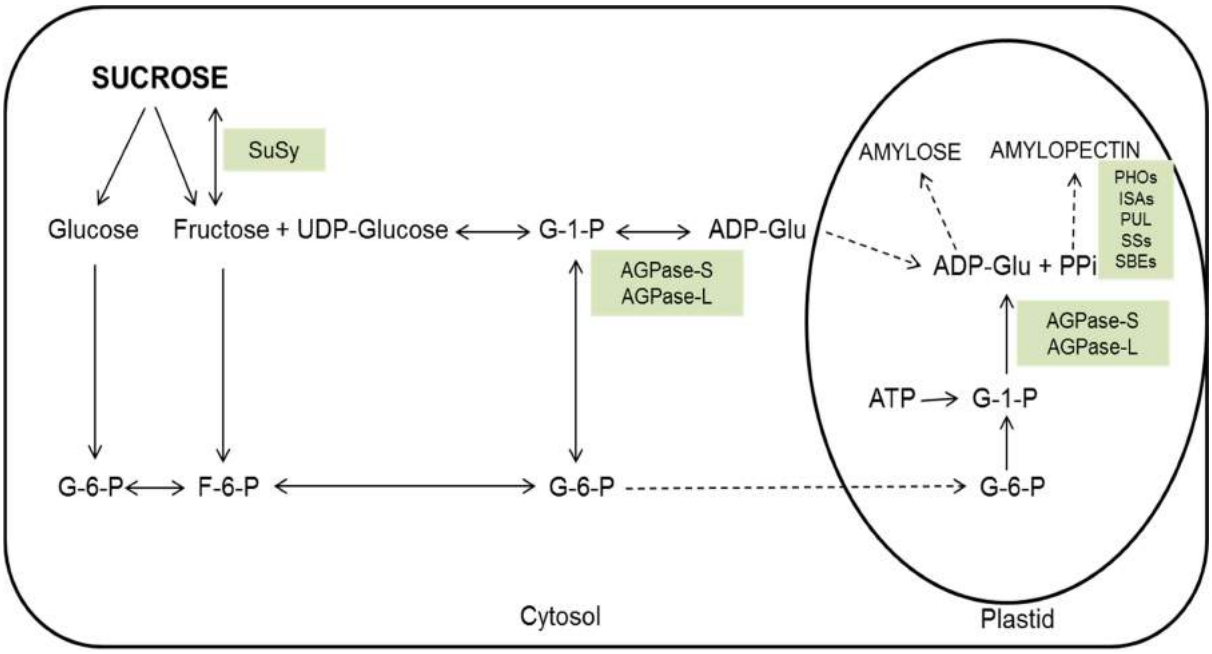


Figure 1 The coordination of different starch biosynthesis genes in rice (modified from Thitisaksakul et al., 2012). Abbreviations: SuSy (sucrose synthase), G-1-P (glucose-1-phosphate), G-6-P (glucose-6-phosphate), ADP-Glu (ADP-glucose), PPi (inorganic diphosphate); F-6-P (fructose-6-phosphate); ATP (adenosine triphosphate); ADP (adenosine diphosphate), UDP (uridine diphosphate), PHO (α-glucan phosphorylase); AGPase (ADP-glucose pyrophosphorylase); ISAs (isoamylase-type starch debranching enzymes); PUL (pullulanase); SS (starch synthases); GBSSI (granule bound starch synthase 1); SBEs (starch-branching enzymes).

There are two major groups of starch synthases (**Figure 2**). The first group is the classical starch synthases (SS) and comprises four isoforms, some represented by multiple paralogs: SSI, SSIIa/b/c, SSIIIa/b and SSIVa/b (Nakamura, 2002). These synthesize the linear chains of amylopectin and their distribution between granular and stromal fractions can vary between species, tissues and developmental stages (Ball and Morell, 2003). The second group is the granule-bound starch synthases (GBSS) which are restricted to the granule matrix. There are two isoforms (GBSSI and GBSSII), the first expressed mainly in the endosperm and the second mainly in the leaves (Ohdan et al., 2005). GBSSI catalyzes the extension of long glucan chains primarily in amylose (Maddelein et al., 1994). The GBSSI protein is 609 amino acids in length, with a catalytic site spanning residues 78–609 composed of α-helices and β-sheets that form a substrate binding cleft. An N-terminal transit peptide outside the catalytic center is required to import the protein into the starch granules (Momma and Fukimoto, 2012).

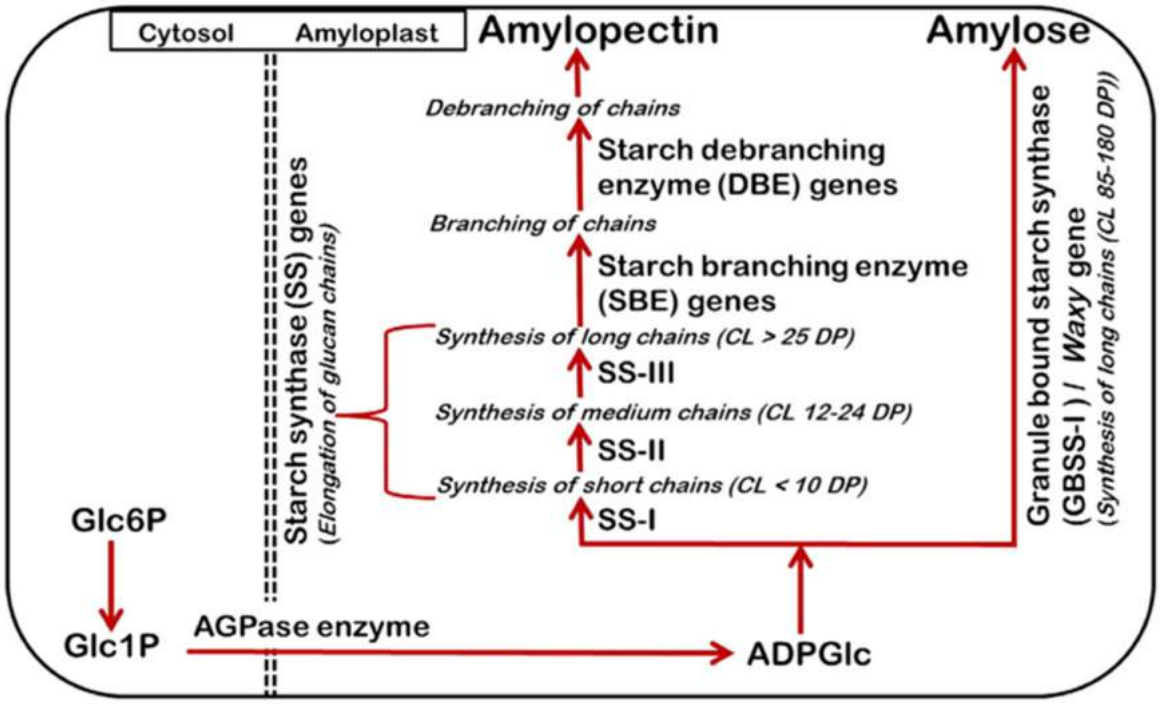


Figure 2 Steps in the starch biosynthesis pathway that generate the different components of starch found in rice endosperm.

In rice, GBSSI is encoded by the *Waxy* (*Wx*) gene, so named because of the waxy appearance of the amylose-free grain in *Wx* null mutants (Hirano, 1993). As well as the *Wx* null allele, two other natural alleles are common in rice (Sano, 1984). The *Wxa* allele is predominant in the indica subspecies, which has strong GBSSI activity and thus more amylose in the endosperm, whereas *Wxb* has weaker GBSSI activity and the amount of amylose and amylopectin is more variable (Hirano and Sano, 1998; Umenmoto and Terashima, 2002). Because of its impact on grain quality, *Wx* is a key target in rice quality improvement programs and in studies of starch biosynthesis and metabolism (Tran et al., 2011; Zhang et al., 2012). The *Wx* gene has been modified by conventional mutagenesis strategies such as irradiation, chemical mutagenesis and T-DNA/transposon insertional mutagenesis but these approaches generate random lesions and large populations must be screened to isolate useful mutants. Such techniques have largely been supplanted by targeted mutagenesis using designer nucleases, particularly CRISPR/Cas9, because this can modify the target gene without altering other agronomic traits (reviewed by Bortesi et al., 2016; Zhu et al., 2017).

Several previous studies have targeted GBSSI in rice using the CRISPR/Cas9 system. In one example, wild-type Cas9 (Cas9WT) was used to target three different sites simultaneously although only plants with one or two target site mutations were recovered. The amylose content in the T1 seeds was reduced from 14.6% to 2.6% (Ma et al., 2015). More recently, four *Wx* frameshift mutants were generated using CRISPR/Cas9 and the proportion of amylose in the seeds was reduced without affecting the overall starch content or agronomic properties such as seed number, yield and size (Zhang et al., 2018). However, these studies did not consider the broader impact on starch metabolism, reflecting feedback mechanisms that may be activated to restore homeostasis in the developing seed. The starch pathway is tightly regulated by the ratio of triose-phosphate to inorganic phosphate, and the disruption of this balance could lead to unanticipated

effects. In a *Wx* mutant generated by irradiation (*GM077*), the loss of GBSSI activity and lower starch and amylose levels was shown to induce the expression of GBSSII, AGPases, starch synthases and isoamylases (Zhang et al., 2012). Given that the mutation of endosperm-specific *GBSSI* is considered a strategy to modulate amylose production in the endosperm without affecting amylose metabolism in vegetative tissues, the collateral effect on GBSSII (affecting starch biosynthesis in the leaves) and on enzymes related to amylopectin synthesis and starch degradation needs to be investigated in more detail.

We hypothesized that these collateral effects may differ in mutants characterized by the complete abolition of GBSSI activity and a partial loss of activity, with the latter anticipated to reduce endosperm amylose levels with no further impact on starch metabolism as previously reported (Liu et al., 2009; Zhang et al., 2012; 2018). We therefore used the CRISPR/Cas9 system to generate truncated (nonfunctional) and partly active mutants of GBSSI in rice and carried out a comprehensive analysis of starch and sucrose levels, grain morphology and the expression of other genes involved in starch metabolism.

2.3 Materials and methods

Target sites and sgRNA design

Target sites for Cas9WT (single sgRNA) and Cas9 nickase (Cas9D10A, two sgRNAs targeting adjacent sites) were selected within the rice *Wx* coding sequence (GenBank EU735072.1) using E-CRISP (Heigwer et al., 2014) with the following parameters: only NGG PAM, only G as the 5' base, off-target tolerates many mismatches, non-seed region ignored, introns ignored. These parameters were selected to minimize off-target cleavage. The catalytic efficiency of each sgRNAs was predicted using gRNA scorer (Chari et al., 2015). The following three targets sites were selected: TS1 = 5'-GTCGGCGATGCCGAAGC↓CGGTGG-3', TS2 = 5'-GCTGCTCCGCCACGGGT↓TCCAGG-3' and TS3 = 5'-CCGGCTTCGGCATCGCC↓GACAGG-3', where the arrows indicate the expected site of the DSB.

Vector construction

The Cas9WT vector pJIT163-2NSCas9 and the sgRNA vector pU3-gRNA were obtained from Dr. C. Gao, Chinese Academy of Sciences, Beijing, China (Shan et al., 2014). The nickase vector pJIT163-2NSCas9D10A was constructed in-house by mutating the *cas9* gene in vector pJIT163-2NSCas9 to produce Cas9D10A and combining it with the maize *ubiquitin*-1 promoter and Cauliflower mosaic virus 35*S* terminator. The three sgRNAs described above were prepared as synthetic double-stranded oligonucleotides and introduced

separately into pU3-gRNA at the AarI restriction site allowing all genomic sites with the form 5′-(20)-NGG-3′ to be targeted. The *hpt* selectable marker gene was provided on a separate vector as previously described (Christou et al., 1991).

Rice transformation and recovery of transgenic plants

Seven-day-old mature zygotic embryos (*Oryza sativa* cv. EYI) were transferred to osmotic medium (MS medium supplemented with 0.3 g/L casein hydrolysate, 0.5 g/L proline, 72.8 g/L mannitol and 30 g/L sucrose) 4 h before bombardment with 10 mg gold particles coated with the transformation vectors. The Cas9 vector (wild type or nickase), the corresponding sgRNA vector(s) and the selectable marker *hpt* were introduced at a 3:3:1 molar ratio (Cas9WT:sgRNA:*hpt*) or a 3:3:3:1 molar (Cas9D10A:sgRNA1:sgRNA2:*hpt*) as previously described (Christou et al., 1991; Sudhakar al., 1998; Valdez et al., 1998). The embryos were returned to osmotic medium for 12 h before selection on MS medium supplemented with 0.3 g/L casein, 0.5 g/L proline, 30 g/L sucrose, 50 mg/L hygromycin and 2.5 mg/L 2,4-dichlorophenoxyacetic acid in the dark for 2–3 weeks. Callus was maintained on selective medium for 6 weeks with sub-culturing every 2 weeks as described (Farré et al., 2012). Transgenic plantlets were regenerated and hardened off in soil. For negative controls, we regenerated negative transformants (also bombarded with Cas9WT/Cas9D10A, *hpt* and the appropriate sgRNAs but which did not undergo mutation) at the same time as the mutated lines.

Confirmation of the presence of cas9 transgene and gRNA

Genomic DNA was isolated from the callus and leaves of regenerated plants by phenol extraction and ethanol precipitation (Bassie et al., 2008; Kang and Yang, 2004). The presence of the Cas9WT sequence was confirmed by PCR using primers 5′-GTC CGA TAA TGT GCC CAG CGA-3′ and 5′-GAA ATC CCT CCC CTT GTC CCA-3′. The presence of the Cas9D10A sequence was determined using primers 5′-GCA AAG AAC TTT CGA TAA CGG CAG CAT CCC TCA CC-3′ and 5′-CCT TCA CTT CCC GGA TCA GCT TGT CAT TCT CAT CGT-3′. The presence of the pU3-gRNA vectors was confirmed using the conserved primers 5′-TTG GGT AAC GCC AGG GTT TT-3′ and 5′-TGT GGA TAG CCG AGG TGG TA-3′.

Analysis of induced mutations

The *Wx* mutations induced by Cas9WT and Cas9D10A were detected by PCR using primers 5′-GGG TGC AAC GGC CAG GAT ATT-3′ and 5′-TGA AGA CGA CGA CGG TCA GC-3′. The PCR products were sequenced using

an ABI 3730xl DNA analyzer by Stabvida (http://www.stabvida.com/es/). To confirm the mutations, PCR products generated using the primers listed above were purified using the Geneclean II Kit (MP Biomedicals), transferred to the pGEM-T Easy vector (Promega) and introduced into competent *Escherichia coli* cells. PCR fragments of ~760 bp were selected, purified and sequenced using an ABI 3730xl DNA analyzer by Stabvida. At least five clones per PCR product were sequenced using primer M13Fwd.

Protein structural modeling and phylogenetic analysis

The GBSSI sequences were translated into polypeptides (http://web.expasy.org/translate/) and automated homology modeling was carried out using Phyre2 (Kelley et al., 2015) with rice GBSSI catalytic domain (Protein Databank: 3vuf) as the template. The model of the mutant protein was superimposed on the wild-type version using DS Visualizer (http:/accelrys.com/products/collaborative-science/biovia-discovery-studio/visualization.html). The structure with the ADP and malto-oligosaccharide was docked to the protein using swissdock (http://www.swissdock.ch/docking). Sequence alignment and phylogenetic tree construction was carried out using the phylogeny.fr (Dereeper et al., 2008) server (http://www.phylogeny.fr/index.cgi) with default parameters.

Enzymatic activity assays

Leaves were cut into discs and immersed in 2% paraformaldehyde, 2% polyvinylpyrrolidone 40 (pH 7) for 2.5 h at 4 °C before washing three times in distilled water. AGPase and SuSy activity was then measured using a proprietary kit, according to the manufacturer's instructions (CSIC, 2016).

The GBSS activity in 10 frozen seeds was determined according to Nakamura et al. (1989) and Jiang et al. (2004). The seeds were weighed and homogenized in 10 ml ice-cold 50 mM HEPES–NaOH (pH 7.5) containing 10 mM $MgCl_2$, 2 mM EDTA, 50 mM 2-mercaptoethanol, 12.5% (v/v) glycerol and 5% (w/v) polyvinylpyrrolidone 40. We added 30 µl of the homogenate to 1.8 ml of the same buffer and centrifuged at 2000 x *g* for 20 min at 4 °C. The pellet was resuspended in 2 ml of GBSS assay buffer (100 µl 14 mM ADP-glucose and 700 µl 50 mg/ml amylopectin). After incubation at 30 °C for 5 min, the reaction was initiated by adding 50 µl of the enzyme extract (40 mM phosphoenolpyruvate, 50 mM $MgCl_2$, and 1 IU pyruvate kinase), incubated at 30 °C for 30 min, and was stopped after 20 min by heating in a boiling water bath. A control sample was prepared by boiling the enzyme extract before starting the reaction, to determine the background signal. The ADP produced by GBSS was converted to ATP by the action of pyruvate kinase. . The ATP was measured by adding 5 ml of luciferin reagent to 50 µl of the enzymatic reaction after the production of ATP and measuring the emission at 370–630 nm in a Berthold FB 12 luminometer.

Starch and soluble sugars

Flag leaf samples harvested at 7 pm were homogenized under liquid nitrogen and extracted in perchloric acid to measure the starch content, or in ethanol to measure the content of soluble sugars. The quantity of each carbohydrate was determined by spectrophotometry at 620 nm (Juliano, 1971; Yoshida et al., 1976). To measure the amylose content, milled rice grains were powdered with a faience pestle and mortar and the powder was transferred to a paper envelope and dried for 1 h at 135 °C. We transferred 100±0.01 mg of dried powder to a conical flask and added 1 ml 95% ethanol and 9 ml 1 M NaOH. The suspension was boiled in a water bath for 10 min, cooled at room temperature for 10 min and then topped up to 100 ml with distilled water. A 5-ml aliquot of the solution was transferred to a 100-ml volumetric flask and mixed with 1 ml 1 M acetic acid, 2 ml 0.2% potassium iodide and 92 ml distilled water. Three amylose solutions (3%, 11.5% and 14%) were prepared for comparison. The starch content was determined by measuring the absorbance at 630 nm in a Unicam UV4-100 UV-Vis spectrophotometer after 30 min.

RNA extraction and real-time quantitative RT-PCR analysis

Total leaf RNA was isolated using the RNeasy Plant Mini Kit (Qiagen) and DNA was digested with DNase I from the RNase-free DNase Set (Qiagen). Total RNA was quantified using a Nanodrop 1000 spectrophotometer (Thermo Fisher Scientific) and 2 µg total RNA was used as template for first strand cDNA synthesis with Quantitech reverse transcriptase (Qiagen) in a 20-µl total reaction volume, following the manufacturer's recommendations. Real-time qRT-PCR was performed on a BioRad CFX96 system using 20-µl mixtures containing 5 ng cDNA, 1x iQ SYBR green supermix and 0.5 µM forward and reverse primers. The *OsAPL1*, *OsAPL3*, *OsAPL4*, *OsAPS1*, *OsAPS2a/b, OsAPL2, OsSSI, OsSSIIa, OsSSIIb, OsSSIIc, OsSSIIIa, OsSSIIIb, OsSSIVa, OsSSIVb, OsGBSSI, OsGBSSII, OsBEI, OsBEIIa, OsBEIIb, OsISA1, OsISA2, OsISA3, OsPUL, OsDPE1, OsDPE2, OsPHOH* and *OsPHOL* cDNAs were amplified using appropriate primers (Ohdan et al. 2005) as described by Tang et al. (2016). Serial dilutions of cDNA (80–0.0256 ng) were used to generate standard curves for each gene. PCR was performed in triplicate using 96-well optical reaction plates. Values represent the mean of three biological replicates ± SE. Amplification efficiencies were compared by plotting the ΔCt values of different primer combinations of serial dilutions against the log of starting template concentrations using the CFX96 software. The rice housekeeping *OsUBQ5* (LOC_Os01g22490) was used as an internal control.

Seed phenotype and microscopy

The seed hull was removed to observe the external appearance of the grain in mutant lines using a magnifying lens. Thin sections (2 µm) of rice endosperm and leaves were prepared with a diamond knife using a Reichter Jung Ultramicrotome Ultracut E and were mounted on glass slides for analysis under a Zeiss Axioplan light microscope coupled to a Leica DC 200 digital camera. Tissue for sectioning was prepared by embedding in glycol methacrylate using the Technovit 7100 kit according to the manufacturer's protocol (Kulzer). Grains were cut through the center to expose the endosperm. Drops of 1.0% Richardson Blue were placed on the endosperm surface and images were captured after 3–5 min. Rice seeds were fractured in half and mounted in stubs with carbon tape to keep them vertical. Then they were processed for scanning electron microscopy (SEM) by dehydration at 60 °C for 24 h followed by carbon coating using an Edwards Auto 306 and gold sputtering using a Balzers SCD050 Sputter Coater. The samples were stored at 60 °C or in a vacuum chamber before analysis on a Jeold JSM-6300.

Statistical analysis

A general linear model was used to determine statistically significant differences in normalized expression of starch pathway genes. All the analyses were performed using the JMP Pro (JMP, SAS Institute Inc., Cary, NC, 2013). Five-factorial analysis of variance (ANOVA) with tissue, gene, genotype, gene_type and isoform as random factors was applied per normalized expression on log-transformed data. Other statistical analysis were determine by Student's test.

Accession numbers

GenBank EU735072.1 (*OsWaxy* gene sequence)

UniProtKB/Swiss-Prot: P04713.1 (Waxy protein *Oryza sativa* subsp. *Indica)*

UniProtKB/Swiss-Prot: Q0DEV5 (Waxy protein *Oryza sativa* subsp. *japonica)*

UniProtKB/Swiss-Prot: P09842 (Waxy protein *Hordeum vulgare* (barley)*)*

UniProtKB/Swiss-Prot: P27736 (Waxy protein *Triticum aestivum* (wheat)*)*

UniProtKB/Swiss-Prot: Q9MAQ0 (Waxy protein *Arabidopsis thaliana)*

UniProtKB/Swiss-Prot: A1YZE0 (Waxy protein *Glycine max* (soybean)*)*

UniProtKB/Swiss-Prot: M9Q2A3 (Waxy protein *Nicotiana tabacum* (tobacco)*)*

UniProtKB/Swiss-Prot: D2D315 (Waxy protein *Gossypium hirsutum* (cotton)*)*

UniProtKB/Swiss-Prot: K4CPX6 (Waxy protein *Solanum lycopersicum* (tomato)*)*

2.4 Results

Recovery and characterization of mutants

We analyzed six induced mutant lines in detail (**Table 1**). PCR analysis revealed that line 1 featured the substitution of a single nucleotide at site TS1 (**Figure 3a**). The remaining five lines all featured mutations at sites TS2/TS3. Lines 2, 4 and 5 featured deletions of 55, 56 and 246 nucleotides, respectively, line 3 featured an insertion of 28 nucleotides, and line 6 featured a 2-nucleotide substitution (**Figure 3 b-f**). In T0 plants, all the lines contained heterozygous mutations except line 2, in which the mutation was homozygous. As well as testing for on-target mutations, E-CRISP identified potential off-target cleavage sites at three loci based on the number of mismatches allowed in the target sequence and 2 bp upstream of the double-strand break (DSB). The three potential off-targets were identified for TS1 in chromosomes 1, 5 and 7, but sequencing of these loci revealed no evidence of off-target mutations. No off-target sites were predicted for TS2 and TS3.

Table 2 Characteristics of the six mutated lines generated in this study and the irradiation mutants KUR and Musa, showing the DNA-level changes and the effect on the GBSSI protein.

Line	Mutation	Protein changes	Strategy
Line 1	1bp changed	**Missense mutation**	Wild-type Cas9
Line 2	55bp deletion	**Change the frameshift**	Cas9D10A nickase
Line 3	28bp Insertion	**Change the frameshift**	Cas9D10A nickase
Line 4	56bp deletion	**Change the frameshift**	Cas9D10A nickase
Line 5	246bp deletion	**Deletion**	Cas9D10A nickase
Line 6	2bp changed	**Synonymous mutation**	Cas9D10A nickase
KUR	Loss of the gene	**No protein**	Neutron irradiation
Musa	23bp duplication	**Change the frameshift**	Gamma irradiation

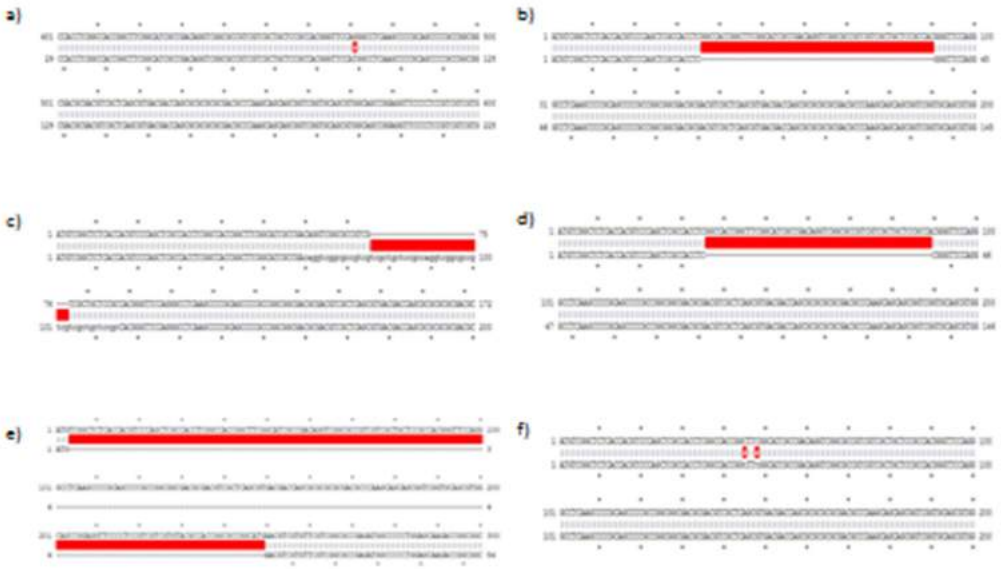


Figure 3 The gRNA target sites and sequencing results showing the nature of our six *Wx* mutants.

Structural comparisons and phylogenetic analysis of protein sequence

In order to investigate potential changes at the protein level, we translated the mutant sequences (**Figure 4**) and generated 3D models using the SWISS-MODEL program (**Figure 5**). The line 1 missense mutation resulted in the amino acid substitution Q33H, lines 2–4 featured indels and concurrent frameshifts that generated a severely truncated and nonfunctional protein, the indel in line 5 removed the N-terminal portion of the protein without affecting the catalytic site, and line 6 was a synonymous substitution with no effect on protein structure.

a)
MSALTTSQLATSATGFGIADRSAPSSLLRHGFQGLKPRSPAGGDATSLSVTTSAR
ATPKQQRSVQRGSRRFPSVVVYATGAGMNVVFVGAEMAPWSKTGGLGDVLG
GLPPAMAANGHRVMVISPRYDQYKDAWDTSVVAEIKVADRYERVRFFHCYK
RGVDRVFIDHPSFLEKVWGKTGEKIYGPDTGVDYKDNQMRFSLLCQAALEAPR
ILNLNNNPYFKGTYGEDVVFVCNDWHTGPLASYLKNNYQPNGIYRNAKVAFCI
HNISYQGRFAFEDYPELNLSERFRSSFDFIDGYDTPVEGRKINWMKAGILEADR
VLTVSPYYAEELISGIARGCELDNIMRLTGITGIVNGMDVSEWDPSKDKYITAK
YDATTAIEAKALNKEALQAEAGLPVDRKIPLIAFIGRLEEQKGPDVMAAAIPEL
MQEDVQIVLLGTGKKKFEKLLKSMEEKYPGKVRAVVKFNAPLAHLIMAGADV
LAVPSRFEPCGLIQLQGMRYGTPCACASTGGLVDTVIEGKTGFHMGRLSVDCK
VVEPSDVKKVAATLKRAIKVVGTPAYEEMVRNCMNQDLSWKGPAKNWENVL
LGLGVAGSAPGIEGDEIAPLAKENVAAP

b)
MSALTTSQLATSATGFGIADRSAPSSLLRHGFHGLKPRSPAGGDATSLSVTTSAR
ATPKQQRSVQRGSRRFPSVVVYATGAGMNVVFVGAEMAPWSKTGGLGDVLG
GLPPAMAANGHRVMVISPRYDQYKDAWDTSVVAEIKVADRYERVRFFHCYK
RGVDRVFIDHPSFLEKVWGKTGEKIYGPDTGVDYKDNQMRFSLLCQAALEAPR
ILNLNNNPYFKGTYGEDVVFVCNDWHTGPLASYLKNNYQPNGIYRNAKVAFCI
HNISYQGRFAFEDYPELNLSERFRSSFDFIDGYDTPVEGRKINWMKAGILEADR
VLTVSPYYAEELISGIARGCELDNIMRLTGITGIVNGMDVSEWDPSKDKYITAK
YDATTAIEAKALNKEALQAEAGLPVDRKIPLIAFIGRLEEQKGPDVMAAAIPEL
MQEDVQIVLLGTGKKKFEKLLKSMEEKYPGKVRAVVKFNAPLAHLIMAGADV
LAVPSRFEPCGLIQLQGMRYGTPCACASTGGLVDTVIEGKTGFHMGRLSVDCK
VVEPSDVKKVAATLKRAIKVVGTPAYEEMVRNCMNQDLSWKGPAKNWENVL
LGLGVAGSAPGIEGDEIAPLAKENVAAP

c)
MSALTTSQLATSGSRASSPAAPPAATRRRSA

d)
MSALTTSQLATSATGFGIADRSAPSSLLRQVGAVVAAPPRVPGPQAPQPRRRRR
DVAQRDDQRARDAQAAAVGAAWQPEVPLRRRVRHRRRHERRVRRRRDGPLE
QDRRPR

e)
MSALTTSQLATSVPGPQAPQPRRRRRDVAQRDDQRARDAQAAAVGAAWQPE
VPLRRRVRHRRRHERRVRRRRDGPLEQDRPR

f)
MSALTTSQLATSATGFGIADRSAPSSLLRHGFQGLKPRSPAGGDATSLSVTTSAR
ATPKQQRSVQRGSRRFPSVVVYATGAGM

MNVVFVGAEMAPWSKTGGLGDVLGGLPPAMAANGHRVMVISPRYDQYKDA
WDTSVVAEIKVADRYERVRFFHCYKRGVDRVFIDHPSFLEKVWGKTGEKIYGP
DTGVDYKDNQMRFSLLCQAALEAPRILNLNNNPYFKGTYGEDVVFVCNDWHT
GPLASYLKNNYQPNGIYRNAKVAFCIHNISYQGRFAFEDYPELNLSERFRSSFDF
IDGYDTPVEGRKINWMKAGILEADRVLTVSPYYAEELISGIARGCELDNIMRLT
GITGIVNGMDVSEWDPSKDKYITAKYDATTAIEAKALNKEALQAEAGLPVDRK
IPLIAFIGRLEEQKGPDVMAAAIPELMQEDVQIVLLGTGKKKFEKLLKSMEEKY
PGKVRAVVKFNAPLAHLIMAGADVLAVPSRFEPCGLIQLQGMRYGTPCACAST
GGLVDTVIEGKTGFHMGRLSVDCKVVEPSDVKKVAATLKRAIKVVGTPAYEE
MVRNCMNQDLSWKGPAKNWENVLLGLGVAGSAPGIEGDEIAPLAKENVAAP

Figure 4 GBSSI predicted protein sequences encoded by each of the six mutated *Wx* alleles.

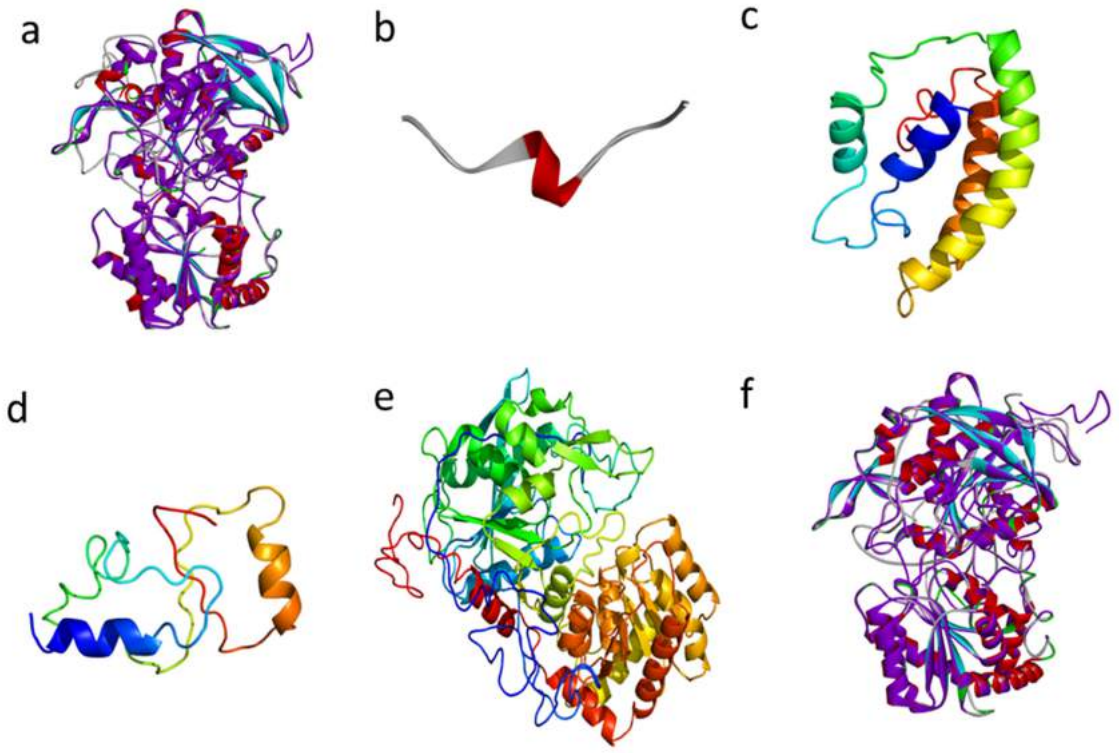


Figure 5 GBSSI predicted protein structures encoded by each of the six mutated *Wx* alleles, superimposed over the wild-type structure

The wild-type protein structure contains binding clefts for ADP and malto-oligosaccharides. Although there was little overall structural change in the line 1 mutant, superimposing the mutant sequence over that of the wild-type protein revealed surface changes that constricted the malto-oligosaccharide pocket and prevented these substrates accessing the catalytic center (**Figure 6**). Furthermore, amino acids 1–77 are essential for the import of GBSSI into starch granules so we compared the sequence of GBSSI enzymes from other plants to determine whether the Q33H substitution removed a functionally critical residue. We constructed a phylogenetic tree from the GBSSI sequences of the japonica and indica subspecies of rice, as well as barley, wheat, tomato, tobacco, cotton, soybean and *Arabidopsis thaliana*, which revealed that Q33 is highly conserved among cereals but not in dicots (**Figure 7**). These data suggest that Q33 may be required for the efficient import of GBSSI in cereals and that the line 1 mutant may therefore suffer from the inefficient import of the enzyme into starch granules.

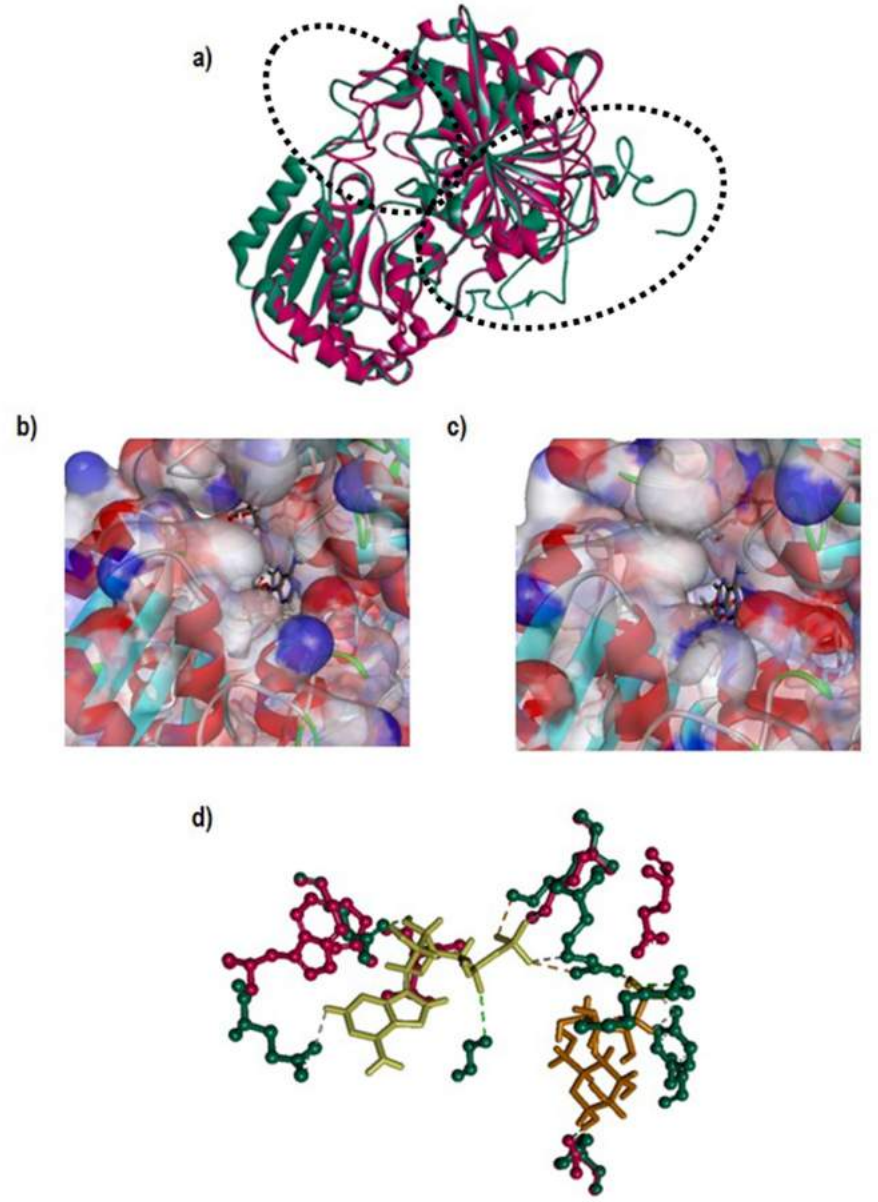


Figure 6 Structure of GBSSI in wild-type rice and mutant line 1. (a) Superimposition of wild-type GBSSI (pink) and the line 1 mutant (green). The principal structural differences are encircled by dashed lines. (b) Wild-type surface model with ADP and malto-oligosaccharide in the catalytic site. (c) Line 1 mutant surface model with ADP and malto-oligosaccharide in the catalytic site. d) Superimposition of wild-type GBSSI (pink) and the line 1 mutant (green) with ADP (yellow) and malto-oligosaccharide (orange) in the catalytic site.

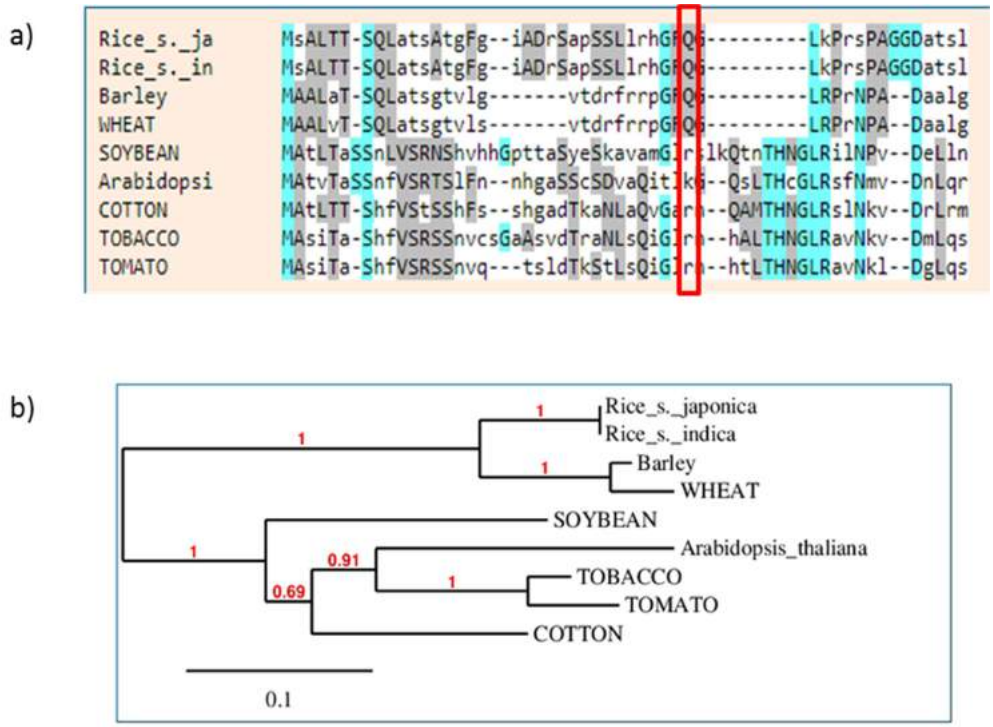


Figure 7 Sequence analysis of the GBSSI protein. (a) Sequence alignment of GBSSI proteins from various monocot and dicots plants. Highly conserved residues are shaded in blue and moderately conserved residues are shaded in gray. The red box highlights residue 33, which is mutated in line 1. (b) Phylogenetic tree generated from the aligned sequences using the Phylogeny.fr server with default parameters. The UniProt accession numbers of the sequences are: *Oryza sativa* subsp. *japonica* (rice) Q0DEV5; *Oryza sativa* subsp. *indica* (rice) P04713; *Hordeum vulgare* (barley) P09842; *Triticum aestivum* (wheat) P27736; *Arabidopsis thaliana* Q9MAQ0, *Glycine max* (soybean) A1YZE0; *Nicotiana tabacum* (tobacco) M9Q2A3; *Gossypium hirsutum* (cotton) D2D315; and *Solanum lycopersicum* (tomato) K4CPX6.

Enzymatic activity

GBSS activity was measured in the T2 seeds of the six mutant lines and compared to wild-type seeds as well as two *Wx* mutants generated by irradiation, namely KUR generated by exposure to neutrons (Yatou and Amano, 1991) and Musashimochi (Musa) generated by exposure to gamma rays (Itoh et al., 1997). As anticipated, there was no significant difference between line 6 and wild-type seeds, given that the mutation had no effect on the protein sequence or structure. In the other five mutant lines, the GBSS activity fell by 61–71% compared to wild-type activity, which was similar to the GBSS activity in Musa and slightly lower than the activity in KUR (**Figure 8a**).

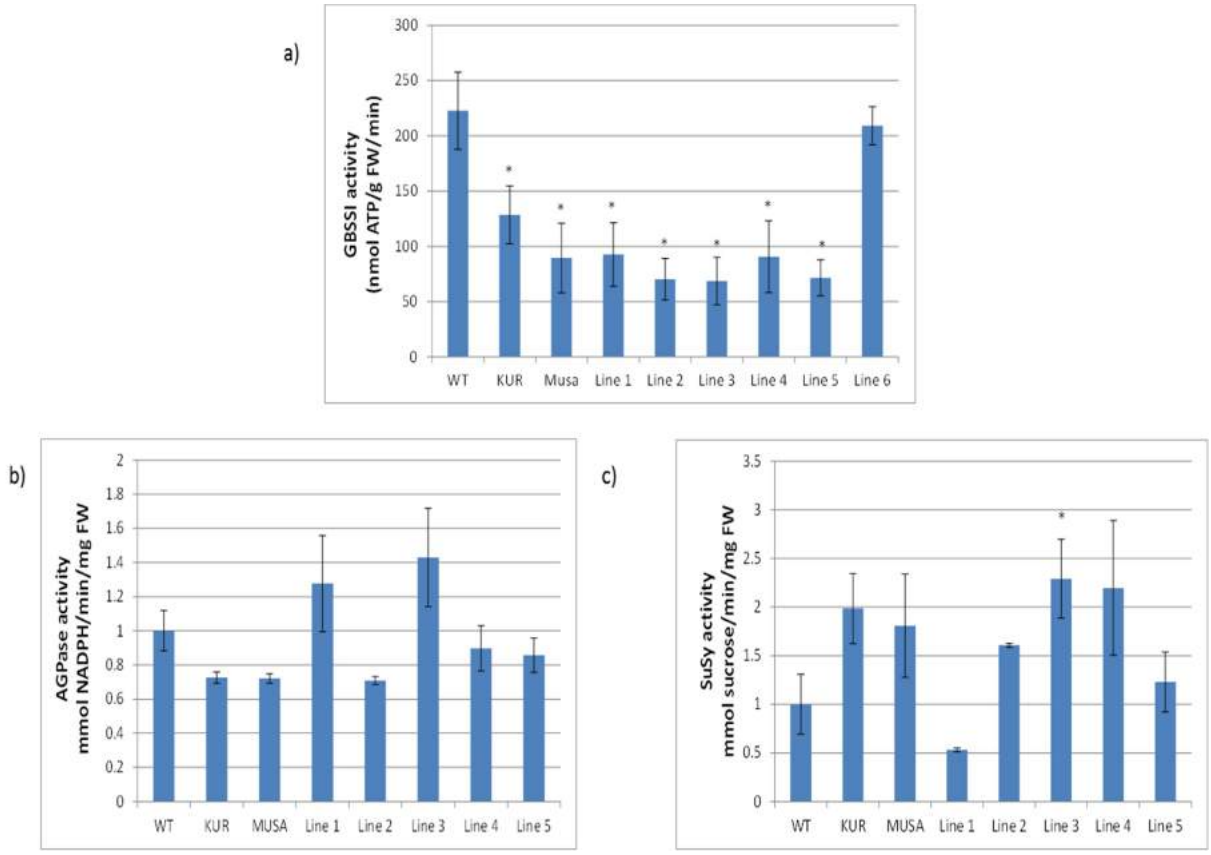


Figure 8 Analysis of enzyme activity in wild-type and mutant rice plants. (a) GBSS1 activity in T2 seeds of wild-type (WT) plants, our *Wx* mutant lines and the two *Wx* irradiation mutants KUR and Musa. (b) AGPase activity the T0 flag leaves of WT and mutant lines. (c) Sucrose synthase activity in the T0 flag leaves of WT and mutant lines. Values are means ± SDs (n = 3 biological replicates, 2 technical replicates for each biological replicate). The asterisk indicates a statistically significant difference between WT and mutant, as determined by Student's t-test (*P < 0.05, **P<0.001).

We also investigated whether the loss of GBSS activity in the mutants affected the activity of AGPase and sucrose synthase, excluding line 6 given that there was no loss in this line. Most of the remaining lines, as well as the KUR and Musa mutants, showed a 10–30% drop in AGPase activity, but in lines 1 and 3 the activity of AGPase increased by 30–40%. Although the trends were clear, these differences were not statistically significant. Conversely, almost all our lines, as well as KUR and Musa, showed a >50% increase in soluble sucrose synthase activity, with a statistically significant >100% increase in line 3. Line 5 showed a less striking increase of 28%, and line 1 showed a 50% decrease compared to wild-type plants, but these differences were only statistically significant in line 3 (**Figure 8 b,c**).

Analysis of starch-related family gene expression

Next, we measured the expression of the rice *Wx* gene in the T0 flag leaves and T1/T2 endosperm and compared the expression of other genes involved in starch biosynthesis (*OsAPL1*, *OsAPL3*, *OsAPL4*, *OsAPS1*, *OsAPS2a/b, OsAPL2, OsSSI, OsSSIIa, OsSSIIb, OsSSIIc, OsSSIIIa, OsSSIIIb, OsSSIVa, OsSSIVb, OsGBSSII, OsBEI, OsBEIIa, OsBEIIb, OsISA1, OsISA2, OsISA3, OsPUL, OsDPE1, OsDPE2, OsPHOH* and *OsPHOL*). We created a heat map of the expression profiles based on percentiles to visualize the most significant changes. The intense red color represents extreme (≥10-fold) downregulation compared to wild-type and the red gradient represents moderate (<10-fold but ≥1.12-fold) downregulation compared to wild-type. The intense green color represents extreme (≥10-fold) upregulation compared to wild-type and the green gradient represents moderate (<10-fold but ≥2-fold) upregulation compared to the wild-type. All other values close to 1-fold (no change) are colored yellow. In leaves (**Supplementary Figure S5**), the relative expression levels of *OsSSI, OsSSIIIa, OsBEI, OsPUL, OsPHOL, OsDPE1* and *OsDPE2* increased significantly except in line 6 where all genes responded in a similar manner to wild-type plants as expected. *OsAPS1, OsAPL3, OsSSIVb, OsPHOH* and *OsGBSSI* were strongly downregulated. In contrast, *OsAPS2a/b, OsBEI, OsAPL4, OsSSIIIa, OsPUL, OsGBSSII, OsPHOH* and *OsPHOL* were strongly upregulated and *OsAPL1, OsSSIIIb, OsSSIVb, OsBEIIa* and *OsGBSSI* were strongly downregulated in the T1 seeds (**Figure 9**). The T2 seeds showed a similar expression profile, as did the KUR and Musa mutants (**Figure 10**).

An analysis of variance for normalized expression on log-transformed data was carried out using Tissue (Leave/Seed); Genotype (Wild-type and separate mutant lines); Gene (Family of genes); Gene_Type (Function of genes) and isoform (different isoforms of the same gene) as independent factors. Gene_Type, Gene (Gene_Type) and Isoform (Gene_Type,Gene) were the most significant factors, although their interactions with tissue were also highly significant (**Table 2**). For the highest order interaction, Tissue*Isoform (Gene_Type, Gene) is the most significant interaction and the data was graphically represented, where Tissue and OsAPL1, OsAPL3, OsAPL4, OsAPS2a, OsDPE1, OsGBSSII, OsSSI, OsSSIIIa, OsSSIIIb and OsSSIVb isoforms were highly significant. Factor Gene (family) was statistically significant except on ISA and Gene_Type was statistically significant (**Figure 11**).

Table 3 Analysis of variance for normalized expression of different gene types, genes, isoforms and genotypes in different tissues on log-transformed data.

Source	DF	Sum of Squares	Semi-partial R^2	Mean Square	F Ratio	Prob > F
C. Total	391	779.27				
Model	145	558.56	71.7%	3.85	4.29	<.0001*
Tissue	*1*	*3.90*	*0.7%*	*3.90*	*4.35*	*0.0381**
Genotype	*6*	*14.92*	*2.7%*	*2.49*	*2.77*	*0.0126**
*Tissue*Genotype*	*6*	*3.46*	*0.6%*	*0.58*	*0.64*	*0.6963*
Gene_Type	*5*	*42.07*	*7.5%*	*8.41*	*9.38*	*<.0001**
*Tissue*Gene_Type*	*5*	*16.65*	*3.0%*	*3.33*	*3.71*	*0.0029**
*Genotype*Gene_Type*	*30*	*30.94*	*5.5%*	*1.03*	*1.15*	*0.2778*
*Tissue*Genotype*Gene_Type*	*30*	*13.31*	*2.4%*	*0.44*	*0.49*	*0.9885*
Gene[Gene_Type]	*3*	*61.05*	*10.9%*	*20.35*	*22.68*	*<.0001**
*Gene*Tissue[Gene_Type]*	*3*	*11.47*	*2.1%*	*3.82*	*4.26*	*0.0059**
*Gene*Genotype[Gene_Type]*	*18*	*30.38*	*5.4%*	*1.69*	*1.88*	*0.0180**
Isoforme[Gene_Type,Gene]	*19*	*251.65*	*45.1%*	*13.24*	*14.76*	*<.0001**
*Isoforme*Tissue[Gene_Type,Gene]*	*19*	*113.05*	*20.2%*	*5.95*	*6.63*	*<.0001**
Error	246	220.71	28.3%	0.90		

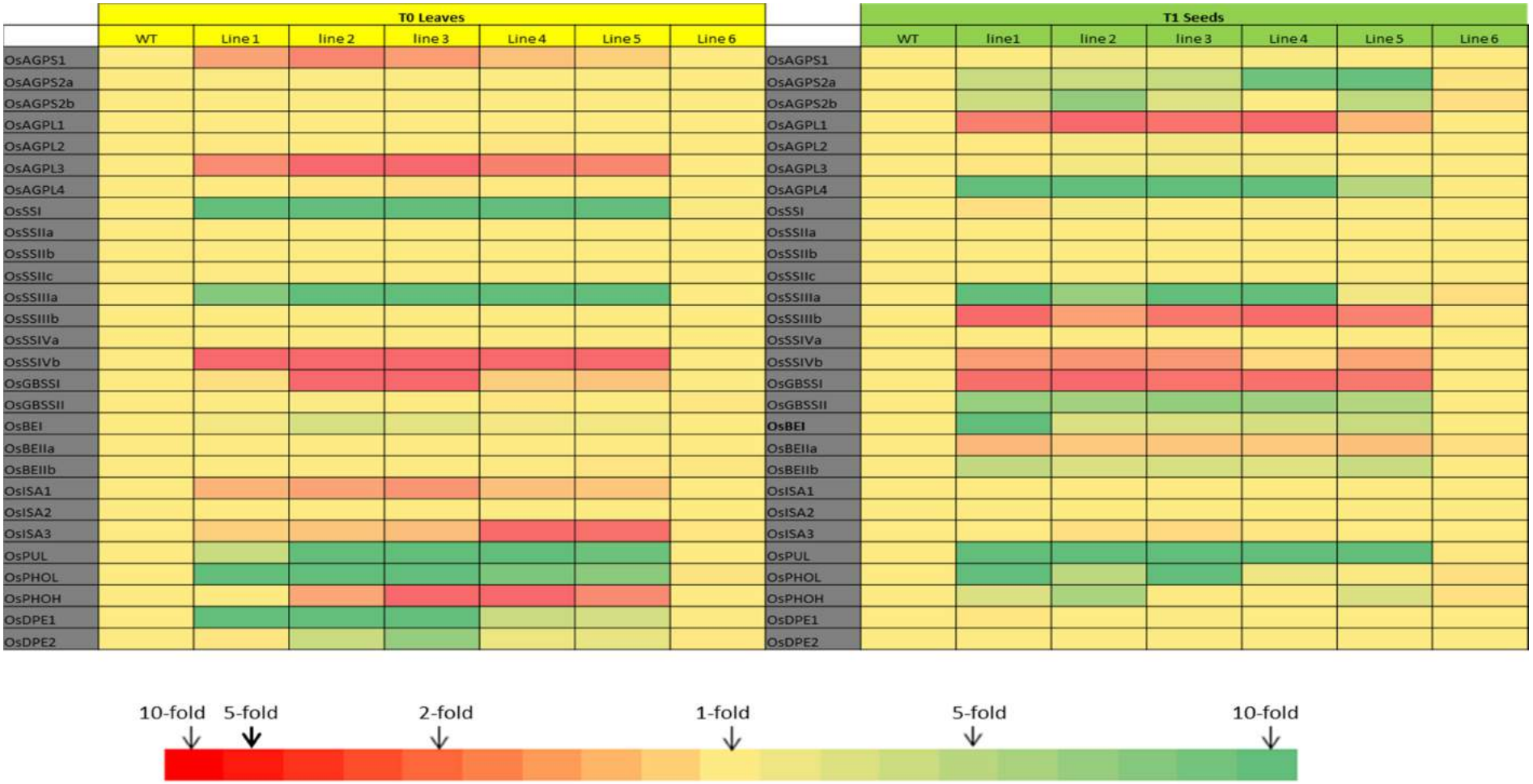


Figure 9 Heat map showing fold-change values for the expression of starch biosynthesis and degradation pathway genes in T0 leaves and T1 seeds of wild-type and mutant rice plants

Figure 10 Heat map showing fold-change values for the expression of starch biosynthesis and degradation pathway genes in T2 seeds in wild-type and mutant rice plants. The red gradient shows increasing degrees of downregulation and the green gradient shows increasing degrees of upregulation, with yellow indicating no change in expression. The red gradient is expanded in the lower ranges because this is where most of the values lie while the green gradient is linear.

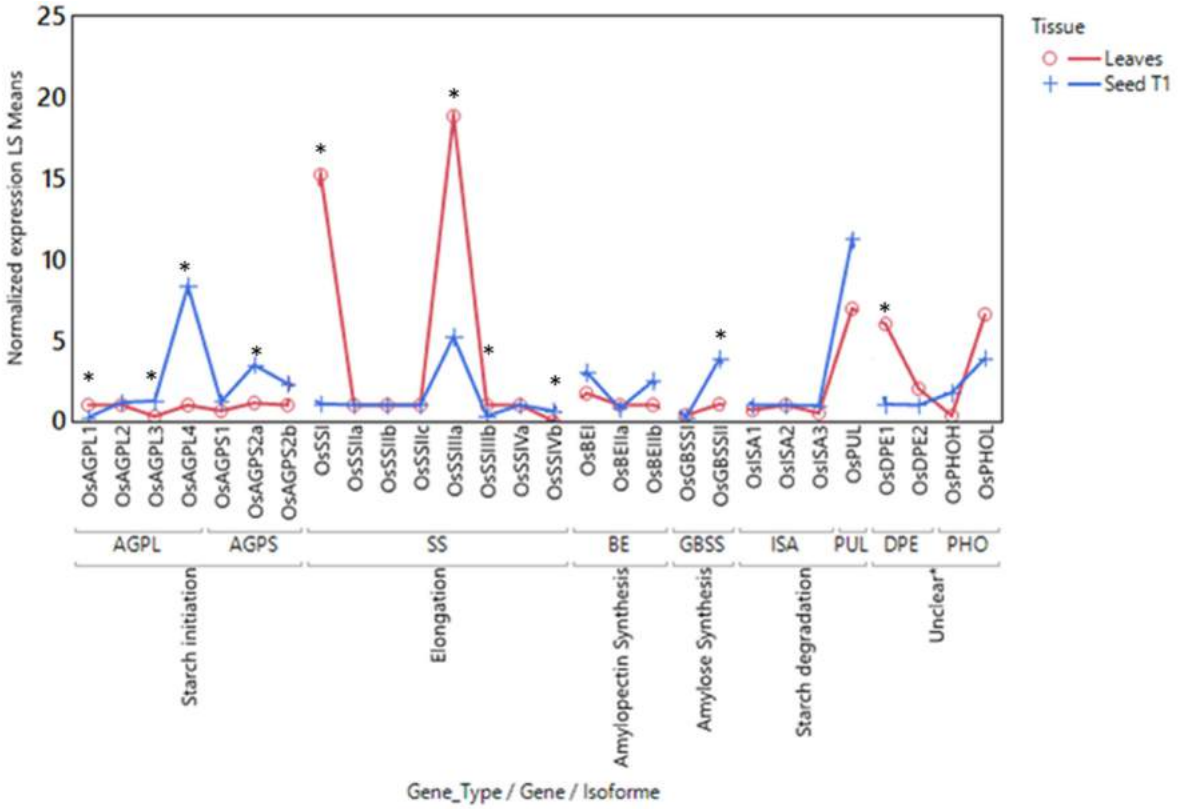


Figure 11. Mean normalized expression for different gene types, genes, and isoforms at different plant tissues. The asterisks indicate a statistically significant difference, as determined by t-test (*P < 0.05) on log-transformed data

Analysis of starch, amylose and soluble sugar levels

The starch content of the T2 seeds was similar to wild-type levels in all six mutant lines and in the KUR and Musa mutants (**Figure 12a**). In contrast, the soluble sugar content was significantly (23–57%) higher than wild-type levels in all the mutant lines except lines 1 and 6, the latter being effectively wild-type due to the synonymous nature of the mutation (**Figure 12b**). Ignoring line 6, the amylose content of the T1 seeds in most of our lines was 40–50% below wild-type levels, but >75% lower in line 2, probably reflecting the homozygous nature of the mutation (**Figure 12c**). Like the Musa mutant, the T1 seeds of line 2 contained less than 5% amylose which can be considered as a Wx phenotype, whereas the other lines (except line 6) contained 8–12% amylose. In T2 seeds, the mutations in lines 1–5 were homozygous and the amylose content was therefore lower than in T1 seeds. Lines 1, 2 and 4 (like Musa) can thus be considered as *Wx* lines because they contained less than 5% of amylose, whereas lines 3 and 5 (like KUR) can be considered as very-low-amylose lines because they contained 6–11.5% amylose (**Figure 12d**).

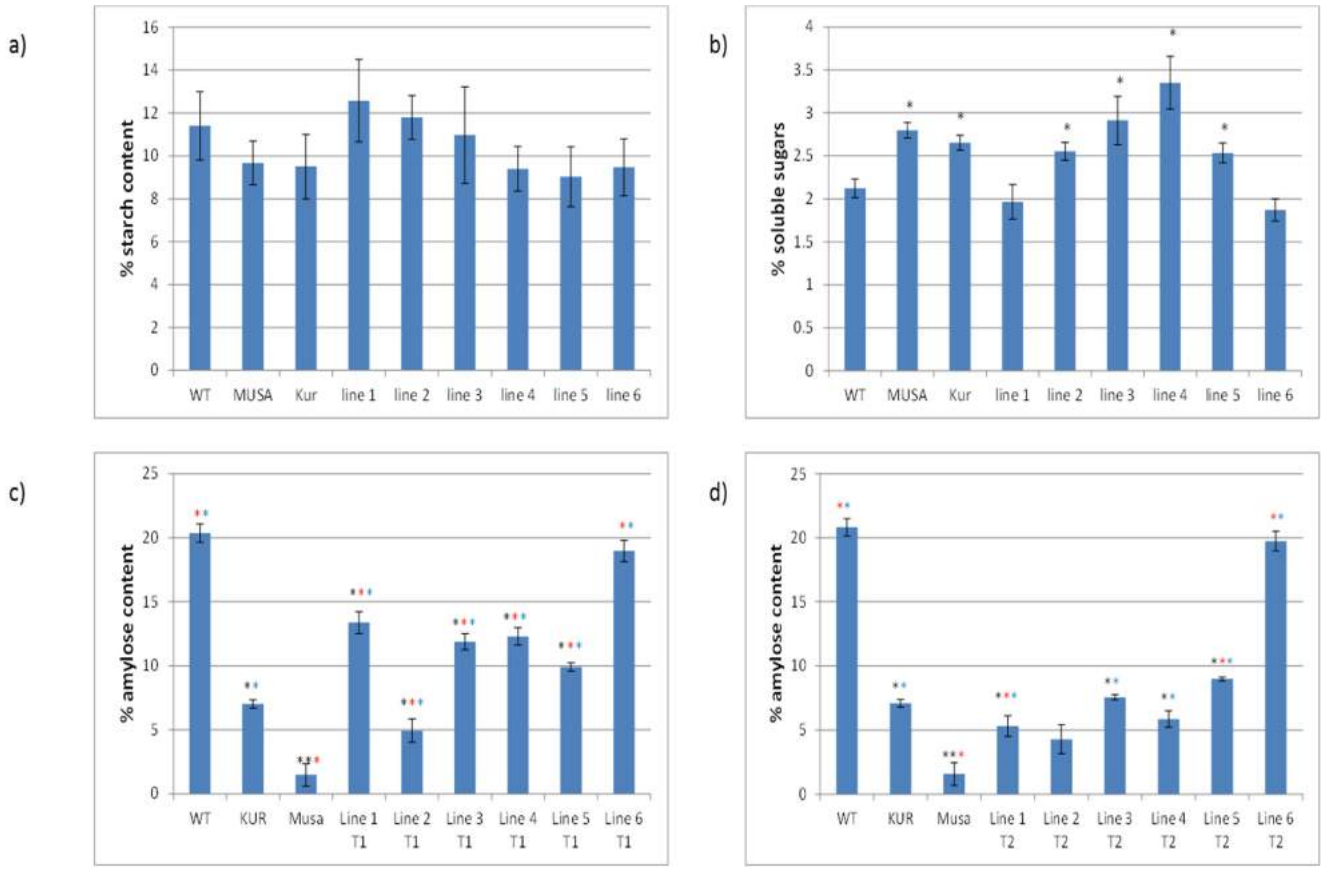


Figure 12 Carbohydrate content of the seeds in wild-type and mutant rice plants. (a) Total starch content of T2 seeds from wild-type (WT) plants, our *Wx* mutant lines and the two *Wx* irradiation mutants KUR and Musa. (b) Soluble sugar content of T2 seeds. (c) Amylose content of T1 seeds. (d) Amylose content of T2 seeds. Values are expressed as means ± SDs (n = 3 biological replicates, 2 technical replicates for each biological). The asterisks indicate a statistically significant difference, as determined by Student's t-test (*P < 0.05; **P < 0.01). Asterisks in black indicate a statistically difference between wild-type and mutant lines. Asterisks in blue indicate a statistically difference between KUR and other lines. Asterisks in red indicate a statistically difference between Musa and other lines.

Phenotype and microscopy

The KUR and Musa *Wx* lines have an opaque seed, characteristic of the Wx phenotype, allowing us to compare our six mutant lines to both wild-type and Wx varieties (**Figure 13**). The T1 seeds of lines 1, 2 and 5 were almost completely opaque, whereas those of lines 3 and 4 were semi-opaque and those of line 6 were indistinguishable from wild-type seeds. Richardson blue staining of fixed T1 mutant seeds showed changes in the cell structure and cell organization in the aleurone layer similar to lines KUR and Musa (**Figure 14**). More detailed morphological analysis by SEM revealed that wild-type starch granules are angular with sharply-defined edges whereas KUR/Musa seeds feature more rounded granules with softer, less-defined edges. In the T1 seeds of our mutant lines, the homozygous line 2 was similar to the KUR/Musa morphology whereas the other lines were more reminiscent of the wild-type morphology (**Figure 15**).

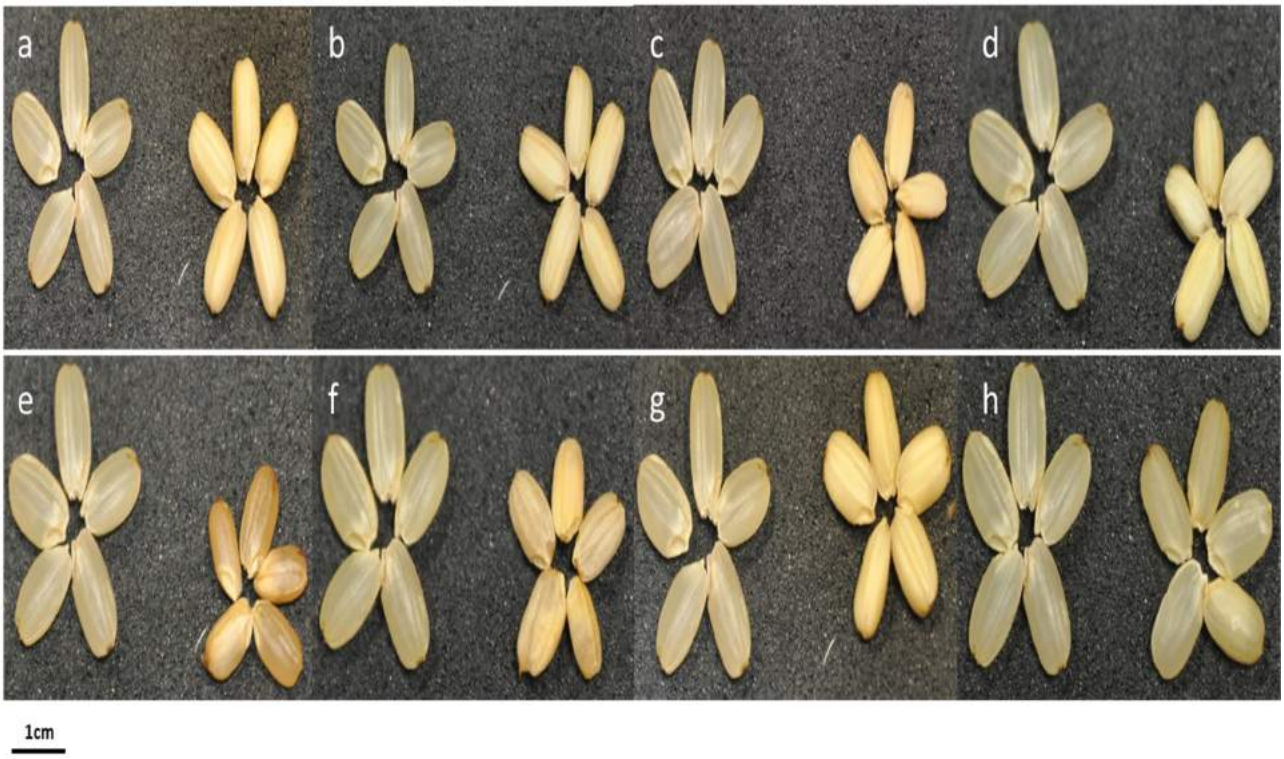


Figure 13 Seed phenotypes of wild-type (WT) plants, our *Wx* mutant lines and the two *Wx* irradiation mutants KUR and Musa. (a) Wild-type vs KUR. (b) Wild-type vs Musa. (c) Wild-type vs line 1. (d) Wild-type vs line 2. (e) Wild-type vs line 3. (f) Wild-type vs line 4. (g) Wild-type vs line 5. (h) Wild-type vs line 6.

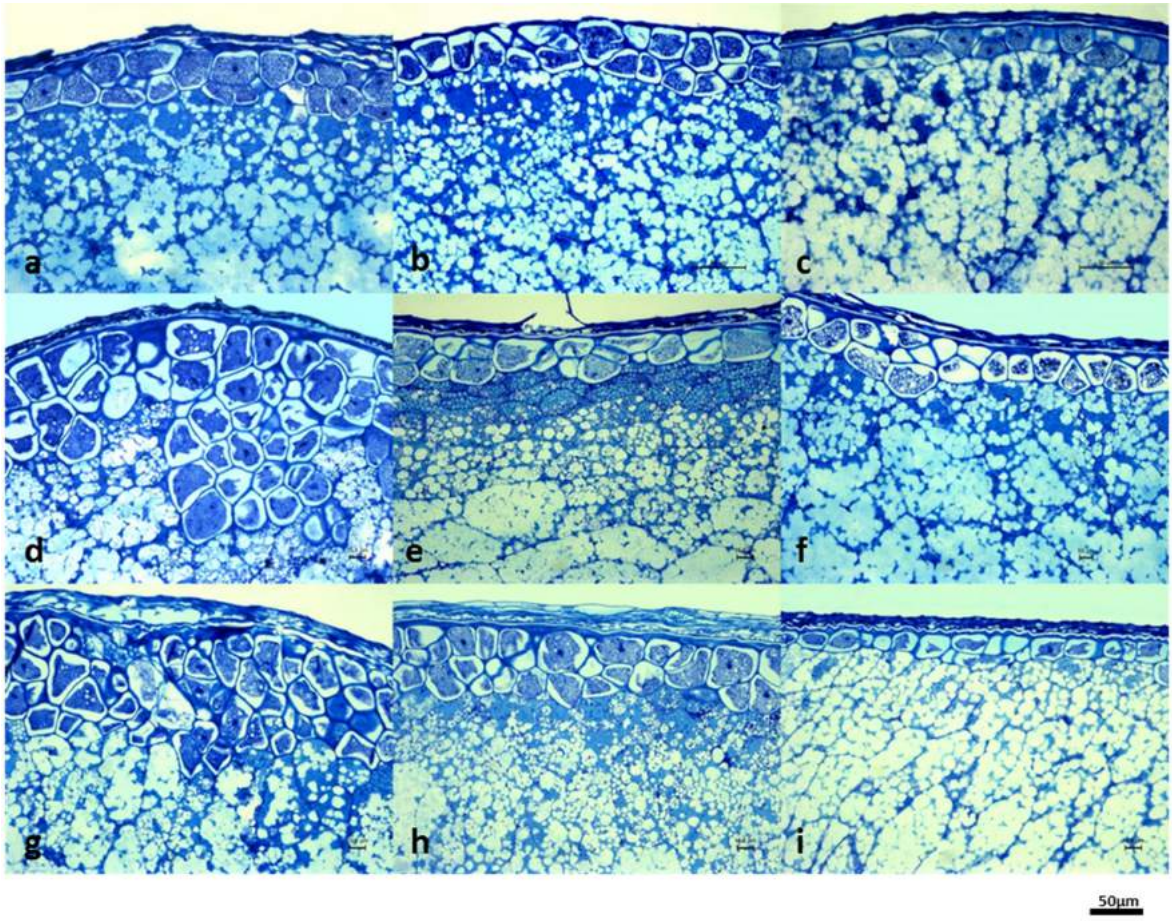


Figure 14 Optical microscopy showing the structure of the aleurone layer in wild-type and mutant rice seeds.

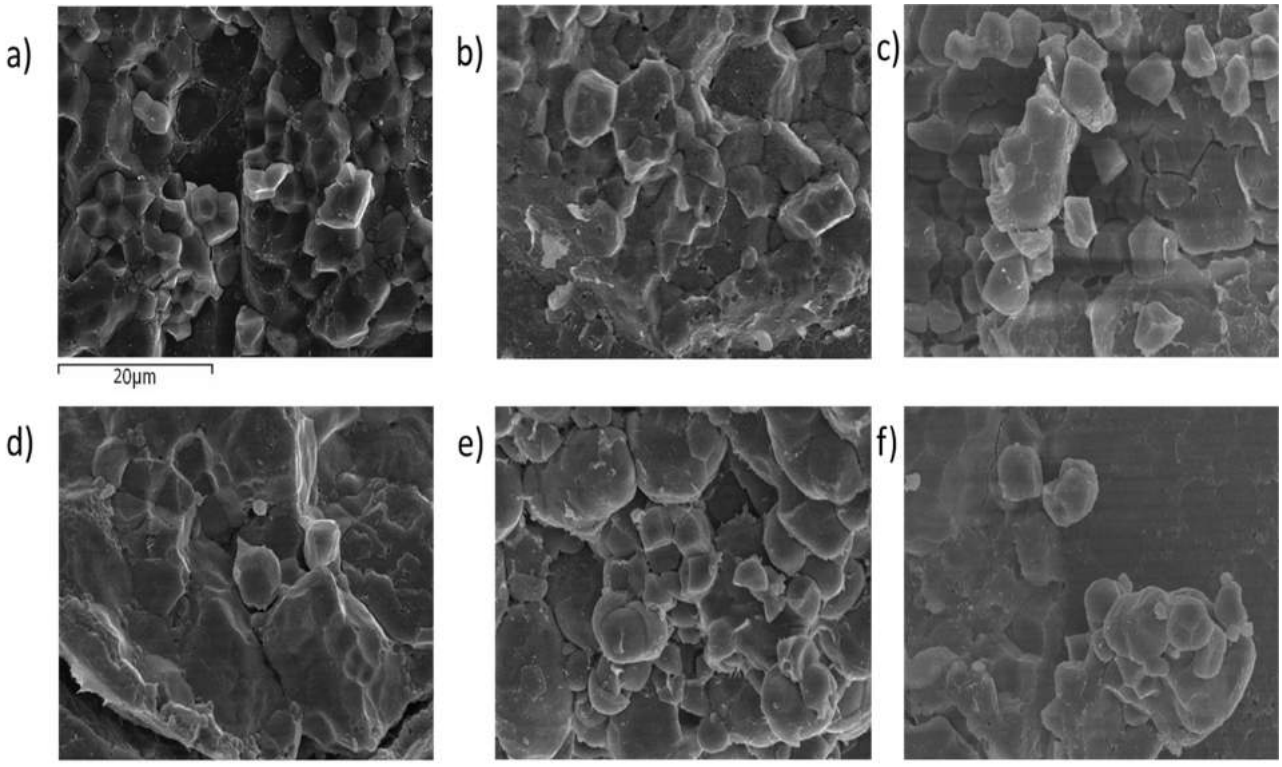


Figure 15 Scanning electron microscopy showing the structure of starch granules in wild-type and mutant rice seeds.

2.5 Discussion

Starch accounts for ~90% of the dry weight of rice grains. It has two components: linear amylose with a small number of long branches, and amylopectin with a large number of short branches. Starch synthesis begins when AGPases catalyze the formation of short chains of ADP-glucose monomers (Li et al., 2017) and continues with the elongation of amylose and amylopectin by granule bound starch synthase (GBSS) and soluble starch synthase (SS), respectively (Ohdan et al., 2005). The frequency of amylopectin branching is controlled by the balance between starch branching and debranching activities (Tang et al., 2016). In rice, there are two GBSS isoforms: GBSSI is expressed in the endosperm and uses malto-oligosaccharides or short amylopectin chains as primers to synthesize amylose (Jeon et al., 2010; Momma and Fujimoto, 2012) whereas GBSSII is expressed in leaves and other vegetative tissues that accumulate transient starch (Tetlow, 2011). GBSSI, encoded by the *Waxy* (*Wx*) locus, is therefore the primary determinant of amylose levels in rice endosperm.

The *Wx* gene is mainly expressed in the endosperm but also in reproductive tissues (Mérida et al., 1999) thus *Wx* rice mutants are easily identified by staining pollen grains with iodine (Itoh et al., 1997). Loss-of-function *Wx* mutants reduce the amylose and protein content of the endosperm (Itoh et al., 2003) resulting in differences in starch gel consistency and also changes in the cellular ultrastructure of the aleurone layer (Tran et al., 2011). Whereas most studies of *Wx* mutants focus on starch quality and agronomic traits, an important but little investigated aspect is the impact on other genes in the starch biosynthesis pathway, as first noted by Zhang et al. (2012).

Waxy mutants have been generated by conventional mutagenesis using chemical mutagens, irradiation or T-DNA/transposons, including the lines KUR and Musashimochi (Musa) induced by neutron irradiation and gamma irradiation, respectively (Yatou and Amano, 1991; Itoh et al., 1997). KUR has an amylose content of 7.5% and Musa has an amylose content of 1.3%, with no changes in the overall amount of starch compared to wild-type plants in either case. Similar phenotypes have been generated by the targeted mutation or knockdown of *Wx* in rice, confirming that GBSSI is the key determinant of amylose levels in the endosperm (Terada et al., 2000; Itoh et al., 2003; Fujita et al., 2006; Tran et al., 2011). More recently, CRISPR/Cas9 has been used to target *Wx*. For example, Ma et al. (2015) targeted three different sites simultaneously. One or two of the sites were mutated in the resulting lines and the amylose content inT1 seeds was reduced from 14.6% to 2.6%. However, these studies have looked solely at the impact on starch and agronomic properties such as seed number and length (Li et al., 2017, Zhang et al., 2018) overlooking impact on other enzymes in the starch biosynthesis pathway.

We had previously used CRISPR/Cas9 to target starch biosynthetic genes active in the endosperm and have observed a profound impact on the wider starch biosynthetic pathway in vegetative tissues. For example, when we targeted *OsAPL2* (an endosperm-specific subunit of AGPase) and induced a partial loss-of-function

mutation, the starch content of the endosperm was reduced but we observed the unexpected ectopic expression of both *OsAPL2* and *OsAPS2b* (also typically endosperm-specific) in the leaves, and compensation for the loss of one functional *OsAPL2* allele by the upregulation of the remaining allele (Pérez et al., 2018). We also observed effects on the broader starch pathway when we targeted *OsBEIIb* using two different sgRNAs with different activity scores and different degrees of conservation with the paralogous gene *OsBEIIa* to confirm the absence of off-target mutations (Baysal et al., 2016). To investigate whether similar effects would occur when we targeted *Wx*, we designed sgRNAs targeting three sites in exon 1 and obtained three different types of mutation: a missense mutation caused by the substitution Q33H resulting in a moderate structural change compared to the wild-type enzyme (line 1), a synonymous substitution caused by two nucleotide exchanges that did not change the sense of the corresponding codon (line 6), and larger indels causing frameshifts and complete loss of function due to early truncation (lines 2–4) or the abolition of protein import to the starch granules (line 5). All lines except line 2 were heterozygous in the T1 generation (line T2 was a homozygous mutant) and all six lines were homozygous mutants in the T2 generation.

The major consequence of mutating the *Wx* gene was the modification of the relative abundance of amylose and amylopectin without changing the overall starch content, so we investigated this phenotype first. In previous studies, the amylose content was reduced when GBSS was targeted (Zhang et al., 2012; 2018) but it was increased when SS was targeted, due to the suppression of amylopectin synthesis (Ryoo et al., 2007; Zhang et al., 2011). We compared the amylose content of our mutants with wild-type plants and the KUR and Musa irradiation mutants. Line 6 showed no difference to wild-type plants as expected because the GBSSI enzyme retained its normal activity. In the T1 generation, the heterozygous seeds of lines 1 and 3–5 had a lower amylose content than wild-type seeds but higher than both irradiation mutants, whereas the amylose content of the homozygous seeds of line 2 was between that of KUR and Musa. In contrast, the homozygous T2 seeds of lines 1–5 all lay between the KUR and Musa mutants, with amylose levels of 4–9%. The mutation of other starch pathway genes increases the abundance of soluble sugars, e.g. as shown for AGPases (Rösti et al., 2007; Tang et al., 2016). The analysis of our mutant lines likewise showed an increase in soluble sugars in lines 1, 2, 4 and 5 but not 3, which may reflect a metabolic bottleneck caused by the mutation or an increase in starch degradation due to feedback regulation, as discussed in more detail below. The KUR and Musa lines also showed higher levels of soluble sugars. The phenotypes described above were concordant with the GBSS activity of the seeds measured directly. Our mutants and the two irradiated lines showed only 42.4–69.16% of the GBSS activity of the wild-type lines.

Mutations in starch biosynthesis genes are often recognized by their characteristic grain phenotype. *APL2* and *APS2* mutants have shrunken seeds (Kawagoe et al., 2005; Tang et al., 2016), mutations in SS or SBEI result in grains with a chalky appearance (Ryoo et al., 2007; Zhang et al., 2011), mutations in *SBEIIb* generate opaque grains (Sun et al., 2017), and mutations in *ISA1* give rise to sugary grains (Wong et al., 2003) in contrast to *ISA2* mutations with no phenotype because the enzyme has negligible activity (Li et al., 2017). Mutating the transcription factor OsZIP58 which regulates starch biosynthesis gives rise to seeds

with a floury-white core (Wang et al., 2013). The phenotype of *PhoL* is temperature dependent, ranging from chalky to shrunken seeds (Satoh et al., 2008). The waxy phenotype that underlies the name of the GBSSI gene reflects the accumulation of amylopectin at the expense of amylose, and is characterized by grains that are white and opaque rather than translucent like wild-type grains (Zhang et al., 2012; 2018). In KUR and Musa, the grains are completely opaque due to the substantial loss of GBSS activity, and the T1 seeds of our lines 1, 2 and 5 were comparable, suggesting a similar degree of GBSS impairment. Wild-type and seeds from line 6 were fully translucent, whereas the remaining lines (3 and 4) were characterized by semi-opaque seeds indicating the wild-type GBSSI allele was more active in these lines or some compensatory mechanism was activated.

The grain phenotypes we observed visually were also reflected by changes in cellular and subcellular organization. The neatly arranged cells of the aleurone layer in the wild-type seeds were disrupted in the mutants, suggesting that cell structure is at least partly dependent on normal starch synthesis. The reason for this became apparent at the subcellular level, where major differences in the shape and structure of starch granules were observed. The wild-type phenotype featured polygonal granules with sharp edges whereas our mutant lines and the KUR/Musa mutants showed rounded and amorphous granules, with greater deviation in the seeds with the lower amylose content. Similar granule structures have been reported in other *Wx* mutants (Liu et al., 2009; Zhang et al., 2018). Furthermore, smaller, rounder, loosely-packed starch granules were observed in *APL2* mutants (Tang et al., 2016). *SSIIA* mutants were characterized by irregular starch granules surrounded by protein bodies in the opaque region of the seed, whereas *SSIIIa* mutants and *SSIIa/SSIIIa* double mutants featured smaller, spherical loosely-packed granules with large air spaces (Zhang et al., 2011). The mutation of OsZIP58 caused a similar phenotype (Wang et al., 2013). The starch granules in *Pho1* mutants were rounder and more amorphous than the wild-type grains (Satoh et al., 2008).

The mutations described above have a dramatic effect on the phenotype of rice grains by perturbing the structure of GBSSI and thus reducing its activity. Amino acids 1–77 of GBSSI correspond to the coiled coil region that interacts with a similar region of the protein PTST and allows both proteins to be imported into starch granules, whereupon the coiled coil region is proteolytically cleaved off (Seung et al., 2015). This means that mutations affecting the PTST protein can phenocopy the loss of GBSSI activity by preventing the import of the enzyme into starch granules (Seung et al., 2015). We targeted the first exon of *Wx* which corresponds to the coiled coil region. In lines 2–4, the resulting frameshift mutation caused the introduction of a nonsense codon and the heavily truncated product was nonfunctional. In contrast, the mutation in line 5 caused part of the coiled coil to be deleted without affecting the remainder of the protein, but still we observed the loss-of-function phenotype because the import of the enzyme was blocked. In line 1, the mutation was a more subtle amino acid exchange, replacing the polar and uncharged glutamine residue at position 33 with the positively charged histidine. Phylogenetic analysis revealed that the glutamine is highly conserved in cereals but not in dicots, suggesting that it may play a role in the

import of GBSSI into cereal starch granules and the mutation may therefore reduce the efficiency of import. This may be one explanation for the partial loss of activity we observed. However, we also found that the tertiary structure of the enzyme was affected. We modeled the tertiary structure of GBSSI in line 1 using the crystal structure of the wild-type enzyme as a template (Momma and Fujimoto, 2012). The wild-type structure features α-helices and β-sheets that form two substrate-binding clefts, one for ADP-glucose and other for malto-oligossacharide precursors. In line 1, the tertiary structure was modified at the N-terminus and C-terminus resulting in a surface change that constricted the malto-oligosaccharide pocket and changed the position of a hydrogen bond, indicating that the mutation may reduce the affinity of the enzyme for its substrate.

One of the key aspects of starch biosynthesis which is rarely investigated in mutational studies is the tight feedback regulation in the pathway, which manifests in the modulation of gene expression among other starch genes when one member is mutated. We therefore analyzed a panel of relevant genes in T0 leaves and T1/T2 seeds relative to wild-type plants and (for seeds only) the two irradiation mutants. We excluded line 6 because it was essentially identical to wild-type plants. In wild-type leaves, the major AGPases are APL1, APL3 and APS2a and the major SEB is SBEI (Ohdan et al., 2005). In our mutant lines, *APL1* and *APL3* were strongly downregulated in the leaves, perhaps due to negative feedback caused by the accumulation of ADP-glucose, whereas *SBEI* was induced, perhaps because the surplus ADP-glucose is used to synthesize amylopectin. In terms of starch degradation the expression of *ISA1* and *ISA3* was downregulated whereas *PUL* was induced, and the *DPE1/2* and *PHOL* genes were upregulated whereas *PHOH*, the major isoform expressed in leaves (Ohdan et al., 2005), was suppressed. The three debranching enzymes are normally expressed in the leaves but whereas *ISA1* has a known role in the maintenance of amylopectin, *ISA2* has no intrinsic activity unless associated with another ISA isoform and the role of *PUL* is unclear (Li et al., 2017), although it may compensate the loss of ISA activity (Jeon et al., 2010). We hypothesized that these changes reflect the capacity of *PUL* to substitute partially for the loss of ISA activity when the absence of GBSSI generates abnormal starch structures that are atypical ISA1 substrates. Similarly, PHOL is more important than PHOH for the maintenance of starch structure. Finally we considered the expression of soluble starch synthases in the leaves. Although *GBSSI* is not expressed in leaves and loss of enzymatic activity in the endosperm should not have any effect, we nevertheless observed an increase in *SSI* and *SSIIIa* expression and a decrease in *SSIVb* expression compared to wild-type plants. *SSI* is strongly expressed in leaves and it forms a complex with *SSIIIa* and other proteins to synthesize short-chain glucans for amylopectin biosynthesis (Crofts et al., 2015). *SSIV* regulates the number of starch granules, and its modulation may reflect a response to the effect of abnormal starch on granule structures (Li et al., 2017).

A different set of starch-related genes was modulated in seeds compared to the leaves. *APS2a/b* and *APL4* were upregulated and *APL1* was downregulated in our mutants and in the KUR and Musa lines. The rice *apl1* mutant showed no change in AGPase activity in the endosperm or in the leaves, and the leaves were reported to contained <5% of normal starch levels but normal levels of soluble sugars (Rösti et al., 2007). These data

suggest that the APL1 subunit is necessary for starch synthesis in leaves but not in non-photosynthetic organs. The increase in *APL4* and *APS2a* expression allows these proteins to form a functional heterotetrameric structure that is not normally found at significant levels in seeds. APS2b is the major AGPase small subunit in endosperm and its activity is needed to form ADP-glucose (Ohdan et al., 2005). We also observed the upregulation of *SBEI* and *SBEIIB*, whose products form a complex to facilitate endosperm starch synthesis (Tetlow et al., 2004), but the downregulation of *SBEIIa* needed to maintain the short-chain content of leaf starch, which does not appear to play a role in endosperm starch synthesis (Nakamura, 2002). Concerning starch degradation *PHO* and *PUL* were upregulated in our mutants; *PUL* debranches pullulan and amylopectin thus its upregulation may help to deal with unusual starch structures by debranch these wear structures (Nakamura, 2002). Finally, we found that the loss of *GBSSI* expression induced a compensatory increase in *GBSSII* expression, which is normally restricted to non-storage tissues. However, given the overall decrease in GBSS enzyme activity we observed in the mutant seeds, this compensation is clearly not enough to restore the normal phenotype. The loss of amylose in the mutant seeds also strongly induced the expression of *SSIIIa* but inhibited *SSIIIb* and *SSIVb*. In wild-type plants *SSIIIa* is more strongly expressed in seeds whereas *SSIIIb* is expressed at the onset of grain formation and declines rapidly thereafter, and SSIVb regulates the number of starch granules (Ohdan et al., 2005). The expression profile in the mutant therefore mirrors but exaggerates the normal situation, with *SSIVb* expression not required because the granule number is already limited in the mutants. The model of variance analysis of normalized expression showed statistically significant differences. Gene_Type, Gene (Gene_Type) and Isoform (Gene_Type, Gene) were the most significant factors, but Genotype was also highly significant. Our statistical analysis further supported our conclusion that the nature of the induced mutations influenced changes in gene expression of other starch biosynthetic genes in a tissue dependent manner.

Given the differences between mutant and wild-type plants in terms of AGPase gene expression, we also measured the overall AGPase activity and the activity of sucrose synthase (SuSy), which provides an alternative pathway for starch synthesis in the absence of AGPase. We found that the opposing changes in the expression levels of different AGPase subunits did not resul in any statistically significant differences in overall AGPase activity between wild-type and mutant lines. However, there was an increase in SuSy activity in the mutants which was statistically significant at least in line 3. This may reflect the accumulation of precursors that cannot be converted into amylose by AGPase and the ability of SuSy to use ADP as substrate for the synthesis of ADP-glucose (Baroja-Fernández et al., 2012).

Our collective results provide a basis to suggest that the overall GBSS activity in lines 3 and 5 was lower than in lines 1 and 4 because there was a smaller compensatory increase in *GBSSII* expression to address the loss of *GBSSI* in the former lines. The increase in AGPase activity in lines 1 and 3 reflected the more profound increase in *APL4* and *APS2a/b* expression and the less severe suppression of *APL1* compared to the other mutants. AGPase activity in line 2 was particularly low due to the severe suppression of *APL1* and relatively weak induction of *APL4*. In lines 4 and 5, the small subunit genes (*APS2a/b*) were upregulated but without a

corresponding increase in *APL4* expression the quantity of the heterotetrameric enzyme could not increase. In lines 2–5, the accumulation of soluble sugars due to the loss of GBSSI activity resulted in an increase in sucrose synthase activity, but in line 1 the weak AGPase activity does not produce enough soluble sugars to induce sucrose synthase, hence the low activity in that line.

2.6 Conclusions

Taking the results of all the experiments together, we propose that the overall GBSS activity in lines 3 and 5 was lower than in lines 1 and 4 because there was a smaller compensatory increase in *GBSSII* expression to address the loss of *GBSSI*. The increase in AGPase activity in lines 1 and 3 reflected the more profound increase in *APL4* and *APS2a/b* expression and the less severe suppression of *APL1* compared to the other mutants. AGPase activity in line 2 was particularly low due to the severe suppression of *APL1* and relatively weak induction of *APL4*. In lines 4 and 5, the small subunit genes (*APS2a/b*) were upregulated but without a corresponding increase in *APL4* expression the quantity of the heterotetrameric enzyme could not increase. In lines 2–5, the accumulation of soluble sugar due to the loss of GBSSI activity resulted in an increase in sucrose synthase activity, but in line 1 the weak AGPase does not produce enough soluble sugars to induce sucrose synthase, hence the low activity in that line. In summary, mutating the first exon of the rice *Wx* gene encoding GBSSI resulted in the expected partial loss of GBSS activity and the corresponding loss of amylose, but also caused the unexpected expression of other downstream starch pathway genes, partly to deal with abundant intermediates and unusual starch structures (amylopectin has more branches and takes up more space than amylose, so the replacement of amylose with amylopectin results in hyperbranched starch that occupies a greater volume than normal). The increase in *GBSSII* expression did not compensate for the loss of *GBSSI*, reflected in the amylose content of the mutant lines. Modifying the peptide signal to enter in starch granules is enough to block *GBSSI* activity without influencing the catalytic center. The mutation of starch pathways genes generates a phenotype that can be used as a first screen to detect mutated lines and provides insight into the complex feedback relationship among genes, enzymes, products and intermediates in the starch biosynthesis and degradation pathways.

2.7 References

Ball, S. G. and Morell, M. K. (2003) From bacterial glycogen to starch: understanding the biogenesis of the plant starch granule. *Ann. Rev. Plant Biol.* 54, 207–233.

Baroja-Fernández, E., Muñoz, F. J., Li, J., Bahaji, A., Almagro, G., Montero, M.,Etxeberria, E., Hidalgo, M., Sesma, T. and Pozueta-Romero, J. (2012) Sucrose synthase activity in the sus1/sus2/sus3/sus4 Arabidopsis mutant is sufficient to support normal cellulose and starch production. *Proc. Natl Acad. Sci. USA*, 109, 321–326.

Bassie, L., Zhu, C., Romagosa, I., Christou, P. and Capell, T. (2008) Transgenic wheat plants expressing an oat arginine decarboxylase cDNA exhibit increases in polyamine content in vegetative tissue and seeds. *Mol. Breeding*, 22, 39–50.

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

Bortesi, L., Zhu, C., Zischewski, J., Perez, L., Bassié, L., Nadi, R., Forni, G., Boyd-Lade, S., Soto, E., Jin, X., Medina, V., Villorbina, G., Muñoz, P., Farré, G., Fischer, R., Twyman, R., Capell, T., Christou, P. and Schillberg, S. (2016) Patterns of CRISPR/Cas9 activity in plants, animals and microbes. *Plant Biotechnol J.* 14, 2203–2216.

Chari, R., Mali, P., Moosburner, M. and Church, G. M. (2015) Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach. *Nature Meth.* 12, 823.

Christou, P., Ford, T. L. and 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 immature zygotic embryos. *Nature Biotechnol*. 9, 957–962.

CISC (Consejo Superior de Investigaciones Científicas) (2016) Método para la determinación “in situ” de actividades enzimáticas relacionadas con el metabolismo del carbono en hojas. ES Patent no 7.915.111.623.106, 2016-05-06.

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

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.F., Guindon, S., Lefort, V., Claverie, J. M. and Gascuel, O. (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucl. Acids Res.* 36, W465–W469.

Farré, G., Sudhakar, D., Naqvi, S., Sandmann, G., Christou, P., Capell, T. and 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 Res.* 21, 1093–1097.

Fujita, N., Yoshida, M., Asakura, N., Ohdan, T., Miyao, A., Hirochika, H. and Nakamura, Y. (2006) Function and characterization of starch synthase I using mutants in rice. *Plant Physiol*. 140, 1070–1084.

Heigwer, F., Kerr, G. and Boutros, M. (2014) E-CRISP: fast CRISPR target site identification. *Nature Meth.* 11, 122.

Hirano, H. Y. (1993) Genetic variation and gene regulation at the *wx* locus in rice. *Gamma-Field Symp.* 24, 63–79.

Hirano, H. Y. and Sano, Y. (1998) Enhancement of *Wx* gene expression and the accumulation of amylose in response to cool temperatures during seed development in rice. *Plant Cell Physiol.*39, 807–812.

Hirano, H. Y. and Sano, Y. (1991) Molecular characterization of the waxy locus of rice (*Oryza sativa*). *Plant Cell Physiol.*, 32, 989–997.

Itoh, K., Nakajima, M. and Shimamoto, K. (1997) Silencing of *waxy* genes in rice containing *Wx* transgenes. *Mol. Gen. Genet.* 255, 351–358.

Itoh, K., Ozaki, H., Okada, K., Hori, H., Takeda, Y. and Mitsui, T. (2003) Introduction of *Wx* transgene into rice *wx* mutants leads to both high-and low-amylose rice. *Plant Cell Physiol*. 44, 473–480.

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

Jiang, D., Cao, W. X., Dai, T. B. and Jing, Q. (2004) Diurnal changes in activities of related enzymes to starch synthesis in grains of winter wheat. *Acta Bot. Sin.* 46, 51–57.

Jiang, W., Zhou, H., Bi, H., Fromm, M., Yang, B. and Weeks, D. P. (2013) Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. *Nucl. Acids Res.* 41, e188-e188.

Jobling, S. (2004) Improving starch for food and industrial applications. *Curr. Opin. Plant Biol.* 7, 210–218.

Juliano, B.O. (1971) A simplified assay for milled-rice amylose. *Cereal Sci. Today*, 16, 334–336.

Kang, T. J. and Yang, M. S. (2004) Rapid and reliable extraction of genomic DNA from various wild-type and transgenic plants. *BMC Biotechnol.* 4, 20.

Kawagoe, Y., Kubo, A., Satoh, H., Takaiwa, F. and Nakamura, Y. (2005) Roles of isoamylase and ADP-glucose pyrophosphorylase in starch granule synthesis in rice endosperm. *Plant J.* 42, 164–174.

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

Li, C., Powell, P. O. and Gilbert, R. G. (2017) Recent progress toward understanding the role of starch biosynthetic enzymes in the cereal endosperm. *Amylase*, 1, 59–74.

Liu, L., Ma, X., Liu, S., Zhu, C., Jiang, L., Wang, Y., Shen, Y., Ren, Y., Dong, H., Chen, L., Liu, X., Zhao, Z., Zhai, H. and Wan, J. (2009) Identification and characterization of a novel *Waxy* allele from a Yunnan rice landrace. *Plant Mol. Biol.* 71, 609–626.

Ma, X., Zhang, Q., Zhu, Q., Liu, W., Chen, Y., Qiu, R., Wang, B., Yang, Z., Li, H., Lin, Y., Xie, Y., Shen, R., Chen, S., Wang, Z., Chen, Y., Guo, J., Chen, L., Zhao, X. and Liu, Y. G. (2015) A robust CRISPR/Cas9 system for convenient, high-efficiency multiplex genome editing in monocot and dicot plants. *Mol. Plant*, 8, 1274–1284.

Maddelein, M. L., Libessart, N., Bellanger, F., Delrue, B., D'Hulst, C., Van den Koornhuyse, N., Fontaine, T., Wieruszeski, J. M., Decq, A. and 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. *J. Biol. Chem.* 269, 25150–25157.

Martin, C. and Smith, A. M. (1995) Starch biosynthesis. *Plant Cell*, 7, 971–985.

Mérida, A., Rodríguez-Galán, J. M., Vincent, C. and Romero, J. M. (1999) Expression of the granule-bound starch synthase I (*Waxy*) gene from snapdragon is developmentally and circadian clock regulated. *Plant Physiol.* 120, 401–410.

Momma, M. and Fujimoto, Z. (2012) Interdomain disulfide bridge in the rice granule bound starch synthase I catalytic domain as elucidated by X-ray structure analysis. *Biosci. Biotechnol. Biochem.* 76, 1591–1595.

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

Nakamura, Y., Yuki, K., Park, S. Y. and Ohya, T. (1989) Carbohydrate metabolism in the developing endosperm of rice grains. *Plant Cell Physiol.* 30, 833–839.

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

Pérez, L., Soto, E., Villorbina, G., Bassie, L., Medina, V., Muñoz, P., Capell, T., Zhu, C., Christou, P. and 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 Res.* https://doi.org/10.1007/s11248-018-0089-7

Ryoo, N., Yu, C., Park, C. S., Baik, M. Y., Park, I. M., Cho, M. H., Bhoo, S. H., An, G., Hahn, T. R. and Jeon, J. S. (2007) Knockout of a starch synthase gene *OsSSIIIa/Flo5* causes white-core floury endosperm in rice (*Oryza sativa* L.). *Plant Cell Rep.* 26, 1083–1095.

Rösti, S., Fahy, B. and Denyer, K. (2007) A mutant of rice lacking the leaf large subunit of ADP-glucose pyrophosphorylase has drastically reduced leaf starch content but grows normally. *Func. Plant Biol.* 34, 480–489.

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. and Nakamura, Y. (2008) Mutation of the plastidial α-glucan phosphorylase gene in rice affects the synthesis and structure of starch in the endosperm. *Plant Cell*, 20, 1833–1849.

Sano, Y. (1984) Differential regulation of *waxy* gene expression in rice endosperm. *Theor. Appl. Genet.* 68, 467–473.

Seung, D., Soyk, S., Coiro, M., Maier, B. A., Eicke, S. and Zeeman, S. C. (2015) PROTEIN TARGETING TO STARCH is required for localising GRANULE-BOUND STARCH SYNTHASE to starch granules and for normal amylose synthesis in Arabidopsis. *PLOS Biol.* 13, e1002080.

Shan, Q., Wang, Y., Li, J. and Gao, C. (2014) Genome editing in rice and wheat using the CRISPR/Cas system. *Nature Prot*. 9, 2395.

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

Sun, Y., Jiao, G., Liu, Z., Zhang, X., Li, J., Guo, X., Du, J., Francis, F., Zhao, Y. and Xia, L. (2017) Generation of high-amylose rice through CRISPR/Cas9-mediated targeted mutagenesis of starch branching enzymes. *Front. Plant Sci.* 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. and Wan, J. M. (2016) ADP-glucose pyrophosphorylase large subunit 2 is essential for storage substance accumulation and subunit interactions in rice endosperm. *Plant Sci.* 249, 70–83.

Terada, R., Nakajima, M., Isshiki, M., Okagaki, R. J., Wessler, S. R. and Shimamoto, K. (2000) Antisense *waxy* genes with highly active promoters effectively suppress *waxy* gene expression in transgenic rice. *Plant Cell Physiol.* 41, 881–888.

Tetlow, I. J., Morell, M. K. and Emes, M. J. (2004) Recent developments in understanding the regulation of starch metabolism in higher plants. *J. Exp. Bot.* 55, 2131–2145.

Tetlow, I. J. (2011) Starch biosynthesis in developing seeds. *Seed Sci. Res.* 21, 5–32.

Thitisaksakul, M., Jiménez, R. C., Arias, M. C. and Beckles, D. M. (2012) Effects of environmental factors on cereal starch biosynthesis and composition. *J. Cereal Sci.* 56, 67–80.

Tran, N. A., Daygon, V. D., Resurreccion, A. P., Cuevas, R. P., Corpuz, H. M. and Fitzgerald, M. A. (2011) A single nucleotide polymorphism in the *Waxy* gene explains a significant component of gel consistency. *Theor. Appl. Genet.* 123, 519–525.

Umemoto, T. and Terashima, K. (2002) Activity of granule-bound starch synthase is an important determinant of amylose content in rice endosperm. *Funct. Plant Biol.* 29, 1121–1124.

Valdez, M., Cabrera-Ponce, J. L., Sudhakar, D., Herrera-Estrella, L. and Christou, P. (1998) Transgenic Central American, West African and Asian elite rice varieties resulting from particle bombardment of foreign DNA into mature seed-derived explants utilizing three different bombardment devices. *Ann. Bot.* 82, 795–801.

Wang, J. C., Xu, H., Zhu, Y., Liu, Q. Q. and Cai, X. L. (2013) OsbZIP58, a basic leucine zipper transcription factor, regulates starch biosynthesis in rice endosperm. *J. Exp. Bot.* 64, 3453–3466.

Wong, K. S., Kubo, A., Jane, J. L., Harada, K., Satoh, H. and Nakamura, Y. (2003) Structures and properties of amylopectin and phytoglycogen in the endosperm of *sugary-1* mutants of rice. *J. Cereal Sci.* 37, 139–149.

Yatou, O. and Amano, E. (1991) DNA Structure of mutant genes in the *waxy* locus in rice. In: *Plant Mutation Breeding for Crop Improvement* volume 2. IAEA, Vienna/FAO, Rome, pp 385–389.

Yoshida, S., Forno, D. A., Cock, J. H. and Gomez, K. A. (1976) Determination of sugar and starch in plant tissue. *Laboratory Manual for Physiological Studies of Rice, third edition.* IRRI, Los Baños, Phillipines, pp 46–49.

Zhang, G., Cheng, Z., Zhang, X., Guo, X., Su, N., Jiang, L., Mao, L. and Wan, J. (2011) Double repression of soluble starch synthase genes *SSIIa* and *SSIIIa* in rice (*Oryza sativa* L.) uncovers interactive effects on the physicochemical properties of starch. *Genome*, 54, 448–459.

Zhang, J., Zhang, H., Botella, J. R. and Zhu, J. K. (2018) Generation of new glutinous rice by CRISPR/Cas9-targeted mutagenesis of the Waxy gene in elite rice varieties. *J. Int. Plant Biol.* 60, 369–375.

Zhang, M. Z., Fang, J. H., Yan, X., Liu, J., Bao, J. S., Fransson, G., Andersson, R., Jansson, C., Aman, P. and Sun, C. (2012) Molecular insights into how a deficiency of amylose affects carbon allocation–carbohydrate and oil analyses and gene expression profiling in the seeds of a rice waxy mutant. *BMC Plant Biol.* 12, 230.

Zhu, C., Bortesi, L., Baysal, C., Twyman, R. M., Fischer, R., Capell, T., Schillberg, S. and Christou, P. (2017) Characteristics of genome editing mutations in cereal crops. *Trends Plant Sci.* 22, 38–52.

CHAPTER 3

The expression of an ectopic mevalonate pathway in rice plastids results in qualitative and quantitative changes in diverse metabolites.

3.1 Abstract

The expression of a partial or complete ectopic mevalonate (MVA) pathway in rice plastids had a profound effect on a broad range of metabolites. Expression of the ectopic pathway also impacted the native MVA and MEP pathways due to the complex feedback mechanisms regulating these pathways. To investigate the wider impact of the genetic interventions we introduced different MVA gene combinations into rice endosperm. We found that expression of ectopic genes increased plastidial terpenoids and affected a number of plant hormones in seeds. Many endogenous MVA and MEP pathway genes were affected, with *OsDXR, OsCMS, OsHDS, OSIDS1 and OsIDI1* strongly downregulated and *OsHMGS, OsHMGR* strongly upregulated in seeds. Phytohormone analysis revealed strong changes in abscisic and jasmonic acids and gibberellins. These changes perturbed the growth and other phenotypic characteristics of the plants.

3.2 Introduction

Isoprenoids were discovered in plants and are essential in all organisms. Organisms unable to synthetize isoprenoids are obligate parasites (Boucher and Doolittle, 2000). These molecules constitute the largest family of compounds found in nature with more than 50,000 known structures. Many isoprenoids are useful components for pharmaceutical and agricultural products. All isoprenoids are derived from isoprene units assembled and modified in many different ways (Vranova et al., 2013).

Humans look for faster, more economical and more sustainable ways to produce chemicals that were originally sourced from nature. Metabolic engineering is the direct manipulation of metabolic pathways for the enhancement of particular products or modulation of cellular physiology. There are two different platforms for the production of terpenoids: increasing the production of isoprene or isoprenoids in engineered organisms such as *E. coli* or other simple single-cell or increasing their production in plants or plant cell cultures, where they will accumulate and subsequently extracted (Zhou et al., 2015). In both cases, the objective is to convert the production organism into an efficient factory for the production of useful chemicals (Ward et al., 2018).

There are two pathways involved in the biosynthesis of isoprenoids (**Figure 1**). The MVA and MEP pathways produce the common precursors, isopentenyl (IPP) and dimethylallyl diphosphate (DMAPP). Most organisms use one of the two pathways. However, plants are able to use both pathways (Vranova et al., 2012). The MVA pathway is present in archaebacterial, some gram-positive bacteria, yeasts and animals,

while most gram-negative bacteria, cyanobacteria and green algae use the MEP pathway for the biosynthesis of terpenoids (Disch et al., 1998; Kovacs et al., 2002).

IPP and DMAPP produced by the MVA pathway are used for the synthesis of isoprenoids in the cytosol and mitochondria; whereas IPP and DMAPP produced by the MEP pathway are used for the synthesis of isoprenoids in plastids. There is a cross-flow of IPP-DMAPP between both pathways (Berthelot et al., 2012).

Precursors for cytosolic and mitochondrial isoprenoids are synthesized via the MVA pathway. The equilibrium between IPP/DMAPP is controlled by IPP isomerase. Mutations in this pathway cause male sterility and arrest embryo development. MVA pathway catalyses the conversion of acetyl-CoA into IPP by 6 enzymatic reactions (Vranova et al., 2013).

Normally, the first steps of isoprenoid biosynthesis are regulated in all organisms. The main limiting step in the pathway is the negative feedback of mevalonate to HMGR. All known plant HMGR proteins bind to the endoplasmic reticulum (ER) with their catalytic site facing the cytosol. The membrane spanning domain in the part of the protein where regulatory motifs are located are regulated by ER-membrane-anchored transcription factors (SREBP). The pathway is also negative regulated by the ratio of AMP/ATP by AMP-activated protein kinase (AMPK); and dephosphorylation restores enzyme activity. HMGR is regulated by the accumulation of sterols where the protein is ubiquitinated and targeted to the proteasome degradation pathway (Burg and Espenshade, 2011). Light is one of the major environmental signals for the differential accumulation of metabolites in the MVA pathway. MVA pathway genes are mostly expressed during the night (light has a negative effect on transcription) in seedlings and in mature plants (in roots, seeds and flowers) (Ghassemian et al., 2006).

Both pathways are regulated by several environmental signals and conditions such as osmotic stress, dehydration, high or low temperature, UV radiation, bacterial pathogens, herbivory, fungal elicitors, wounding and mycorrhiza (Cordoba et al., 2009).

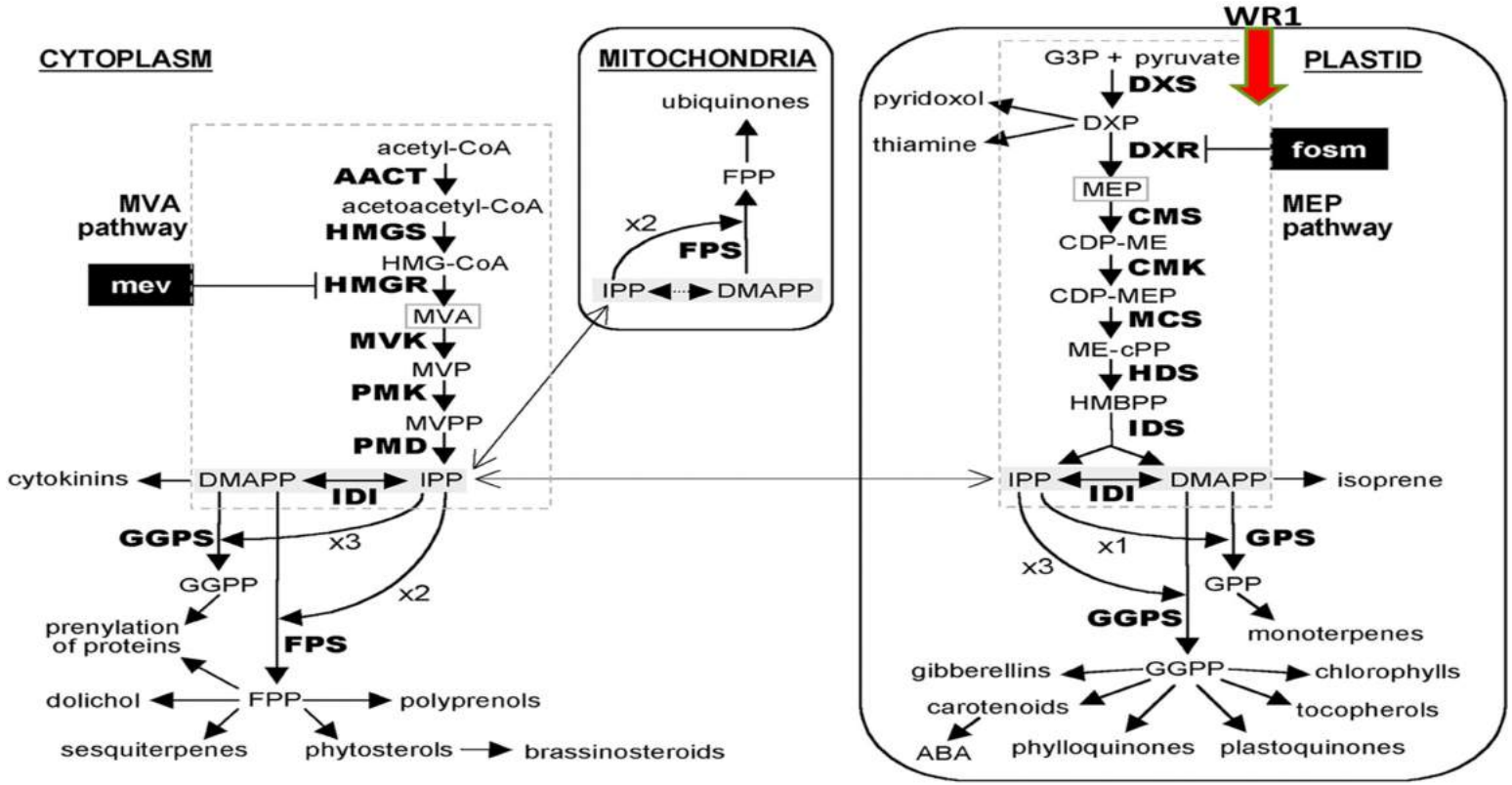


Figure 5. Steps in the terpenoids biosynthesis pathway that generate the different components of terpenoids in the plant cell (Rodrıguez-Concepción and Boronat, 2002). Abreviations: AACT (Acyl-coenzyme A-cholesterol acyltransferase); HMGS (3-hydroxy-3-methylglutaryl-CoA synthase): HMGR (3-hydroxy-3-methylglutaryl-CoA reductase); MVK (mevalonate kinase); PMK (5-phosphomevalonate decarboxylase); PMD (mevalonate diphosphate decarboxylase); IDI (Isopentenyl diphosphate isomerase); GGPS (Glucosylglycerol-phosphate synthase); FPS (farnesyl-diphosphate synthase); FPP (Farnesyl diphosphate); GGPP (Geranylgeranyl diphosphate); mev (mevalonate); DMAPP (Dimethylallyl diphosphate); IPP (Isopentenyl pyrophosphate); HMG-CoA (3-hydroxy-3-methylglutaryl-CoA); MVA (mevalonic acid); MVP (mevalonate-3-phosphate); MVPP (mevalonate- 3,5-diphosphate); DXS (1-deoxy-D-xylulose-5-phosphate synthase); DXR (1-deoxy-D-xylulose-5-phosphate reductoisomerase); CMS (4-Diphosphocytidyl-2C-methyl-d-erythritol 4-phosphate synthase); CMK (4-(cytidine-5'-diphospho)-2-C-methyl-d-Erythritol kinase); MCS (2C-methyl-d-erythritol 2,4-cyclodiphosphate synthase); HDS (4-Hydroxy-3-methylbut2-en-yl-diphosphate synthase); IDS (1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate reductase); GPS (Geranyl diphosphate synthase); G3P (glyceraldehyde 3 phosphate); MEP (2C-methylerythritol 4-phosphate); CDP-ME (4-(cytidine-5'-diphospho)-2-C-methyl-d-Erythritol); CDP-MEP (4-Diphosphocytidyl-2C-methyl-d-erythritol 4-phosphate); ME-cPP (2C-methyl-d-erythritol 2,4-cyclodiphosphate); HMBPP ((E)-4-Hydroxy-3-methyl-but-2-enyl pyrophosphate) ; fosm (fosmidomicin); WR1 (WRINKLED 1).

There were some attempts to increase the flux of MVA pathway by increasing the precursor of the pathway Acetyl-CoA (Tippman et al., 2017), and the cofactor NADPH which is needed for HMGR synthesis (Alper et al., 2005; Zhao et al., 2013). Another approach was to improve productivity by resolving bottlenecks of the pathway. HMGR is the most regulated enzyme in the MVA pathway (Tsuruta et al., 2009; Ma et al., 2011) and it is inhibited by both of its substrates (HMG-CoA and acetoacetyl-CoA) (Wang et al., 2012). Subsequent approaches created novel bypass pathways to alleviate bottlenecks by offering new entry points into endogenous pathways (Kumar et al., 2012). Expressing the truncated *HMGR* gene abolishes the gene's strict regulation and removes the bottleneck in the pathway. HMGR catalytic activity is located in the C terminal while N-terminal is formed by seven α-helices crossing the endoplasmic reticulum membrane with several sites regulating protein activity. AtHMGR protein contains the C-terminal and one α-helix in N-terminal, eliminating the phosphorylating regulation sites of the protein (Vranova et al., 2013).

An increase of precursors for the ectopic MVA pathway has been reported by overexpressing the Arabidopsis thaliana WRINKLED1 transcription factor (WR1) (Cernac and Benning; 2004). WR1 controls the expression of genes for plastidial glycolysis and fatty acid biosynthesis and by controlling expression, availability of Acetyl-CoA increases.

Phytohormones are a group of naturally occurring organic substances which influence physiological processes in plants at low concentrations (Davies, 2010). They play crucial roles in almost all stages of plant growth and development, from embryogenesis to senescence. In addition, they also regulate response of plant to biotic and abiotic stress (Kumar, 2013). Phytohormones have been categorized into several groups based on their structures and physiological functions, including auxins, cytokinins (CKs), abscisic acid (ABA), jasmonates (JAs), salicylates, gibberellins (GAs), ethylene (ET), brassinosteroids (BRs), polyamines, signal peptides and the more-recently-discovered hormones, strigolactones (SLs) (Davies, 2010). Each class of phytohormone has characteristic biological functions.

In the metabolic engineering studies reported here, we employed a multi-gene approach to express *tHMGR*, early part of the pathway (*HMGS*, *tHMGR* and *CrMK*) and the complete ectopic MVA pathway in plastids, and *WR1*. We wished to test the hypothesis that our engineering approaches and the specific use of *tHMGR* when engineered in rice plastids would not only abolish the tight regulation of the pathway, in part or entirely but also increase the pools of IPP and DMAPP. We further hypothesized that such increases in IPP and DMAPP levels might modulate levels of other related metabolites.

3.3 Materials and methods

Construction of transformation vectors

All promoters driving transgene expression were endosperm specific. *HMGS* was from *Brassica juncea*, under the control of the barley D-hordein promoter and the *Agrobacterium tumefaciens* nopaline synthase terminator (T-Nos). *tHMGR* was from Arabidopsis thaliana and it is a deregulated, truncated version of *HMGR* (the catalytic domain of the protein is located at the C-terminal, separated from the membrane spanning domain in which regulatory motifs are located) with an N-terminal transit peptide to direct the protein to the plastids. The gene was under the control of the wheat low molecular weight glutenin promoter and the rice ADPGPP terminator. *CrMK* was from *Catharanthus roseus*, the gene was under the control of the barley D-hordein promoter and the T-Nos terminator. *PMK* was from *Catharanthus roseus*, the gene was under the control of the maize γ–zein gene (GZ63) promoter and T-Nos terminator. *MVD* was from *Catharanthus roseus*, the gene was under the control of the rice prolamin (RP5) promoter and T-Nos terminator. The *WR1* transcription factor was from *Oryza sativa*, the transcription factor was under the control of RP5 and T-Nos terminator. The sequences were rice codon optimized and the vectors were synthesized by Genscript. An extra plasmid was used containing the hygromicin phosphotransferase (*hpt*) selectable marker gene (Sudhakar et al., 1998) was used for the generation of transgenic rice plants.

Rice transformation and recovery of transgenic plants

Seven-day-old mature zygotic embryos (Oryza sativa cv. EYI) were transferred to osmotic medium (MS medium supplemented with 0.3 g/L casein hydrolysate, 0.5 g/L proline, 72.8 g/L mannitol and 30 g/L sucrose) 4 h before bombardment with 10 mg gold particles coated with the transformation vectors. The *tHMGR*, *HMGS:tHMGR:CrMK*, *PMK:MVD* and *WR1* vectors and the selectable marker hpt were introduced at a 3:1 for (*tHMGR;hpt* and *HMGS:tHMGR:CrMK;hpt*) and 3:3:3:1 (*HMGS:tHMGR:CrMK;PMK:MVD;WR1;hpt*) molar ratio as previously described (Christou et al. 1991; Sudhakar al. 1998; Valdez et al. 1998). The embryos were returned to osmotic medium for 12 h before selection on MS medium supplemented with 0.3 g/L casein, 0.5 g/L proline, 30 g/L sucrose, 50 mg/L hygromycin and 2.5 mg/L 2,4 dichlorophenoxyacetic acid in the dark for 2–3 weeks. Callus was maintained on selective medium for 6 weeks with sub-culturing every 2 weeks as described (Farré et al. 2012). Transgenic plantlets were regenerated and hardened off in soil.

Confirmation of the presence of MVA pathway genes

Genomic DNA was isolated from the callus and leaves of regenerated plants by phenol extraction and ethanol precipitation (Bassie et al. 2008; Kang and Yang 2004). The presence of *HMGS, tHMGR, CrMK, PMK, MVD* and *WR1* sequence was confirmed by PCR using the primers listed in **Table 1**.

Table 4. Primers for detection of exogenous MVA genes by PCR and expression by real-time PCR

Gene	Forward Sequence	Reverse sequence
BjHMGS	5′- CATCTACTTCCCGCCAACCTG 3′	5′- TAGCCGCTTCATCAATGCTGG-3′
AtHMGR	5′- GGTTGCGTCCACGAATAGAGG-3′	5′- TTCCCTGCGTGAGCGTTGA-3′
CrMK	5′- CCTTTTGCGTGGCTTTGTCAG -3′	5′- AAACGCCCATGCAGAGCAAA-3′
CrPMK	5′- GCAGATGTCCAGGGAGACGATGT-3′	5′- ACATGATCGCGCAGACAGCC-3′
CrMVD	5′- GAAGGGCATCAAGATCACGAAGA-3′	5′- ATCCTCATCGCCGTGGTGTCCTC-3′
OsHMGS1	5′- GCCTACGCCTTCCTCCCAAT-3′	5′- ATGCCCACGTCCTTCCTCTC-3′
OsHMGS2	5′- GGGATGGACGCTACGGTCTT-3′	5′- TAGCAGCAGCACCACCTGTT-3′
OsHMGS3	5′- TTGTTGCCTCCTGGGACGTT-3′	5′- GATCTCCTCGTCGGCCTTCC-3′
OsHMGR1	5′-A GCTCTTGCCTCTGGTCACCT-3′	5′- GATGCGAAGGGCTGCTCAAG-3′
OsHMGR2	5′- CTCAGCTTCTTCGGCATCGC-3′	5′- GCGGGAGTCGATCAGGAAGT-3′
OsHMGR3	5′- GCGTGCAGAATGTGCTGGAT-3′	5′- CGCGCCCTTCGATCCAATTC-3′
OsMK1	5′- CACTAGTGGATCGCGACCGT-3′	5′- TGGCCTGCACAATGCAAACT-3′
OsMK2	5′- TTGCAGAGCTGGAGTCGCAT-3′	5′- GAGCAGCCTCCTTGGCAAAC-3′
OsMK3	5′- CGGCCTTTGGGTTGTTGCTT-3′	5′- TGCTCGCCGGCAAGGATAAT-3′
OsPMK1	5′- ACTTGGGTCATCAGCTGCCA-3′	5′- GATCTCGTCCAGCTGCGTTG-3′
OsPMK2	5′- GCGCTTCTACGCCATCGTC-3′	5′- CGGGAGAGCTGAGGAGAGGT-3′
OsMVD1	5′- CCGTCAACGACAGCATCAGC-3′	5′- ATCTCCTTGCCGTTGAGCCA-3′
OsMVD2	5′- CATGTGGCTCAACGGCAAGG-3′	5′- TTGAGCACGCTTCCGGATCT-3′

OsDXS1	5′-CTCAAGGGAGGGAAGAACAA-3′	5′-ACACCTGCTTGTTGTCGTTG-3′
OsDXS2	5′-TGTTGTGGAGCTCGCTATTG-3′	5′-TCCTCCCACCTAGATCCCTT-3′
OsDXS3	5′- ACCTCCTCGGGAAGAAGAAC-3′	5′- GAGGGACACCTGCTTGTTGT-3′
OsDXR	5′- GGTGACCTCTGAGCAATA-3′	5′- TCCACCCACAATCTATCA-3′
OsIPPI1	5′-AATGCTGCTCAGAGGAAGC-3′	5′-GGACGATGAACAGCAGGTA-3′
OsIPPI2	5′-GCAGACCTTTGCTGAAGTAAC-3′	5′-TTTAGTGGACGAACAGGACA-3′
OsHDR1	5′-GCTGAATGAGAAGAAGGTGC-3′	5′-AAAGGAAGCAGTGGCAAC-3′
OsHDR2	5′-GAGTAGGCGTTGTTGTGAATCAAAC-3′	5′-CCTTTCCTGAGTAGCATTGC-3′
OsHDS	5′-GGGTTGCATTGTCAATGG-3′	5′-CTGAATCAAGGCGTCAGT-3′
OsMDS	5′-TCGTCGGTGTTCATGAGG-3′	5′-TTGTCTCCTTGAATGGGC-3′
OsMCT	5′-GACTTGAGGTCACTGATG-3′	5′-AGCAAGTCATCATCAGGAGTC-3′
OsCMK	5′-ATGATGATGACTACAAGG-3′	5′-TCAGTCAGAAACTGATGC-3′
OsActin	5′- CATCGCCCTGGACTATGACC-3′	5′-CGGTGTGAGCTGCACGAACGA-3′

Gene expression analysis

Total seed RNA was isolated using the RNeasy Plant Mini Kit (Qiagen) and DNA was digested with DNase I from the RNase-free DNase Set (Qiagen). Total RNA was quantified using a Nanodrop 1000 spectrophotometer (Thermo Fisher Scientific) and 2 µg total RNA was used as template for first strand cDNA synthesis with Quantitech reverse transcriptase (Qiagen) in a 20-µl total reaction volume, following the manufacturer's recommendations. Real-time qRT-PCR was performed on a BioRad CFX96 system using 20-µl mixtures containing 5 ng cDNA, 1x iQ SYBR green supermix and 0.5 µM forward and reverse primers. The ectopic transgenes BjHMGS, AtHMGR, CrMK, CrPMK, CrMVD and OsWR1 and the native MVA and MEP genes, OsHMGS, OsHMGR, OsMK, OsPMK, OsMVD, OsDXS, OsDXR, OsCMS, OsCMK, OsMCS, OsHDS, OsIDS and OsIDI genes cDNAs were amplified using appropriate primers (Table 1). Serial dilutions of cDNA (80–0.0256 ng) were used to generate standard curves for each gene. PCR was performed in triplicate using 96-well optical reaction plates. Values represent the mean of three biological replicates ± SE. Amplification efficiencies were compared by plotting the ΔCt values of different primer combinations of serial dilutions

against the log of starting template concentrations using the CFX96 software. The rice housekeeping OsActin1 (ABF98567.1) was used as an internal control.

Metabolomic analysis

Carotenoids extraction and sample preparation.

Dried seed were ground to a fine powder using a mortar and pestle. 30 mg of the dried powder was then used for carotenoid extraction and analysis. Extractions were carried out following a modification of the Bligh and Dyer method (Bligh and Dyer., 1959). Briefly, 400 µl of methanol and 800 µl of chloroform were added to the dried material and shake it for 1 hr in dark conditions. Subsequently, 400 µl of water was added to the mixture and centrifuged at top speed for 5 min to allow phase separation. The organic layer (hypophase) containing carotenoids was collected and dried under N2 stream. Dried extracts were then stored at -20° C until chromatographic analysis. For UPLC analysis dried extracts were dissolved in 50 µl of ethyl acetate and centrifuged at maximum speed for 10 min to clean up the samples prior to analysis. 30 µl of clear supernatant was transferred to glass vials with glass inserts and 3 µl used as injection volume.

Chromatographic Analysis

Carotenoids were separated and identified by ultrahigh performance liquid chromatography with photodiode array detection. An Acquity ultrahigh performance liquid chromatography system (Waters) was used with an Ethylene Bridged Hybrid (BEH C18) column (2.1 × 100 mm, 1.7 µm) with a BEH C18 VanGuard precolumn (2.1 × 50 mm, 1.7 µm). The mobile phase used was A, methanol/water (50/50), and B, acetonitrile (ACN)/ethyl acetate (75:25). All solvents used were HPLC grade and filtered prior to use through a 0.2-µm filter. The gradient was 30% A:70% B for 0.5 min and then stepped to 0.1% A:99.9% B for 5.5 min and then to 30% A:70% B for the last 2 min. Column temperature was maintained at 30°C and the temperature of samples at 8°C. Online scanning across the UV/visible range was performed in a continuous manner from 250 to 600 nm, using an extended wavelength photo diode array detector (Waters). Carotenoids were quantified from dose–response curves. The HPLC separation, detection, and quantification of carotenoids, tocopherols, and chlorophylls have been described in detail previously (Fraser et al., 2000).

Phytohormone analysis

The main classes of plant hormones, cytokinins (trans-zeatin, tZ, zeatin riboside, ZR and isopentenyl adenine, iP), gibberellins (GA1, GA3 and GA4), indole-3-acetic acid (IAA), abscisic acid (ABA), salicylic acid

(SA), jasmonic acid (JA) and the ethylene precursor 1-aminocyclopropane-1-carboxylic acid (ACC) were analysed (Albacete et al., 2008) with some modifications. Briefly, 0.1 g of fresh plant material was homogenized in liquid nitrogen and dropped in 0.5 ml of cold (-20°C) extraction mixture of methanol/water (80/20, v/v). Solids were separated by centrifugation (20,000 g, 15 min, 4ºC) and re-extracted for 30 min at 4ºC in additional 0.5 ml of the same extraction solution. Pooled supernatants were passed through Sep-Pak Plus †C18 cartridges (SepPak Plus, Waters, USA) to remove interfering lipids and part of plant pigments, and evaporated at 40ºC under vacuum to near dryness. The residue was dissolved in 1 ml methanol/water (20/80, v/v) solution using an ultrasonic bath. The dissolved samples were filtered through 13 mm diameter Millex filters with 0.22 µm pore size nylon membrane (Millipore, Bedford, MA, USA).

Ten µl of filtrated extract were injected in a U-HPLC-MS system consisting of an Accela Series U-HPLC (ThermoFisher Scientific, Waltham, MA, USA) coupled to an Exactive mass spectrometer (ThermoFisher Scientific, Waltham, MA, USA) using a heated electrospray ionization (HESI) interface. Mass spectra were obtained using the Xcalibur software version 2.2 (ThermoFisher Scientific, Waltham, MA, USA). For quantification of the plant hormones, calibration curves were constructed for each analysed component (1, 10, 50, and 100 µg l-1) and corrected for 10 µg l-1 deuterated internal standards. Recovery percentages ranged between 92 and 95%.

Phenotypic analysis

The leaves of transgenic lines were analysed at 12 weeks of growing in soil to measure different parameters compared to wild-type plants as described below.

Number of leaves was counted. Height of the plants was measured from the beginning of the root until the maximum length of the flag leaf.

Transgenic plants expressing *tHMGR* alone, or the full MVA pathway/WR1 and wild-types were used to measured leaf chlorophyll content with a SPAD meter. Six measurements were performed in the last expanded leaf of each plant, in a total of 10 plants.

Length and maximum width from the last expanded leaf were measured and the area was calculated using the formula:

Leaf area = K x length x width

Where K is the “adjustment factor”. K varies with the shape of the leaf which in turn is affected by the variety, nutritional status and growth stage of the leaf. In the case of rice, K=0.75 for all stages (Tsunoda, 1962; Murata, 1967).

Photos of the plants and leaves expressing *tHMGR*, full MVA-WR1 and wild-type were taken at 12 weeks to compare the height and the state of the leaves (senescence). N=5 plants.

Starch and soluble sugars

Flag leaf samples harvested at 7 pm and seeds were homogenized under liquid nitrogen and extracted in perchloric acid to measure the starch content, or in ethanol to measure the content of soluble sugars. The quantity of each carbohydrate was determined by spectrophotometry at 620 nm (Juliano, 1971; Yoshida et al., 1976). To measure the amylose content, milled rice grains were powdered with a faience pestle and mortar and the powder was transferred to a paper envelope and dried for 1 h at 135 °C. We transferred 100±0.01 mg of dried powder to a conical flask and added 1 ml 95% ethanol and 9 ml 1 M NaOH. The suspension was boiled in a water bath for 10 min, cooled at room temperature for 10 min and then topped up to 100 ml with distilled water. A 5-ml aliquot of the solution was transferred to a 100-ml volumetric flask and mixed with 1 ml 1 M acetic acid, 2 ml 0.2% potassium iodide and 92 ml distilled water. Three amylose solutions (3%, 11.5% and 14%) were prepared for comparison. The starch content was determined by measuring the absorbance at 630 nm in a Unicam UV4-100 UV-Vis spectrophotometer after 30 min.

Microscopy analysis

For light microscopy (LM), semithin sections (2µm) from samples (seed pieces) as above processed, once dried on a slide, where stained with Richardson's blue, covered by a drop of slide mounting medium (DPX Mountant for histology) and a coverslip. The slides were observed under a Leica DM4000 B microscope (Leica Microsystems, Mannheim, Germany) and photographed using a Leica DFC300 FX 1.4-megapixel digital color camera equipped with the Leica software application suite LAS V3.8 (Leica Microsystems).

We used a standard protocol. Thus, seed pieces (~1mm3) of each treatment were fixed in 2.5% (v/v) glutaraldehyde in 0.1 M sodium phosphate buffer (pH 7.2) overnight at 4°C and washed three times with same buffer. They were post-fixed in 1% (w/v) osmium tetroxide in 0.1 M sodium phosphate buffer (pH 7.2) for 1h and washed three times in redistilled water and dehydrated in an alcohol series (30–100%) before embedding in epoxy resin (Epoxi, Electron Microscopy Sciences, Hatfield, PA 19440, USA.) and polymerizing at 60ºC for 48h. Semithin and ultrathin sections of samples were obtained in a Reichert-Jung Ultracut E Ultramicrotome Ultra Microtome 701704 (Scotia, NY, USA).

For transmission electron microscopy (TEM), ultrathin sections (70-90nm) were mounted on Formvar©-carbon coated copper grids, stained with uranyl acetate (10min) and Reynold´s lead citrate (2min) prior to examination using a EM 910 Transmission Electron Microscope (Zeiss, Oberkochen, Germany).

3.4 Results

Recovery and characterization of plants expressing ectopic MVA pathway

We recovered ca: 50 independent transformants and we focused on 10 lines selected for high levels of gene expression.

Gene expression

We measured the expression of the *BjHMGS, AtHMGR, CrMK, CrPMK, CrMVD* and *OsWR1* genes in the T1/T2 endosperm and compared the expression of the endogenous genes involved in MVA pathway and MEP pathway (*OsHMGS, OsHMGR, OsMK, OsPMK, OsMVD, OsDXS, OsDXR, OsCMS, OsCMK, OsMCS, OsHDS, OsIDS* and *OsIDI*). We identified in T1 seeds, over 20 lines expressing the transgenes at high levels. In T2 seeds, we selected 10 lines derived from the three different experiments with the highest expression of the input transgenes to perform all subsequent analysis (**Figure 2**). We created a heat map of the expression profiles based on percentiles to visualize the most significant changes. The intense red color represents extreme (≥10-fold) downregulation compared to wild-type and the red gradient represents moderate (<10-fold but ≥1.12-fold) downregulation compared to wild-type. Expression of *OsDXR, OsCMS, OsHDS, OSIDS1 and OsIDI1* genes from the endogenous MEP pathway were decreased between 20-80% with respect to wild-type and expression of *OsHMGS, OsHMGR* from the MVA pathway increased approximately 20-60% (**Figure 3**).

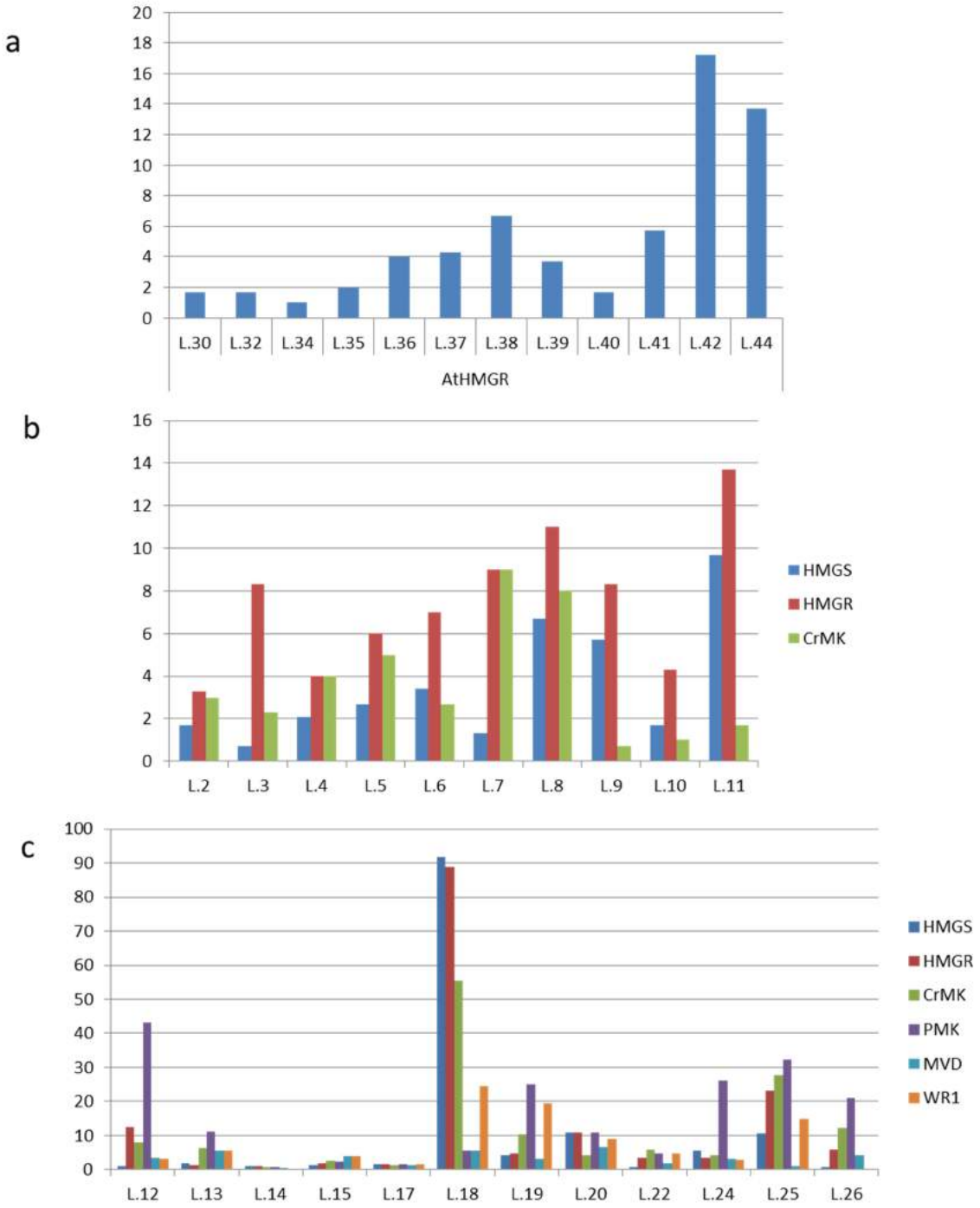


Figure 6. Expression profiles of ectopic MVA pathway genes: a) expression of *AtHMGR* normalized with respect to *OsActin*. b) expression of the early part of MVA pathway genes normalized with respect to *OsActin*. c) expression of the entire MVA pathway and wrinkled 1 transcription factor normalized with respect to *OsActin*.

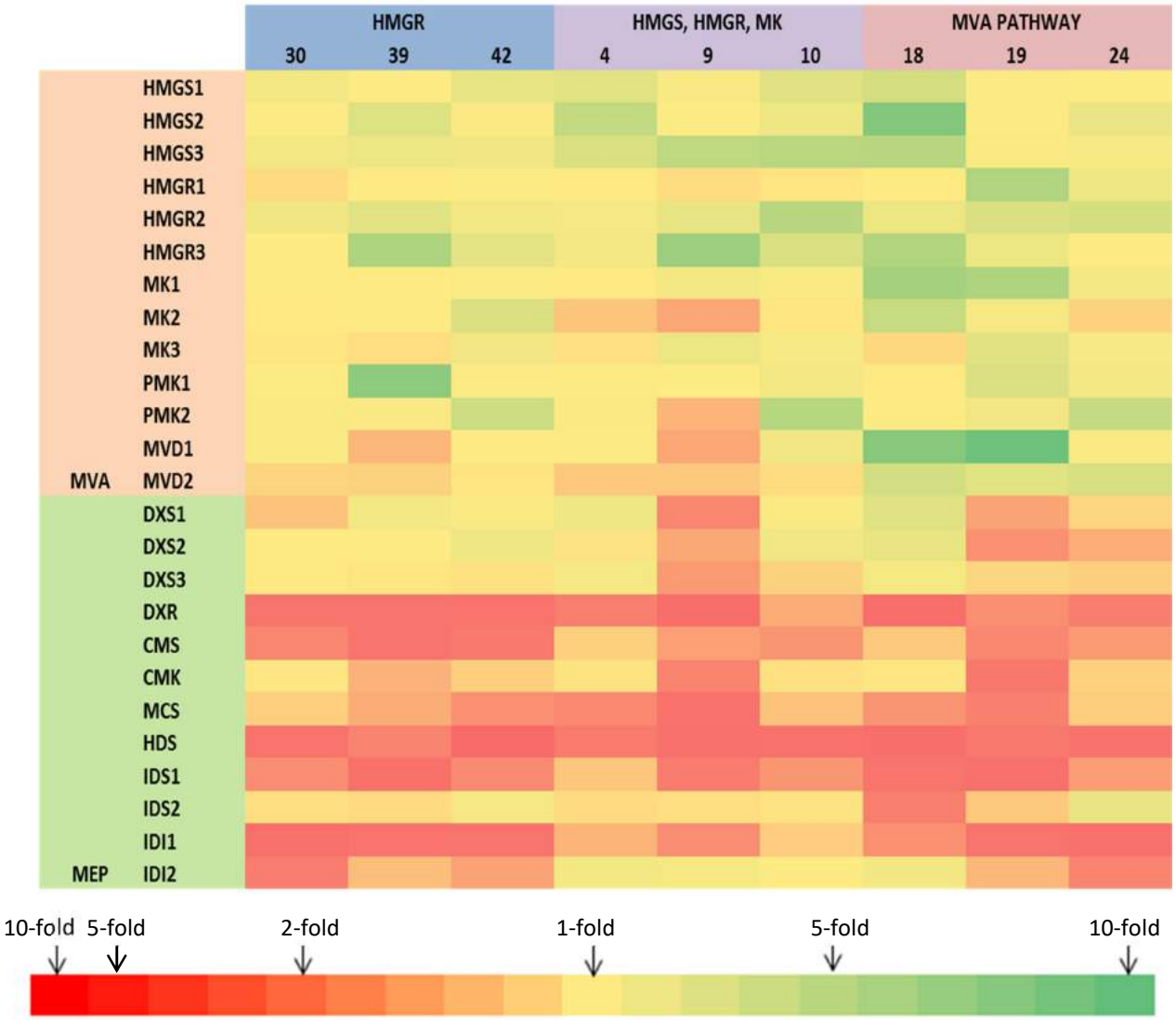


Figure 7. Heat map showing fold-changes in the expression in T2 seeds in wild-type and plants expressing ectopic MVA pathway genes. The red gradient shows increasing degrees of downregulation and the green gradient shows increasing degrees of upregulation, with yellow indicating no change in expression. The red gradient is expanded in the lower ranges because this is where most of the values lie while the green gradient is linear.

Metabolomic analysis

Increases in the levels of lutein, tocopherols and pheophytins (from chlorophyll degradation) (**Figure 4**) were measured by UPLC. All transgenic lines from the three different experiments accumulated more tocopherols compared with the wild-type, through the range 9 to 13650 ng more than wild-type. With respect to lutein and pheophytins, ca: 50% of the lines contained higher levels of these products. Lutein was not detected in wild-type seeds, while in all transgenic lines it accumulated in the range of 900-5000ng.

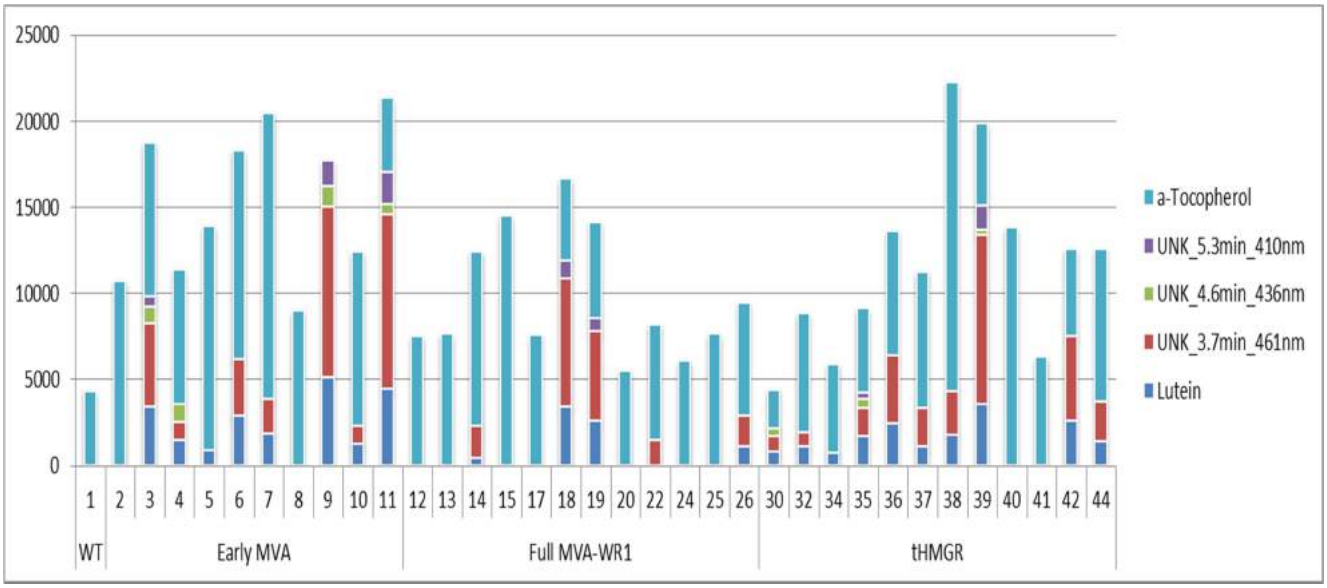


Figure 8. UPLC analysis where lutein, a-tocopherol and three types of pheophytins were detected in dry rice seeds. Wild-type sample showed around 4000 ng of tocopherol content while transgenic plants had increased levels of tocopherol and/or lutein and pheophytins content.

Phytohormone analysis

Analysis of phytohormones was performed in leaves to provide a basis for the early senescence phenotype we observed in all transgenic plants. We measured increases in abscisic acid which is related with the increases in plastidial carotenoids (lutein) and jasmonic acid derived from fatty acids. Gibberellins hormones do not maintain same levels and proportion than wild-type (related with the decrease in cytosolic metabolites) was measured. A decreased in precursors of ethylene (1-aminocyclopropane-1-carboxylic acid) for maturation was detected and changes in cytokinins that are related with cell division and delay of senescence (**Figure 5**). Jasmonic acid, cytokinins like isopentenyladenine and abscisic acid changes can be related with the early senescence produced in plants expressing *tHMGR* and full MVA-WR1 and no differences in plants expressing the early MVA.

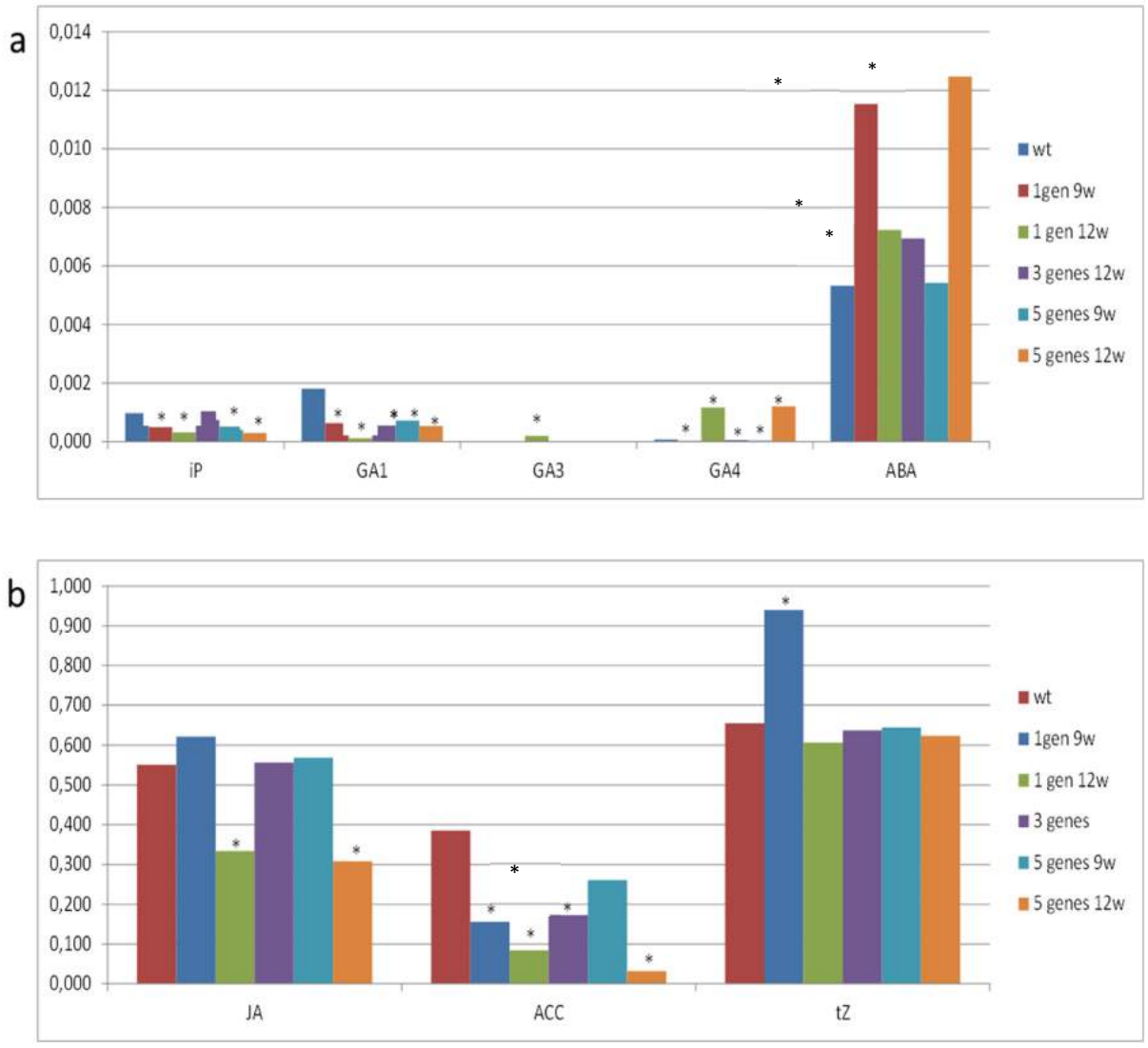


Figure 9. HPLC analysis of phytohormes. Abscisic acid (ABA), Isopentenyladenine (iP), jasmonic acid (JA), 1-Aminocyclopropane-1-carboxylic acid (ACC), trans-Zeatin (tZ) and gibberelins (GBs) exhibited changes with respect to the wild-type in tHMGR, early MVA and full MVA-WR1 transgenic plants. 1 gen 9w and 5 genes 9w refers to plants at 9 weeks growing in soil without senescence effects; 1 gen 12w and 5 genes 12w refers to plants at 12 weeks growing in soil with senescence effect. (n = 5 biological replicates, the asterisks indicate a statistically significant difference between wild-type (WT) and mutant, as determined by a Student's t-test (*P < 0.05). This is consistent with the observed senescence phenotype.

Phenotypic analysis

We observed clear differences in height between wild-type plants and *tHMGR* and full MVA -WR1 transgenic plants (**Figure 6a and 6c**). No such differences were observed in plants expressing the early part of the pathway. Leaves of *tHMGR* and full MVA-WR1 transgenic plants at 12 weeks entered into senescence, whereas wild-type plants remained healthy (**Figure 6b and 6d**).

There were statistically significant differences between wild-type, *tHMGR* and full MVA-WR1 transgenic plants in terms of height, number of leaves, leaf length, leaf width, surface and chlorophyll content (**Figure 7**). These changes co-related with phytohormone levels that limited growth and development and initiated senescence in the *tHMGR* and full MVA-WR1 plants compared to plants expressing only the early MVA plants.

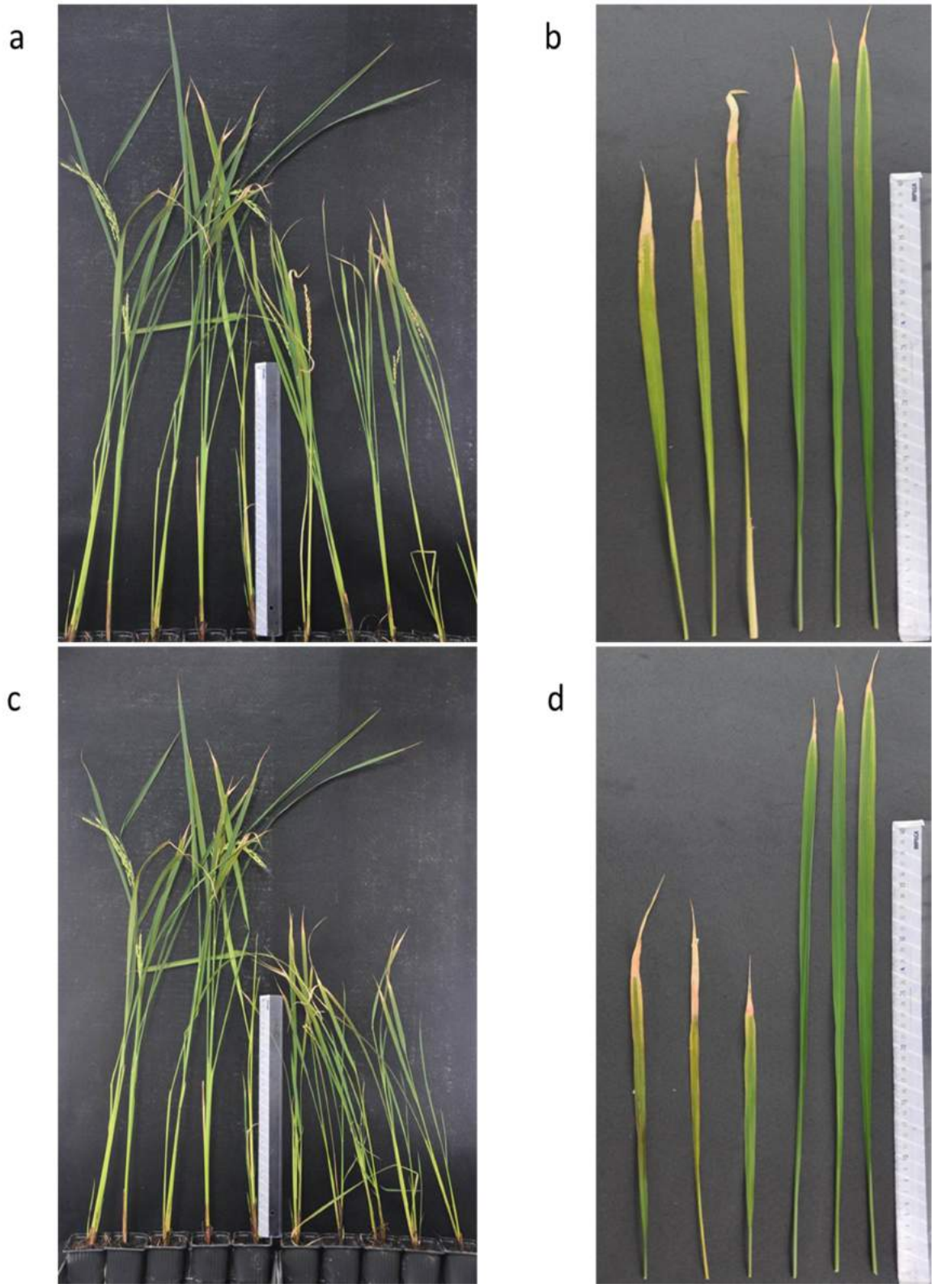


Figure 10. a) Plant phenotypes; 5 wild-type and 5 plants expressing *tHMGR*. b) Leaf phenotype, 5 wild-type and 5 expressing *tHMGR* c) Plants phenotype, 5 wild-type plants and 5 expressing full MVA+WR1. d) Leaves phenotype 5 wild-type and 5 expressing full MVA+WR1.

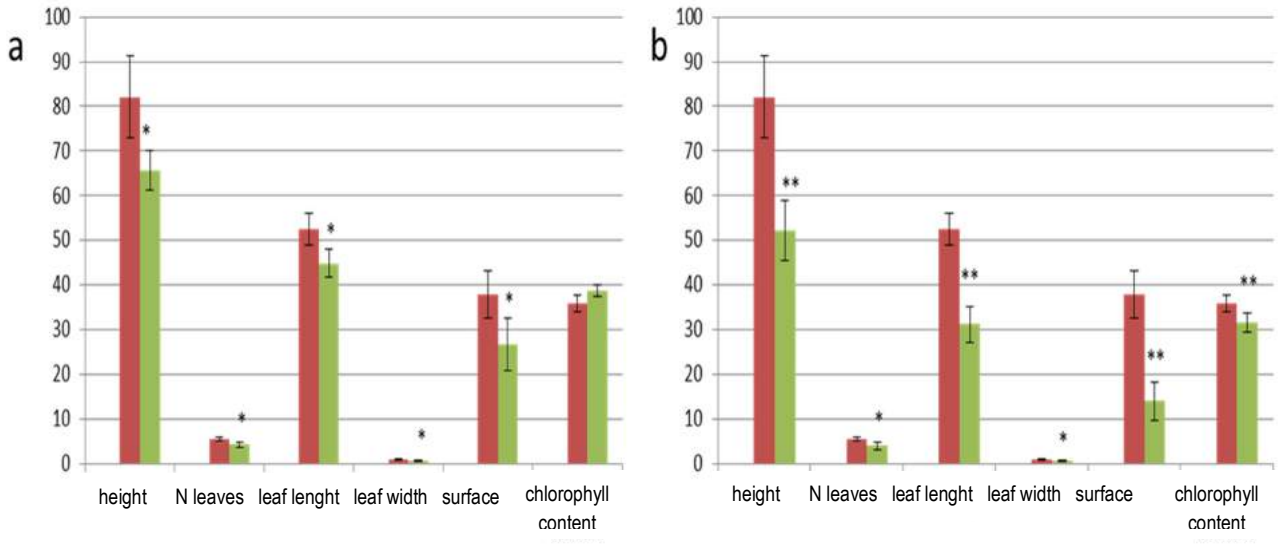


Figure 11. a) Phenotypic analysis of plants expressing *tHMGR*. B) Phenotypic analysis of plants expressing full MVA+WR1. Red colour represent wild-type phenotype and green colour represent mutant lines. Values are means ± sd (n = 10 biological replicates, 3 technical replicates for each biological replicate, except chlorophyll content with 6 technical replicates). The asterisks indicate a statistically significant difference between wild-type (WT) and mutant, as determined by a Student's t-test (*P < 0.05; **P < 0.01).

Microscopy

The seeds expressing *tHMGR* (**Figure 8 b, and c**) have less abundant starch grains by light microscopy while seeds expressing full MVA-WR1 and early MVA have more starch grains (**Figure 8 d, e, f and g**). The structure and composition of the aleurone layer does not have changes respect to the wild-type. More detailed plastid morphological analysis by TEM revealed that the three different experiments have different plastid morphology. While *tHMGR* line 39 showed a deformed, granulated, unstructured and smaller plastids and a bigger holds of starch; line 4 expressing early MVA accumulation of fatty acids in form of black spots in the plastids with a more structured membrane; and line 19 expressing full MVA-WR1 has the closest structure compare with wild-type but having small spots of fatty acids accumulation and a semi-structured membrane (**Figure 9**).

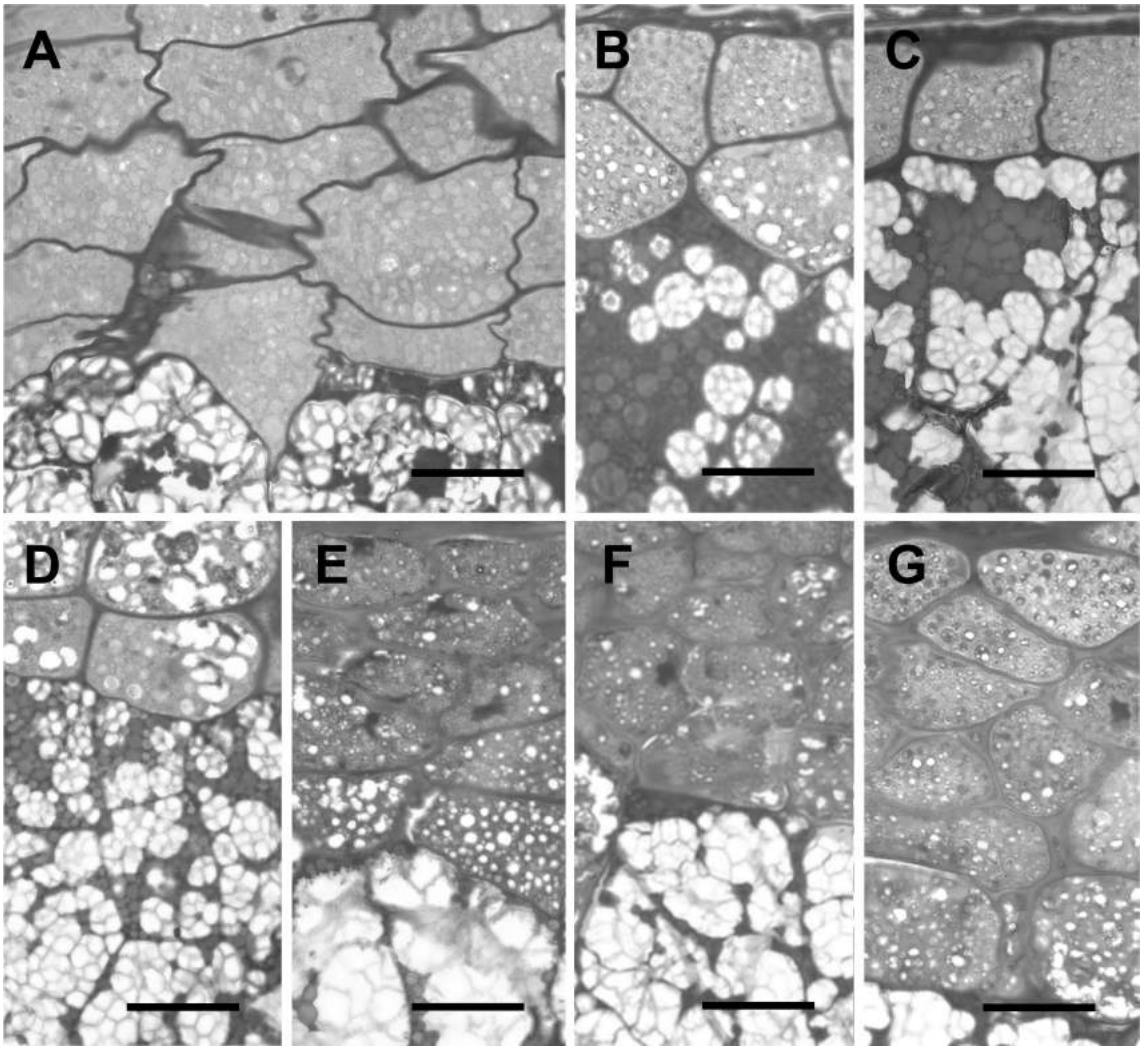


Figure 12. Light microscopy of rice seeds. a) aleurone layer and starch grains of wild-type seed. b) aleurone layer and starch grains of line 39 expressing *tHMGR.* c) aleurone layer and starch grains of line 40 expressing *tHMGR.* d) aleurone layer and starch grains of line 3 expressing early MVA. e) aleurone layer and starch grains of line 4 expressing early MVA. f) aleurone layer and starch grains of line 18 expressing full MVA-WR1. g) aleurone layer and starch grains of line 24 expressing full MVA-WR1. There are changes in aleurone layer structure and plants expressing ectopic MVA pathway genes showed less starch grains.

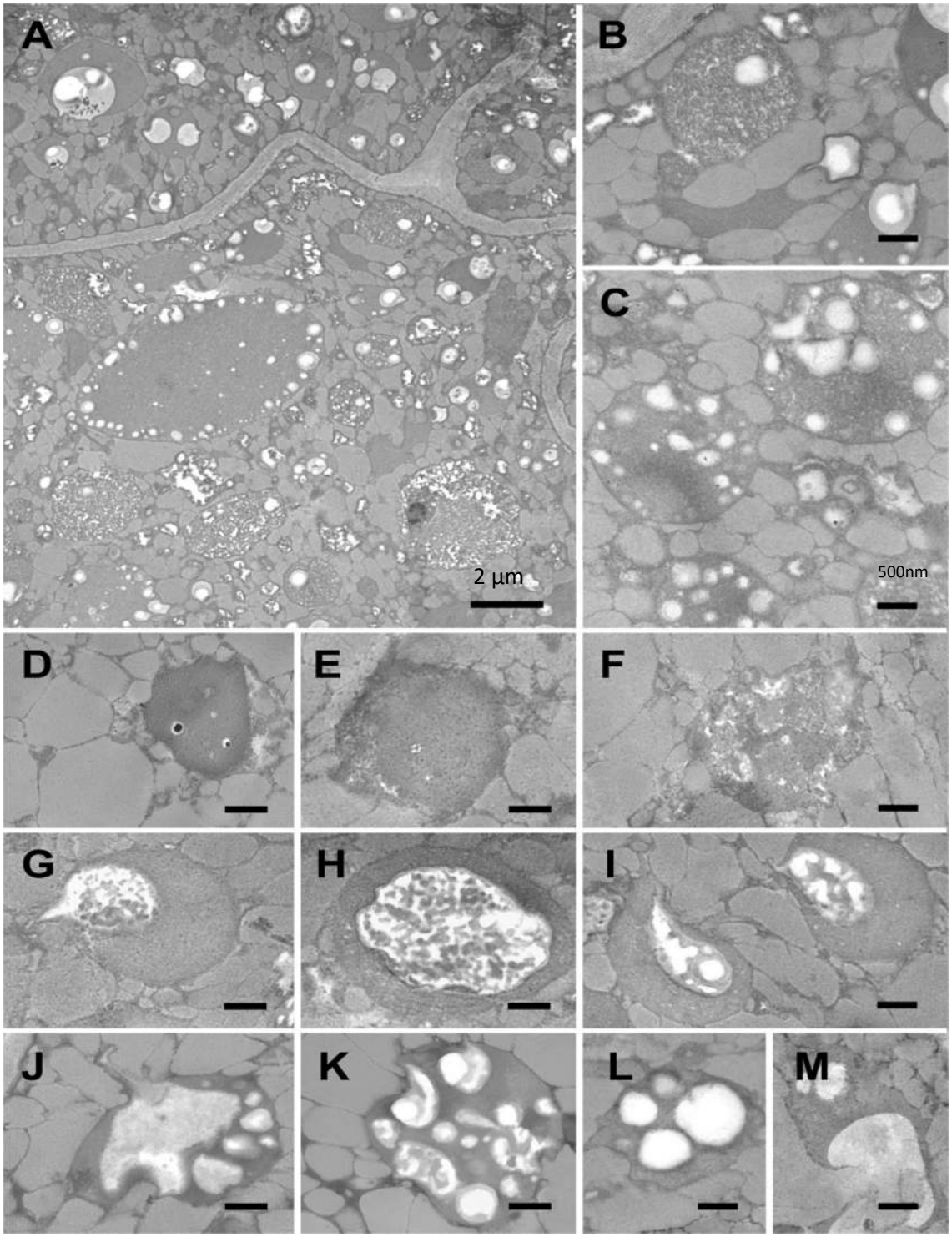


Figure 13. Electron microscopy (TEM) for plastid visualization. a)b)c) wild-type seed at different magnifications (2 µm-) showing diverse types of plastids and starch granules. d)e)f) plastids of line 19, expressing full MVA-WR1*,* with small black spots (fatty acid accumulation) and with deformed and unstructured membrane. g)h)i) line 4, expressing early MVA, showing granulated plastids with high accumulation of black spots (fatty acids) and starch. j)k)l)m) line 39, expressing *tHMGR*, with a complete deformed and unstructured membrane with big white holds of starch.

Starch and soluble sugars

Analysis of starch and soluble sugars was performed to see changes at global metabolism apart from fatty acids and isoprenoids modified by the expression of the ectopic genes. Heat Map of starch and soluble sugars of the independent lines was created to show patterns in the three experiments (**Figure 10**). In leaves, the three different experiments followed a pattern were starch is decreased compared with wild-type (average between 50 to 68% decrease), while soluble sugars are increased respect to wild-type (average between 156-176% increase) (**Figure 11**). In seeds, the three different experiments followed different patterns. Lines expressing *tHMGR* have highly increased levels of soluble sugars (110%) and a decrease in starch (20%) (**Figure 11**). Lines expressing early MVA have a slightly increased in soluble sugars (32%) and starch (25%) (**Figure 11**). Lines expressing full MVA-WR1 have a highly increased in starch (80%) and a decreased of 25% in soluble sugars (**Figure 11**).

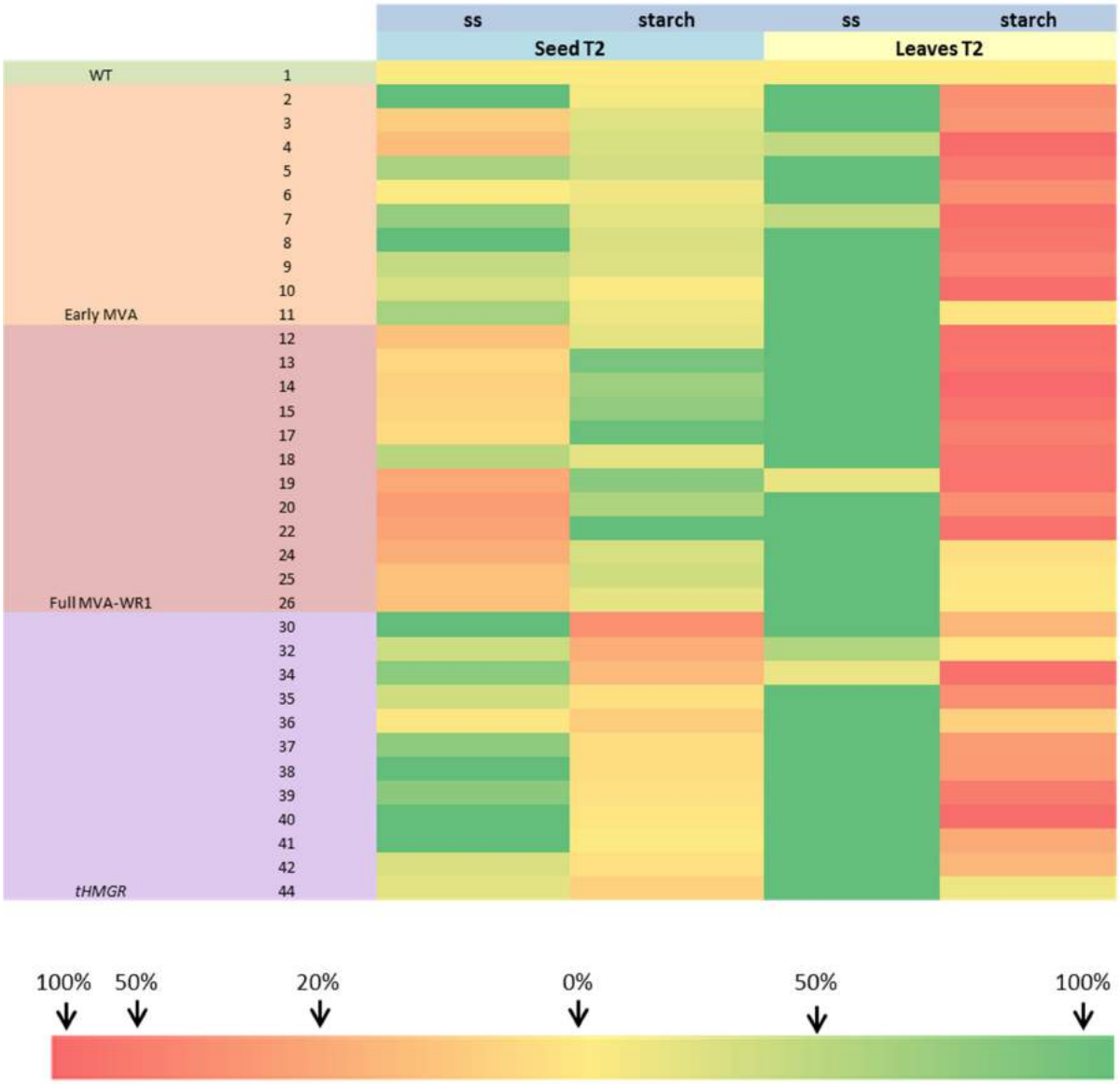


Figure 14. Heat Map showing changes at soluble sugars and starch levels in percentage (%) of rice T2 seeds and leaves.The red gradient shows increasing degrees of downregulation and the green gradient shows increasing degrees of upregulation, with yellow indicating no change in expression. The red gradient is expanded in the lower ranges because this is where most of the values lie while the green gradient is linear.

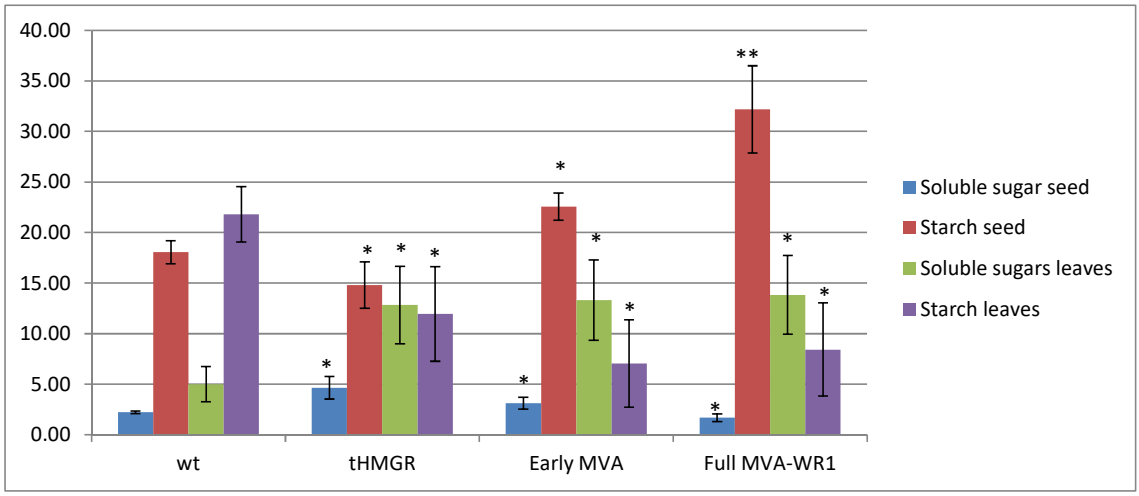


Figure 15. Seed and leaf carbohydrate content in wild-type, tHMGR experiment, early MVA experiment and full MVA-WR1 experiment. Values are expressed as means ± SD (n = 10 biological replicates, 2 technical replicates for each biological replicate). Asterisks indicate a statistically significant difference, as determined by Student's t-test (*P < 0.05; **P < 0.01).

3.5 Discussion

The potential of plants for production of isoprenoids derived from the synthesis of the common precursors isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) via two different pathways separately in cellular compartments (cytosol and plastids). Both of the pathways (MVA and MEP) can be modified producing completely different changes at isoprenoid levels and content (Shi et al., 2014).

No much work has been done in plants do to their complexity; studies in microorganism such as *E.coli* and *S.cerevisiae* for production of isoprenoids have been more widely performed (Ward et al., 2018). For that reason, some studies have been done expressing ectopic genes (typically HMGR, HMGS and DXR) to overcome the bottleneck of one of the pathways and increase isoprenoid content (Singh et al., 2014; Wang et al., 2012 and Shi et al.,2014). But most of the studies were focus in to increase levels of one single isoprenoid or a family of isoprenoids without paying attention to global changes produced by the introduction of an ectopic gene (Hey et al., 2005; Jayashree et al., 2018 and Wang et al., 2011).

In our study, we wanted to investigate the potential of ectopic MVA genes in plastids for crosstalk with endogenous MEP and MVA and changes at amount and composition of isoprenoids. For that reason, we performed three different experiments; expressing *tHMGR* gene that is considered as the most important gene that regulates a rate-limiting step in the MVA pathway (Chappell et al., 1995) because it is regulated at transcriptional, post-transcriptional , translational and post-translational level (Hemmerlin, 2013); expressing the early MVA (the three first genes of the pathway *HMGS*, *HMGR*, *MK*) because this integrate the most regulated genes of the pathway *HMGR* and *HMGS* (Vranova et al., 2013) and *MK* which has a role

in cytokinin-mediated growth and development as well as in substrate feedback and diurnal regulation (Schulte et al., 2000); and expressing the full MVA pathway and a transcription factor (WR1) that will increase Acetyl-CoA precursors (Cernac and Benning, 2004) because deficiency in acetyl-CoA and death of plants has been obtained before only expressing MVA pathway in our laboratory.

In general, our results demonstrate that expression of the MVA pathway in the plastids resulted in an enhanced content of fatty acids and carotenoids. This suggests that increasing the availability of the IPP monomer resulted in enhanced flux through the downstream carotenoid biosynthetic pathway or enhanced expression or activity of the pathway. The complex regulation of carotenoid biosynthesis (Bouvier et al., 2005; Botella-Pavia et al., 2004; Fraser et al., 2009) suggests that many factors may impact levels.

We compared expression of the ectopic MVA pathway genes with a housekeeping gene that express highly in rice, for that reason we selected *actin* gene. Lines expressing *tHMGR* have levels of expressions between 0.5 to 17 times *actin* gene expression. Lines expressing Early MVA have levels of expression of *HMGS* between 0.5 to 9 compared with *actin* gene; *tHMGR* expression levels between 4 to 13 times compared with *actin* gene and *MK* expression levels between 0.5 to 9 times compared with *actin* gene. Lines expressing full MVA-WR1 has levels of expression of *HMGS* between 0.8 to 91 compared with *actin* gene; *tHMGR* expression levels between 1 to 88 times compared with *actin* gene; *MK* expression levels between 0.6 to 55 compared with *actin* gene; *PMK* expression levels between 0.88 to 43 compared with *actin* gene; *MVD* expression levels between 0.4 to 6.5 compared with *actin* gene and *WR1* expression levels between 0.2 to 24 compared with *actin* gene.

For endogenous gene expression analysis in seeds, we selected 3 lines with different levels of expression. From these 3 lines, we select one with low ectopic gene levels of expression, one with intermediate ectopic gene levels of expression and one with high ectopic gene levels of expression. Expressing *tHMGR*, we selected lines 30 (low), 39 (intermediate) and 42 (high); expressing early MVA, we selected lines 4 (intermediate), 9 (high) and 10 (low); and expressing full MVA-WR1, we selected lines 18 (high), 19 (intermediate) and 24 (low). Surprisingly, the three experiments followed similar patterns of endogenous MVA and MEP genes expression independently of ectopic gene levels of expression. In general, increased levels of endogenous MVA pathway genes expression and decreased levels of MEP pathway genes expression. Highlighting in MVA pathway, the increase in HMGS and HMGR genes in the three experiments and the MVD genes which suffered an increase in full MVA-WR1 but a decrease in the experiments of tHMGR and early MVA; in MEP pathway, the great decrease of DXR gene and the following genes, and a great decrease in IDS and IDI only in isoform 1, in isoform 2 there is a slightly decrease. Although the genes encoding the MVA pathway to the point of production of IPP were engineered into rice plastids, plastidial IPP isomerase expression must be determined in further analysis. Tritsch et al. (2010) demonstrated that IPP isomerase acts to dynamically adjust the IPP/DMAPP ratio. It is possible that altering plastidial IPP

isomerase expression would allow optimization of the ratio of IPP and DMAPP to further increase product levels (Martin et al., 2003).

UPLC analysis was performed to ascertain carotenoid content and other metabolites such as tocopherol. The analysis showed that some lines contain higher levels of tocopherols compared to the wild-type. Tocopherol is synthesized through three alternative pathways: precursors from the glycolysis (phosphoenolpyruvate) and pentose phosphate pathway (erythrose-4-phosphate) converging to the shikimic acid pathway; directly from the MEP pathway (phytyl-2-phosphate) or via pheophytins from the chlorophyll degradation pathway (Almeida et al., 2011). Expression of the MVA pathway in plastids enhanced the levels of products associated with fatty acid synthesis and, also, decreased the cytosolic MVA pathway. Most fatty acids (16- and 18-carbon fatty acids) are synthesized from acetyl CoA in plastids. Part of the fatty acid pool that is exported to the cytoplasm is assembled into storage lipids at the endoplasmic reticulum (ER), by donating acyl groups to glycerol-3-phosphate to form TAGs. Under some developmental or stress conditions, plastids synthesize TAGs as well (Kaup et al., 2002). The increase of fatty acid synthesis in response to expression of the MVA pathway in plastids, as evidenced by an increased level of TAGs, indicates that acetyl CoA was freely available in the plastid and not limiting when WR1 is expressed, not like MVA pathway alone where plants died due to the deficiency of Acetyl CoA (Cernac and Benning, 2004). Overproduction of intermediates of MVA or MEP pathway can limit availability of Acetyl-CoA and pyruvate that are essential precursors in other pathways such as glycolysis, pentose phosphate, amino acid and fatty acid biosynthesis (Hemmerlin et al., 2012). Some of these pathways are implicated in tocopherol synthesis. Some transgenic lines exhibited increases not only in tocopherols but also pheophytins and lutein. Rice seeds are green due to the chlorophyll content, once the seeds got dry chlorophyll started to degrade forming pheophytins (Jalink et al., 1999). An increase of lutein was observed due to the accumulation of IPP in the plastids, derived from the cytosol produced by the endogenous and the ectopic MVA pathways. Overexpression of MVA genes in the plastids resulted in the accumulation of IPP and DMAPP that resulted in negative feedback regulation of MEP pathway genes. Cytosol had a deficit of IPP and DMAPP that activated the endogenous MVA pathway genes.

Since we observed accumulation of lutein but not zeaxanthin (β-branch), we further investigated phytohormone concentrations. The β-branch of the carotenoid biosynthetic pathway produces abscisic acid (ABA). Despite the fact that overexpression of MVA pathway enzymes is not generally associated with alterations in the phenotype of plants, we observed developmental perturbations in plants expressing ectopic MVA genes, when compared to wild-type. Plants expressing *tHMGR* or the full MVA-WR1 complement showed a different growth phenotype in terms of height, number of leaves, leaf length, leaf width and leaf surface, and a decrease in chlorophyll content in plants expressing full MVA-WR1. Plants expressing early MVA pathway genes had no visible phenotype changes. The difference between the plants expressing only the early MVA pathway genes versus plants expressing the full transgene complement, including WR1 is that MK overexpression induced cytokinin-mediated growth and development

(Champenoy and Tourte, 1998). In addition, phytohormone analysis showed a clear decrease in IP (Isopentenyl adenine) in *tHMGR* and full MVA-WR1 plants as part of cytokinin family. It has been reported that the overexpression of MK produced an increase in chlorophyll content throughout plant development, followed by senescence retardation (Champenoy and Tourte, 1998). Moreover, jasmonic acid (JA) levels were similar to the wild-type in plants expressing the early MVA pathway genes and they were decreased in plants expressing tHMGR and full MVA-WR1 genes. Gibberellins (GAs) are synthetized from the MEP pathway. In our analysis GA1 and GA4 were decreased in plants expressing tHMGR, early MVA and full MVA-WR1 genes. This can be explained because brassinosteroids (BRs) regulate positively GAs biosynthesis gene expression in rice (Tong et al., 2014). BRs are synthetized by the endogenous MVA pathway, mobilization of IPP-DMAPP from the cytosol to plastids decrease the precursors to form BRs. Lower levels of BRs do not produce the signal for GAs biosynthesis. Abscisic acids levels correlated with an increase of lutein in plants expressing tHMGR, early MVA and full MVA-WR1 genes. 1-Amynocyclopropane-1-carboxylic acid (ACC) is the precursor of ethylene and it can be inhibited by amino-oxyacetic acid, AOA or amino-ethoxy vinyl glycine, AVG. These molecules are synthesized from amino acids when there is an increase of amino acid accumulation (Taiz et al., 2015). As described before, increased levels of Acetyl CoA can be redirected to other pathways such as sugar metabolism, glycolysis, amino acid biosynthesis and fatty acid biosynthesis (Hemmerlin et al., 2012), that can explain the decrease of ACC in our experiments.

Light and transmission electron microscopy (TEM) on seeds was performed to determine whether metabolism changes produce alterations in cell structure. Light microscopy did not indicate any such changes in the aleurone layer. Lines expressing *tHMGR* had less starch grains compared to the wild-type, while plants expressing early MVA genes appeared similar to the wild-type and plants expressing full MVA-WR1 genes exhibited an increase in starch grains. TEM showed that two lines expressing *tHMGR* gene were obviously affected in plastid structure because of the deformed and unstructured membranes. In addition, two lines expressing early MVA genes had clear accumulation of fatty acids which appeared as black spots. Two lines expressing full MVA-WR1 genes had similar structures to the wild-type but showing some small black spots of fatty acids.

Plant cells have delicately responsive mechanisms that sense levels of fatty acids and adjust metabolism to degrade or synthesize them, as needed (Eccleston and Ohlrogge, 1998). In leaves, excess fatty acids are degraded by b-oxidation in peroxisomes. Paradoxically, the loss of fatty acids activates fatty acid synthesis in plastids (Eccleston and Ohlrogge, 1998). The enhancement of TAG levels in the MVA lines, reported earlier reflects the result of an intriguing interplay between enhanced availability of isoprenoid substrates in the plastids, effects of isoprenoid intermediates on fatty acid synthesis (Kizer et al., 2008), and the balance of degradation of excess fatty acids in the cytosol with activation of fatty acid synthesis in the plastid. It is possible that IPP or derived metabolites only exit the cytosol when concentrations exceed normal physiological levels, such as in our study.

Starch and soluble sugars analysis was performed after microscopy experiment showing changes at starch levels. In leaves there is a patter in the three experiments were soluble sugars are increased and starch is decreased. In seeds, each experiment followed a different pattern. Lines expressing *tHMGR,* have a demand of Acetyl CoA that can be obtained from the degradation of starch, increasing soluble sugar levels. Lines expressing early MVA have slightly increased levels of soluble sugars as a precursor to form Acetyl-CoA. Lines expressing full MVA-WR1 synthesized more amount of Acetyl-CoA due to the effect of WR1 and the excess of Acetyl-CoA need to be converted in to starch and decreasing levels of soluble sugars.

Plastid transformation of the full MVA pathway from *S.cerevisiae* has been done one time before in tobacco (Kumar et al., 2012). Overexpression in tobacco did not generate phenotypic changes. The plants enhanced MVA pathway products (squalene and sterols), increased levels of triacylglycerol and a slightly increase of carotenoids. The fundamental differences in experimental systems may explain the discrepancy and would be of interest to study further.

3.6 Conclusions

Changes in metabolism are difficult to correlate due to the complex regulation of plants and low number of published reports in this field. In addition, there is not clear correlation between the ectopic expression of MVA genes and changes in the expression of endogenous genes from the MVA and MEP pathways. Using endosperm specific promoters from the ectopic MVA pathway genes modified indirectly the plant phenotype. These changes resulted from a deregulation in plant phytohormone biosynthesis that resulted in growth problems and early senescence in plants expressing *tHMGR* and full MVA-WR1 genes. Plants expressing early MVA genes do not have decreased levels of cytokinins, in particular iP, which produce the defective phenotype. The ectopic genes of the MVA pathway modified the amount and type of terpenoids in rice seeds, increasing plastidial terpenoids such as carotenoids (lutein) not measurable in wild-type seeds, and tocopherol. Fatty acid levels were increased due to the ectopic mevalonate pathway genes in plastids by increasing amount of Acetyl-CoA as a precursor for fatty acid biosynthesis left over from the high expression of the endogenous mevalonate pathway. Global changes at different metabolomic pathways had been identified and further analysis must be performed.

3.7 References

Albacete A, Ghanem ME, Martínez-Andújar C, Acosta M, Sánchez-Bravo J, Martínez V, Lutts S, Dodd IC and Pérez-Alfocea F (2008) Hormonal changes in relation to biomass partitioning and shoot growth impairment in salinized tomato (Solanum lycopersicum L.) plants. J Exp Bot 59, 4119-4131

Almeida, J., Quadrana, L., Asís, R., Setta, N., de Godoy, F., Bermúdez, L., and Rossi, M. (2011). Genetic dissection of vitamin E biosynthesis in tomato. *Journal of experimental botany*, *62*, 3781-3798.

Alper, H., Jin, Y. S., Moxley, J. F., and Stephanopoulos, G. (2005). Identifying gene targets for the metabolic engineering of lycopene biosynthesis in Escherichia coli. *Metabolic engineering*, *7*, 155-164.

Bassie, L., Zhu, C., Romagosa, I., Christou, P., and Capell, T. (2008). Transgenic wheat plants expressing an oat arginine decarboxylase cDNA exhibit increases in polyamine content in vegetative tissue and seeds. *Molecular breeding*, *22*, 39-50.

Berthelot, K., Estevez, Y., Deffieux, A., and Peruch, F. (2012). Isopentenyl diphosphate isomerase: a checkpoint to isoprenoid biosynthesis. *Biochimie*, *94*(8), 1621-1634.

Bligh, E. G., and Dyer, W. J. (1959). A rapid method of total lipid extraction and purification. *Canadian journal of biochemistry and physiology*, *37*, 911-917

Bligh, E. G., and Dyer, W. J. (1959). A rapid method of total lipid extraction and purification. *Canadian journal of biochemistry and physiology*, *37*, 911-917.

Botella-Pavía, P., Besumbes, O., Phillips, M. A., Carretero-Paulet, L., Boronat, A., and Rodríguez-Concepción, M. (2004). Regulation of carotenoid biosynthesis in plants: evidence for a key role of hydroxymethylbutenyl diphosphate reductase in controlling the supply of plastidial isoprenoid precursors. *The Plant Journal*, *40*, 188-199.

Boucher, Y., and Doolittle, W. F. (2000). The role of lateral gene transfer in the evolution of isoprenoid biosynthesis pathways. *Molecular microbiology*, *37*, 703-716.

Bouvier, F., Rahier, A., and Camara, B. (2005). Biogenesis, molecular regulation and function of plant isoprenoids. *Progress in lipid research*, *44*, 357-429.

Burg, J. S., and Espenshade, P. J. (2011). Regulation of HMG-CoA reductase in mammals and yeast. *Progress in lipid research*, *50*, 403-410.

Cernac, A., and Benning, C. (2004). WRINKLED1 encodes an AP2/EREB domain protein involved in the control of storage compound biosynthesis in Arabidopsis. *The Plant Journal*, *40*, 575-585.

Champenoy, S., and Tourte, M. (1998). Expression of the yeast mevalonate kinase gene in trangenic tobacco. *Molecular Breeding*, *4*, 291-300.

Chappell, J., Wolf, F., Proulx, J., Cuellar, R., and Saunders, C. (1995). Is the reaction catalyzed by 3-hydroxy-3-methylglutaryl coenzyme A reductase a rate-limiting step for isoprenoid biosynthesis in plants?. *Plant Physiology*, *109*, 1337-1343.

Christou, P., Ford, T. L. and 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 immature zygotic embryos. Nature Biotechnol. 9, 957–962.

Cordoba, E., Salmi, M., and León, P. (2009). Unravelling the regulatory mechanisms that modulate the MEP pathway in higher plants. *Journal of Experimental Botany*, *60*, 2933-2943.

Davies, P. J. (2010). The plant hormones: their nature, occurrence, and functions. In *Plant hormones* (pp. 1-15). Springer, Dordrecht.

DISCH, A., SCHWENDER, J., MÜLLER, C., LICHTENTHALER, H. K., and ROHMER, M. (1998). Distribution of the mevalonate and glyceraldehyde phosphate/pyruvate pathways for isoprenoid biosynthesis in unicellular algae and the cyanobacterium Synechocystis PCC 6714. *Biochemical Journal*, *333*, 381-388.

Eccleston, V. S., and Ohlrogge, J. B. (1998). Expression of lauroyl–acyl carrier protein thioesterase in Brassica napus seeds induces pathways for both fatty acid oxidation and biosynthesis and implies a set point for triacylglycerol accumulation. *The Plant Cell*, *10*, 613-621.

Farré, G., Sudhakar, D., Naqvi, S., Sandmann, G., Christou, P., Capell, T. and 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 Res. 21, 1093–1097.

Fraser, P. D., Enfissi, E. M., and Bramley, P. M. (2009). Genetic engineering of carotenoid formation in tomato fruit and the potential application of systems and synthetic biology approaches. *Archives of Biochemistry and Biophysics*, *483*, 196-204.

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

Ghassemian, M., Lutes, J., Tepperman, J. M., Chang, H. S., Zhu, T., Wang, X., and Lange, B. M. (2006). Integrative analysis of transcript and metabolite profiling data sets to evaluate the regulation of biochemical pathways during photomorphogenesis. *Archives of Biochemistry and Biophysics*, *448*, 45-59.

Hemmerlin, A. (2013). Post-translational events and modifications regulating plant enzymes involved in isoprenoid precursor biosynthesis. *Plant science*, *203*, 41-54.

Hemmerlin, A., Harwood, J. L., and Bach, T. J. (2012). A raison d'être for two distinct pathways in the early steps of plant isoprenoid biosynthesis?. *Progress in lipid research*, *51*, 95-148.

Hey, S. J., Powers, S. J., Beale, M. H., Hawkins, N. D., Ward, J. L., and Halford, N. G. (2006). Enhanced seed phytosterol accumulation through expression of a modified HMG-CoA reductase. *Plant biotechnology journal*, *4*, 219-229.

Jalink, H., Van der Schoor, R., Birnbaum, Y. E., and Bino, R. J. (1999, May). Seed chlorophyll content as an indicator for seed maturity and seed quality. In *VI Symposium on Stand Establishment and ISHS Seed Symposium 504* (pp. 219-228).

Jayashree, R., Nazeem, P. A., Rekha, K., Sreelatha, S., Thulaseedharan, A., Krishnakumar, R., ... & Venkatachalam, P. (2018). Over-expression of 3-hydroxy-3-methylglutaryl-coenzyme A reductase 1 (hmgr1) gene under super-promoter for enhanced latex biosynthesis in rubber tree (Hevea brasiliensis Muell. Arg.). *Plant Physiology and Biochemistry*, *127*, 414-424.

Juliano, B.O. (1971) A simplified assay for milled-rice amylose. Cereal Sci. Today, 16, 334–336.

Kang, T. J., and Yang, M. S. (2004). Rapid and reliable extraction of genomic DNA from various wild-type and transgenic plants. *BMC biotechnology*, *4*, 20.

Kaup, M. T., Froese, C. D., and Thompson, J. E. (2002). A role for diacylglycerol acyltransferase during leaf senescence. *Plant physiology*, *129*, 1616-1626.

Kizer, L., Pitera, D. J., Pfleger, B. F., and Keasling, J. D. (2008). Application of functional genomics to pathway optimization for increased isoprenoid production. *Applied and environmental microbiology*, *74*, 3229-3241.

Kovacs, W. J., Olivier, L. M., and Krisans, S. K. (2002). Central role of peroxisomes in isoprenoid biosynthesis. *Progress in lipid research*, *41*, 369-391.

Kumar, P. P. (2013). Regulation of biotic and abiotic stress responses by plant hormones.

Kumar, S., Hahn, F. M., Baidoo, E., Kahlon, T. S., Wood, D. F., McMahan, C. M., and Whalen, M. C. (2012). Remodeling the isoprenoid pathway in tobacco by expressing the cytoplasmic mevalonate pathway in chloroplasts. *Metabolic engineering*, *14*, 19-28.

Ma, S. M., Garcia, D. E., Redding-Johanson, A. M., Friedland, G. D., Chan, R., Batth, T. S., and Lee, T. S. (2011). Optimization of a heterologous mevalonate pathway through the use of variant HMG-CoA reductases. *Metabolic engineering*, *13*, 588-597.

Martin, V. J., Pitera, D. J., Withers, S. T., Newman, J. D., and Keasling, J. D. (2003). Engineering a mevalonate pathway in Escherichia coli for production of terpenoids. *Nature biotechnology*, *21*, 796.

Murata Y (1967) In photosynthesis and utilization of solar energy. Level I Experiments Regort 1, August 1967.

Rodrıguez-Concepción, M., and Boronat, A. (2002). Elucidation of the methylerythritol phosphate pathway for isoprenoid biosynthesis in bacteria and plastids. A metabolic milestone achieved through genomics. Plant physiology, 130, 1079-1089.

Schulte, A. E., Durán, E. M. L., van der Heijden, R., and Verpoorte, R. (2000). Mevalonate kinase activity in Catharanthus roseus plants and suspension cultured cells. *Plant Science*, *150*, 59-69.

Shi, M., Luo, X., Ju, G., Yu, X., Hao, X., Huang, Q., and Kai, G. (2014). Increased accumulation of the cardio-cerebrovascular disease treatment drug tanshinone in Salvia miltiorrhiza hairy roots by the enzymes 3-hydroxy-3-methylglutaryl CoA reductase and 1-deoxy-D-xylulose 5-phosphate reductoisomerase. *Functional & integrative genomics*, *14*, 603-615.

Singh, S., Pal, S., Shanker, K., Chanotiya, C. S., Gupta, M. M., Dwivedi, U. N., and Shasany, A. K. (2014). Sterol partitioning by HMGR and DXR for routing intermediates toward withanolide biosynthesis. *Physiologia plantarum*, *152*, 617-633.

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

Taiz, L., Zeiger, E., Møller, I. M., and Murphy, A. (2015). Plant physiology and development. *Sunderland, MA: Sinauer Associates*.

Tippmann, S., Ferreira, R., Siewers, V., Nielsen, J., and Chen, Y. (2017). Effects of acetoacetyl-CoA synthase expression on production of farnesene in Saccharomyces cerevisiae. *Journal of industrial microbiology & biotechnology*, *44*, 911-922.

Tong, H., Xiao, Y., Liu, D., Gao, S., Liu, L., Yin, Y., and Chu, C. (2014). Brassinosteroid regulates cell elongation by modulating gibberellin metabolism in rice. *The Plant Cell*, tpc-114.

Tritsch, D., Hemmerlin, A., Bach, T. J., and Rohmer, M. (2010). Plant isoprenoid biosynthesis via the MEP pathway: in vivo IPP/DMAPP ratio produced by (E)-4-hydroxy-3-methylbut-2-enyl diphosphate reductase in tobacco BY-2 cell cultures. *Febs Letters*, *584*, 129-134.

Tsunoda, S. (1962). A Developmental Analysis of Yielding Ability in Varieties of Field Crops: IV. Quantrtatrve and spatral development of the stem-system. *Japanese Journal of Breeding*, *12*, 49-56.

Tsuruta, H., Paddon, C. J., Eng, D., Lenihan, J. R., Horning, T., Anthony, L. C., and Newman, J. D. (2009). High-level production of amorpha-4, 11-diene, a precursor of the antimalarial agent artemisinin, in Escherichia coli. *PLoS One*, *4*, e4489.

Valdez, M., Cabrera-Ponce, J. L., Sudhakar, D., Herrera-Estrella, L. and Christou, P. (1998) Transgenic Central American, West African and Asian elite rice varieties resulting from particle bombardment of foreign DNA into mature seed-derived explants utilizing three different bombardment devices. Ann. Bot. 82, 795–801

Vranová, E., Coman, D., and Gruissem, W. (2012). Structure and dynamics of the isoprenoid pathway network. *Molecular plant*, *5*, 318-333.

Vranová, E., Coman, D., and Gruissem, W. (2013). Network analysis of the MVA and MEP pathways for isoprenoid synthesis. *Annual review of plant biology*, *64*, 665-700.

Wang, H., Nagegowda, D. A., Rawat, R., Bouvier-Navé, P., Guo, D., Bach, T. J., and Chye, M. L. (2012). Overexpression of Brassica juncea wild-type and mutant HMG-CoA synthase 1 in Arabidopsis up-regulates genes in sterol biosynthesis and enhances sterol production and stress tolerance. *Plant biotechnology journal*, *10*, 31-42.

Wang, H., Nagegowda, D. A., Rawat, R., Bouvier-Navé, P., Guo, D., Bach, T. J., and Chye, M. L. (2012). Overexpression of Brassica juncea wild-type and mutant HMG-CoA synthase 1 in Arabidopsis up-regulates genes in sterol biosynthesis and enhances sterol production and stress tolerance. *Plant biotechnology journal*, *10*, 31-42.

Wang, Y., Jing, F., Yu, S., Chen, Y., Wang, T., Liu, P., and Tang, K. (2011). Co-overexpression of the HMGR and FPS genes enhances artemisinin content in Artemisia annua L. *Journal of Medicinal Plants Research*, *5*, 3396-3403.

Ward, V. C., Chatzivasileiou, A. O., and Stephanopoulos, G. (2018). Metabolic engineering of Escherichia coli for the production of isoprenoids. *FEMS microbiology letters*, *365*, fny079.

Ward, V. C., Chatzivasileiou, A. O., and Stephanopoulos, G. (2018). Metabolic engineering of Escherichia coli for the production of isoprenoids. *FEMS microbiology letters*, *365*, fny079.

Yoshida, S., Forno, D. A., Cock, J. H. and Gomez, K. A. (1976) Determination of sugar and starch in plant tissue. Laboratory Manual for Physiological Studies of Rice, third edition. IRRI, Los Baños, Phillipines, pp 46–49.

Zhao, J., Li, Q., Sun, T., Zhu, X., Xu, H., Tang, J., and Ma, Y. (2013). Engineering central metabolic modules of Escherichia coli for improving β-carotene production. *Metabolic engineering*, *17*, 42-50.

Zhou, K., Qiao, K., Edgar, S., and Stephanopoulos, G. (2015). Distributing a metabolic pathway among a microbial consortium enhances production of natural products. *Nature biotechnology*, *33*, 377.

GENERAL CONCLUSIONS

GENERAL CONCLUSIONS

1 Mutating one allele of *OsAPL2*, encoding the major endosperm large subunit of AGPase and the only subunit expressed in the cytosol, resulted in the unanticipated expression of *OsAPL2* and *OsAPS2b* (encoding the only cytosolic small subunit) in leaves .

2 The formation of a complete ectopic AGPase in the leaf cytosol did not increase overall starch synthesis.

3 Leaves in mature plants contained less starch than wild-type plants most likely reflecting the lower levels of plastidial *OsAPS2a*, increased SuSy activity, the increase in soluble sugars and/or the inability of *OsAPS1* to replace *OsAPS2a* function completely.

4 The new cytosolic AGPase was not sufficient to compensate for the loss of plastidial AGPase, probably because there is no wider starch biosynthesis pathway in the leaf cytosol and thus no pathway intermediates are shuttled between the two compartments.

5 Mutating the Wx gene, encoding granule-bound starch synthase 1, mainly expressed in rice endosperm, modified the relative abundance of amylose and amylopectin without changing the overall starch content.

6 Waxy mutant lines showed an increase in soluble sugars which may reflect a metabolic bottleneck caused by the mutation or an increase in starch degradation due to feedback regulation.

7 The waxy phenotype reflects the accumulation of amylopectin, at the expense of amylose, and is characterized by grains that are white and opaque rather than translucent like wild-type grains.

8 The overall GBSS activity in Waxy mutant lines exhibited an increase in GBSSII expression to compensate for the loss of GBSSI. The increase in GBSSII expression did not compensate for the loss of GBSSI, as reflected in the amylose content of the mutant lines.

9 Mutating the first exon of the rice Wx gene encoding GBSSI resulted in the expected partial loss of GBSS activity and the corresponding substantial reduction in amylose content, but also caused the unexpected expression of other downstream starch pathway genes, partly to deal with abundant intermediates and unusual starch structures.

10 Introducing an ectopic MVA pathway into rice plastids caused plant death in early stages of growth which may be due to the fact that the ectopic pathway depleted the acetyl CoA pool, essential for fatty acid biosynthesis in plastids and/or limiting the formation of other essential metabolic precursors in the cytosol.

11 Ectopic expression of MVA pathway genes and WR1 transcription factor modified the amount and type of terpenoids in rice seeds, increasing plastidial terpenoids such as carotenoids not present in wild-type seeds.

12 In the same plants fatty acid levels increased most likely as a consequence of increased amounts of Acetyl-CoA.

13 Expression of the ectopic MVA pathway resulted in alteration of phytohormone levels in rice leaves. We measured an increase in abscisic acid, a decrease in jasmonic acid levels and 1-aminocyclopropane-1-carboxylic acid and a different pattern of gibberellins with respect to wild-type.

14 The above hormonal changes are consistent with the early senescence phenotype observed in the plants.

GENERAL DISCUSSION

GENERAL DISCUSSION

Hunger is a global problem. Malnutrition affects more than 2 billion people worldwide and forecasts are rather pessimistic for any improvements in the light of the expected rise in world population to over 9 billion people by 2050. Agricultural productivity increases needed to meet this rise in population ranges between 100-110% compared with levels of 2005 (Tilman et al., 2011). Hidden hunger affects a further two billion people who even though consume enough calories; they do not obtain the essential nutrients needed to maintain a healthy live (Berman et al., 2013). In most developing countries, diet is based on individual crops with low amounts of nutrients (e.g. rice). Zero Hunger can only be achieved by using stable food biofortification. This biofortification can improve crop productivity and nutrient content.

Currently, there are two main strategies to biofortify food, conventional breeding and genetic engineering, or in principle a combination of both. Conventional breeding improves crops without using recombinant DNA technology, sometimes in combination with mutagenesis or marker assisted selection to introgress genes from distant relatives, which similarly would not occur in nature (Bai et al., 2011). It is limited, however, due to the relatively long lead time to have an impact and its dependence on a compatible gene pool. The long time frames to obtain nutritionally improved lines are one of the greatest challenges in conventional breeding (Pérez-Massot et al., 2013). In contrast, genetic engineering can generated biofortified crops by transferring genes directly into breeding lines, obtaining transgenic plants with enhanced nutritional traits. Compared to conventional breeding, genetic engineering has the advantages of speed, direct engineering of breeding lines, simplicity, multiple simultaneous biofortification of different nutrients and unrestricted access to genetic diversity including recombinant genes that do not occur in nature (Zhu et al., 2007). It is limited by issues of public perception and the high cost of deregulating GMO products.

For that reason, in the course of my PhD studies I used genetic engineering. I employed two different techniques to biofortify rice, a staple food crop, genome editing and conventional metabolic engineering.

I used CRISPR-Cas9 to knockout two starch pathway genes (APL2 and GBSSI) in rice. I selected these genes to modify starch content and starch structure because these genes are essential for rice grain quality without altering other agronomic traits. I elucidated AGPase gene function and its impact on the fertility of rice plants. Mutations in AGPase resulted in infertility due to the reduction in starch content.

Knocking out GBSSI resulted in modifications in the relative abundance of amylose and amylopectin without changing total starch content. Accumulation of amylopectin produced a more glutinous rice grain (slow-absorbing sugars) without altering other agronomic traits.

In the second strategy, I introduced an ectopic MVA pathway and overexpressed a transcription factor, Wrinkled1, in rice plastids to increase IPP and DMAPP pools, availability of Acetyl-CoA (precursor of MVA pathway) and improve downstream secondary metabolites of the MVA pathway such as carotenoids which are essential nutrients in the human diet.

In conclusion, I modulated rice grain quality and nutrient composition components contributing towards addressing the challenge of malnutrition. I selected rice because in most developing countries the diet is based principally on this staple. By using CRISPR-Cas9, I modified rice grain quality without altering other agronomic traits by mutating GBSSI and I identified AGPase genes as essential genes for fertility in rice plants. By using metabolic engineering, I introduced an ectopic MVA pathway into rice, together with Wrinkled1 that modified the amount and type of terpenoids in rice seeds and increased fatty acid levels.

References

Berman, J., Zhu, C., Pérez-Massot, E., Arjó, G., Zorrilla-López, U., Masip, G., Bai, C., Bassie, L., Capell, T. and Christou, P. (2013). Can the world afford to ignore biotechnology solutions that address food insecurity?. *Plant molecular biology*, *83*(1-2), 5-19.

Tilman, D., Balzer, C., Hill, J., and Befort, B. L. (2011). Global food demand and the sustainable intensification of agriculture. Proceedings of the National Academy of Sciences, 108: 20260-20264.

Bai, C., Twyman, R. M., Farré, G., Sanahuja, G., Christou, P., Capell, T., & Zhu, C. (2011). A golden era-pro-vitamin A enhancement in diverse crops. In Vitro Cellular & Developmental Biology-Plant, 47: 205-221.

Zhu, C., Naqvi, S., Gomez-Galera, S., Pelacho, A.M, Capell, T., and Christou, P. (2007). Transgenic strategies for the nutritional enhancement of plants. Trends in Plant Science 12: 548–555.

Pérez-Massot, E., Banakar, R., Gómez-Galera, S., Zorrilla-López, U., Sanahuja, G., Arjó, G., Miralpeix, B., Vamvaka, E., Farré, G., Berman, J., Sablaza, M., Yuan, D., Bai, C., Bassie, L., Capell, T., Chistou, P and Zhu, C. (2013). The contribution of transgenic plants to better health through improved nutrition: opportunities and constraints. Genes & nutrition, 8: 29.

OUTPUTS

Pérez, L., Soto, E., Villorbina, G., Bassie, L., Medina, V., Muñoz, P., Capell, T., Zhu, C., Christou, P. and 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 Res.* https://doi.org/10.1007/s11248-018-0089-7

Bortesi, L., Zhu, C., Zischewski, J., **Perez, L.**, Bassié, L., Nadi, R., Forni, G., Boyd-Lade, S., Soto, E., Jin, X., Medina, V., Villorbina, G., Muñoz, P., Farré, G., Fischer, R., Twyman, R., Capell, T., Christou, P. and Schillberg, S. (2016) Patterns of CRISPR/Cas9 activity in plants, animals and microbes. *Plant Biotechnol J.* 14, 2203–2216.

ANNEX

Transgenic Res (2018) 27:423–439
https://doi.org/10.1007/s11248-018-0089-7

ORIGINAL PAPER

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

Lucía Pérez · Erika Soto · Gemma Villorbina · Ludovic Bassie · Vicente Medina · Pilar Muñoz · Teresa Capell · Changfu Zhu · Paul Christou · Gemma Farré

Received: 12 April 2018 / Accepted: 4 August 2018 / Published online: 11 August 2018

Abstract The first committed step in the endosperm starch biosynthetic pathway is catalyzed by the cytosolic glucose-1-phosphate adenylyl transferase (AGPase) comprising large and small subunits encoded by the *OsAPL2* and *OsAPS2b* genes, respectively. *OsAPL2* is expressed solely in the endosperm so we hypothesized that mutating this gene would block starch biosynthesis in the endosperm without affecting the leaves. We used CRISPR/Cas9 to create two heterozygous mutants, one with a severely truncated and nonfunctional AGPase and the other with a C-terminal structural modification causing a partial loss of activity. Unexpectedly, we observed starch depletion in the leaves of both mutants and a corresponding increase in the level of soluble sugars. This reflected the unanticipated expression of both *OsAPL2* and *OsAPS2b* in the leaves, generating a complete ectopic AGPase in the leaf cytosol, and a corresponding decrease in the expression of the plastidial small subunit *OsAPS2a* that was only partially complemented by an increase in the expression of *OsAPS1*. The new cytosolic AGPase was not sufficient to compensate for the loss of plastidial AGPase, most likely because there is no wider starch biosynthesis pathway in the leaf cytosol and because pathway intermediates are not shuttled between the two compartments.

Keywords CRISPR/Cas9 · AGPase large subunit gene · Rice · Ectopic expression · Monoallelic mutations

Lucía Pérez and Erika Soto have contributed equally to the study.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11248-018-0089-7) contains supplementary material, which is available to authorized users.

L. Pérez · L. Bassie · V. Medina · P. Muñoz · T. Capell · C. Zhu · P. Christou (✉) · G. Farré (✉)
Department of Plant Production and Forestry Science, School of Agrifood and Forestry Science and Engineering (ETSEA), University of Lleida-Agrotecnio Center, Av. Alcalde Rovira Roure 191, 25198 Lleida, Spain
e-mail: christou@pvcf.udl.es

G. Farré
e-mail: g.farre@pvcf.udl.cat

E. Soto · G. Villorbina
Department of Chemistry, University of Lleida-Agrotecnio Center, Lleida, Spain

P. Christou
Catalan Institute for Research and Advanced Studies (ICREA), Barcelona, Spain

Introduction

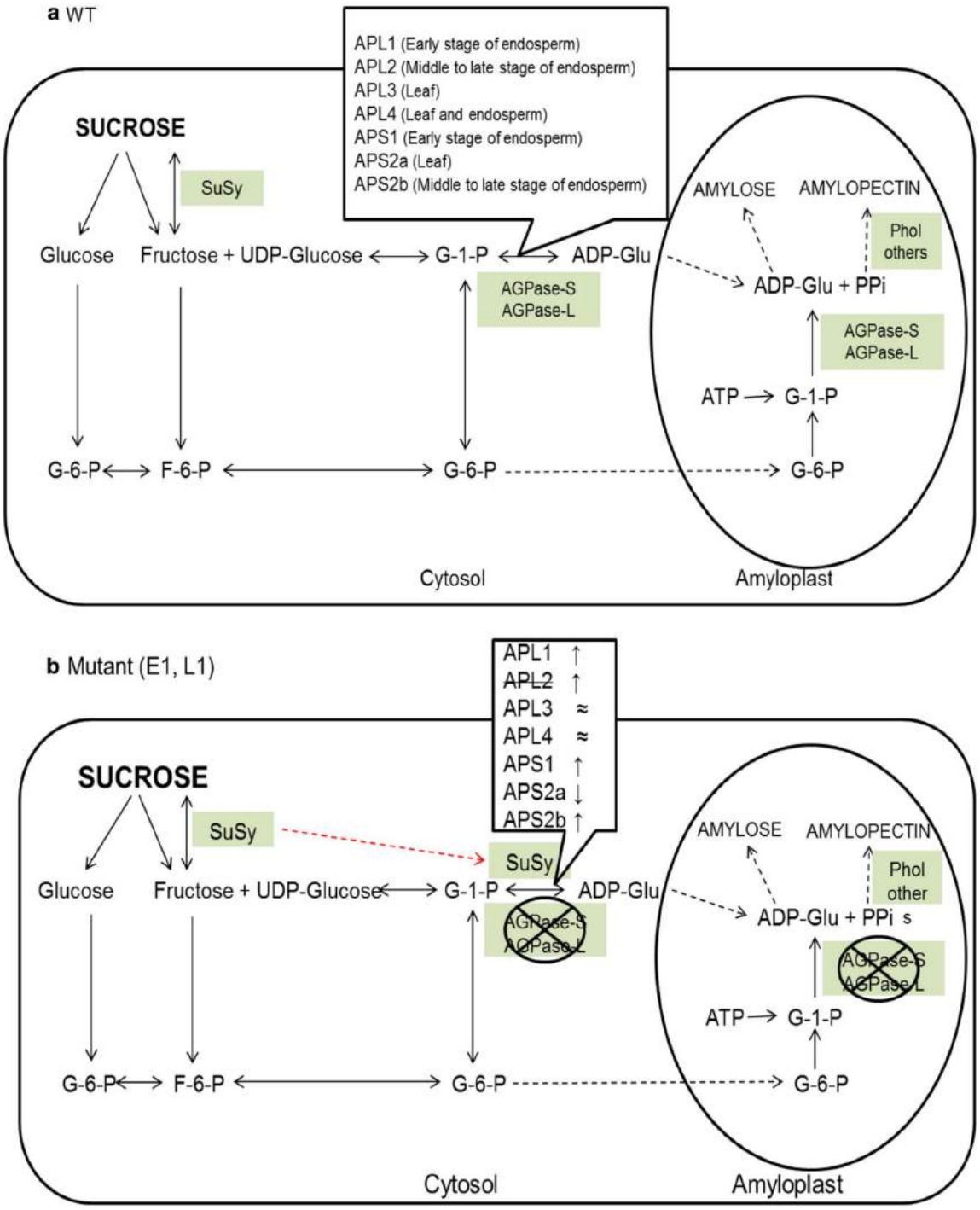
Rice (*Oryza sativa* L.) is one of the most important food crops in the world, accounting for 21% of the calories and 15% of the protein consumed by humans globally, and more than 70% of calories consumed by developing country populations in Asia (Yuan et al. 2011; Zhu et al. 2013). The major energy-providing component of rice grains is starch, a mixture of the two glucose-derived polysaccharides amylose and amylopectin. Amylose predominantly comprises linear chains of α (1,4) linked glucose residues, whereas amylopectin contains additional α (1,6) linked branches every 24–30 residues (Martin and Smith 1995). Starch from different plant species varies significantly in its physicochemical properties due to the relative proportions of amylose and amylopectin, and differences in chain length and/or amylopectin branching density (Jobling 2004).

Starch synthesis in plants involves the conversion of glucose 1-phosphate to ADP-glucose by the ATP-dependent enzyme glucose-1-phosphate adenylyltransferase (AGPase), and then the polymerization of ADP-glucose units by starch synthase to form amylose (Fig. 1). Starch branching enzyme (SBE) introduces α (1,6) linked glycosidic bonds to generate amylopectin, a reaction that is reversed by the starch debranching enzyme isoamylase (DBE). In an alternative route, sucrose synthase in the endosperm cytosol converts sucrose and ADP directly into fructose and ADP-glucose, and the latter is imported into the amyloplasts for starch synthesis (Li et al. 2013). Different forms of starch can be generated by mutating the genes encoding starch biosynthetic enzymes, but the outcome is complicated by the existence of multiple tissue-specific isoenzymes in many plants and the presence of multiple subunits per enzyme.

Conventional mutagenesis such as irradiation, chemical mutagenesis and T–DNA/transposon insertional mutagenesis generate random lesions in DNA sequences and require the screening of large populations to isolate useful mutants. These techniques have been largely supplanted by targeted mutagenesis using designer nucleases, particularly CRISPR/Cas9 (reviewed by Bortesi et al. 2016; Zhu et al. 2017). CRISPR/Cas9 mutagenesis is based on a bacterial defense system that targets invasive DNA by collecting DNA sequences as clustered regularly interspaced short palindromic repeats (CRISPRs) (Doudna et al. 2014). The transcription of these repeats into CRISPR RNAs, which pair with a CRISPR-associated (Cas) nuclease such as Cas9, allows the same invading DNA to be targeted and destroyed if it is encountered again (Lee et al. 2016a, b). This mechanism can be harnessed for targeted mutagenesis if a synthetic guide RNA (sgRNA) is designed to match a genomic target instead of an invasive DNA sequence. The delivery of sgRNA and Cas9 to plant cells results in a double strand break (DSB) at the target site, which is generally repaired by the error-prone non-homologous end joining (NHEJ) pathway, resulting in small insertions or deletions at the site of the DSB that disrupt gene function by causing a frameshift mutation. The wild-type Cas9 generates a blunt DSB at the target site, which is specified by a 20-nt spacer sequence in the sgRNA. An alternative approach is to mutate one of the two endonuclease domains in Cas9 so that the enzyme only cleaves one DNA strand. A DSB therefore requires two such Cas9 nickases, and if these are recruited by different sgRNAs annealing a few base pairs apart, a staggered break is introduced within a 40-nt target sequence, significantly increasing the specificity of targeting and all but eliminating off-target cleavage activity (Fauser et al. 2014; Mikami et al. 2016; Ran et al. 2013).

High-amylose rice mutants have been produced by targeting the genes encoding SBEI and SBEIIb using CRISPR/Cas9 (Sun et al. 2017). Bi-allelic T0 mutants with insertions and deletions at the target sites were generated and the mutations were stably transmitted to progeny. The *OsBEIIb* mutants accumulate higher

Fig. 1 The coordination of different starch biosynthetic genes in rice (modified from Pandey et al. 2012). **a** Expression pattern of AGPase subunits in wild type. **b** Expression patterns of AGPase subunits in leaves in mutants L1 and E1 (↑ upregulated; ↓ downregulated; ≈ similar expression as wild type). The red dotted line represents the alternative pathway and the crossed out circles represent the loss of function of the enzyme. Abbreviations: SuSy (sucrose synthase), G-1-P (glucose-1-phosphate), G-6-P (glucose-6-phosphate), ADP-Glu (ADP-Glucose), PPi (inorganic diphosphate); F-6-P (fructose-6-phosphate); ATP (adenosine triphosphate); ADP (adenosine diphosphate), UDP (uridine diphosphate), PhoI (plastidial α-glucan phosphorylase); AGPase (ADP-glucose pyrophosphorylase); APS1 (AGPase small subunit 1); APS2a (AGPase small subunit 2a); APS2b (AGPase small subunit 2b); APL1 (AGPase large subunit 1); APL2 (AGPase large subunit 2); APL3 (AGPase large subunit 3); APL4 (AGPase large subunit 4). (Color figure online)


a WT
APL1 (Early stage of endosperm)
APL2 (Middle to late stage of endosperm)
APL3 (Leaf)
APL4 (Leaf and endosperm)
APS1 (Early stage of endosperm)
APS2a (Leaf)
APS2b (Middle to late stage of endosperm)
SUCROSE
SuSy
Glucose
Fructose + UDP-Glucose
G-1-P
ADP-Glu
AGPase-S
AGPase-L
AMYLOSE
AMYLOPECTIN
Phol
others
ADP-Glu + PPi
AGPase-S
AGPase-L
ATP
G-1-P
G-6-P
F-6-P
G-6-P
G-6-P
Cytosol
Amyloplast
b Mutant (E1, L1)
APL1 ↑
APL2 ↑
APL3 ≈
APL4 ≈
APS1 ↑
APS2a ↓
APS2b ↑
SUCROSE
SuSy
SuSy
Glucose
Fructose + UDP-Glucose
G-1-P
ADP-Glu
AMYLOSE
AMYLOPECTIN
Phol
other
ADP-Glu + PPi
s
ATP
G-1-P
G-6-P
F-6-P
G-6-P
G-6-P
Cytosol
Amyloplast

proportion of amylose and debranched amylopectin in the seeds than normal (Sun et al. 2017). *OsBEIIb* has also been targeted using two different sgRNAs with different activity scores and different degrees of conservation with the paralogous gene *OsBEIIa* to confirm the absence of off-target mutations (Baysal et al. 2016). Another study using wild-type Cas9 targeted three different sites in *OsWaxy* encoding granule-bound starch synthase (GBSS). Only one or two of the sites were mutated in the resulting primary transformant, but the amylose content in T1 seeds was reduced from 14.6 to 2.6% (Ma et al. 2015).

Although the CRISPR/Cas9 system has been used to target GBSS and SBE, it has not yet been used to target AGPase, which catalyzes the first step in the starch biosynthesis pathway (Tang et al. 2016; Lee et al. 2016a, b). Rather than altering the ratio of amylose and amylopectin, blocking AGPase would therefore prevent starch synthesis all together. AGPase comprises two large and two small subunits that together form a hetero-tetrameric complex (Tuncel et al. 2014; Ballicora 2003). In rice, there are two small subunit genes (*OsAPS1* and *OsAPS2*, the latter producing the two mRNA variants *OsAPS2a* and *OsAPS2b* by alternative splicing) and four large subunit genes (*OsAPL1*, *OsAPL2*, *OsAPL3 and OsAPL4*). AGPase is the key enzyme for starch synthesis. Its regulation is controlled by 3-phosphoglycerate (3-PGA) which upregulates its activity and by inorganic phosphate (Pi) which downregulates AGPase (Preiss 1982). All transcripts are tissue-specific and the corresponding proteins are localized differentially (Ohdan et al. 2005). *OsAPS2a* is mainly expressed in the leaves and the protein is localized in plastids, whereas *OsAPS2b* is only expressed in the endosperm and *OsAPL2* is mainly expressed in the endosperm and the proteins are localized in the cytosol. *OsAPL1* and *OsAPS1* are mostly expressed in early endosperm plastids, whereas *OsAPL3* is expressed in leaf plastids. *OsAPL4* is expressed at high levels in leaf plastids but at low levels in endosperm plastids (Ohdan et al. 2005; Lee et al. 2007). Because *OsAPL2* and *OsAPS2b* are the only cytosolic subunits and *OsAPS2b* is expressed exclusively in endosperm there is no cytosolic AGPase in the leaves. Mutations in *OsAPL2* and *OsAPS2b* cause a marked reduction in starch levels (Lee et al. 2007; Ohdan et al. 2005; Tsai and Nelson 1966; Tester et al. 1993; Johnson 2003; Muller et al. 1992; Giroux 1994, Tang et al. 2016). In the rice *OsAPL1* mutant, the starch content in the leaves was reduced to ~ 5% of wild-type levels, but growth and development were normal (Rosti et al. 2007). Homozygous *OsAPL3* mutants displayed ~ 23% of wild-type AGPase activity and accumulated much less starch than normal in the culm (Cook et al. 2012). The shrunken mutant has a nonfunctional *OsAPS2* and therefore lacks both the *OsAPS2a* and *OsAPS2b* transcripts, and exhibits ~ 20% of wild-type AGPase activity (Kawagoe et al. 2005; Tuncel et al. 2014). Mutation of *OsAPS2b* caused *OSAPS1* and *OsAPL1* transcript levels to increase in leaves whereas *OsAPL3* transcript levels remained unaffected (Ohdan et al. 2005).

The absence of a cytosolic AGPase in leaves suggests that manipulation of *OsAPL2* and *OsAPS2b* might offer a strategy to modulate starch production in the endosperm without affecting starch metabolism in vegetative tissues. Our experiments focused on *OsAPL2* because that is the only AGPase subunit gene that is largely expressed in the endosperm. *OsAPL2* encodes a 518 amino acid polypeptide with a catalytic site spanning residues 88–364. The catalytic site is composed of α-helices and β-sheets that form a substrate binding cleft. A further α-helix and β–sheet outside the catalytic center are required to maintain the correct tertiary conformation and bind with the other subunit, APS2b, to form the tetrameric structure (Figs. 2 and 3) (Zhang et al. 2012). Experiments with null and missense *OsAPL2* mutants indicate that this subunit plays an important role in both the catalytic and regulatory properties of AGPase (Tuncel et al. 2014). We hypothesized that knocking out *OsAPL2* would have no effect on starch levels in leaves because its expression is very low in this tissue. We mutated the *OsAPL2* using two different strategies. We targeted the first exon aiming to truncate the enzyme substantially and abolish its activity entirely. For this purpose, we used Cas9 nickase with two closely-spaced targets in order to generate a deletion. In separate experiments, we targeted an exon downstream of the active site, to maintain some activity but perturb the tertiary structure. In this case we used the wild-type Cas9 with a single target in order to generate indels. We anticipated a reduction in endosperm starch levels in both lines and a corresponding increase in the abundance of soluble sugars, but we anticipated no impact on starch metabolism in leaves.

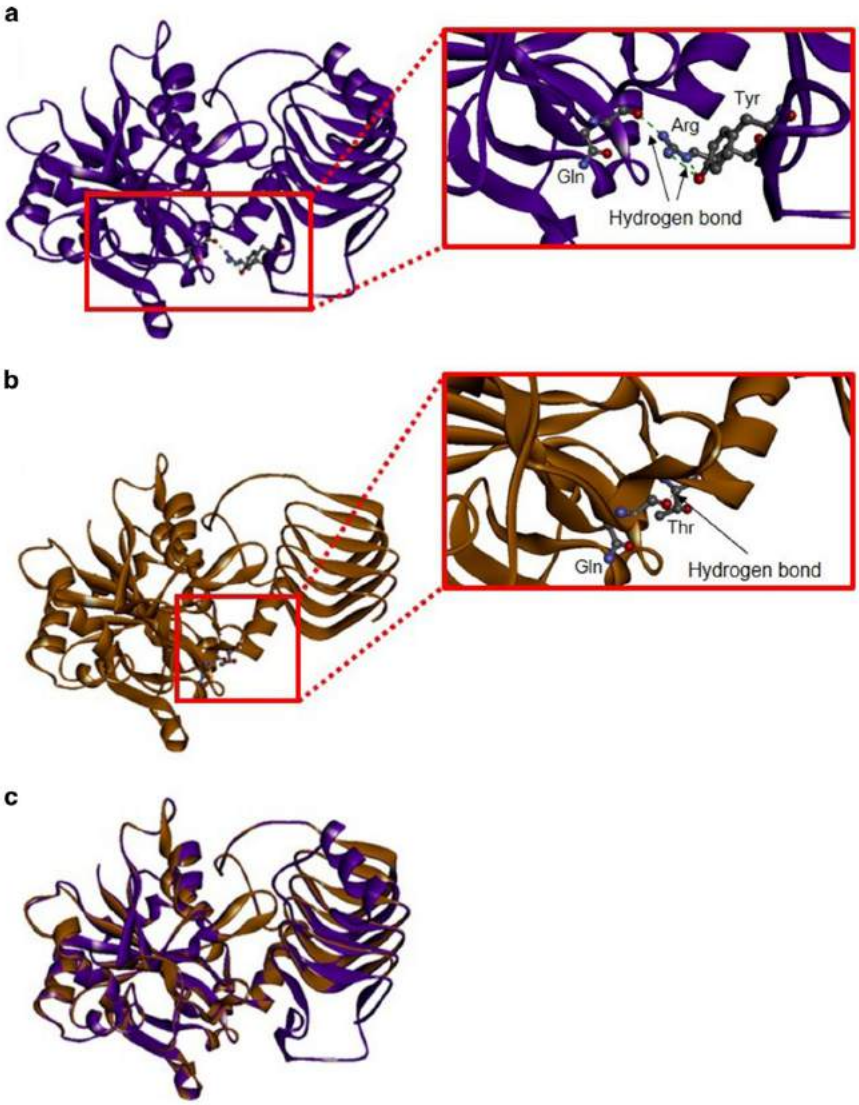


Fig. 2 Comparisons of protein structure. **a** 3-D model of wild-type APL2 and zoom of the homology modeled structure where hydrogen bonding between Gln 98 and Arg 500 and Tyr 498 is shown as a dashed green line. **b** 3-D model of the mutated ADP-glucose pyrophosphorylase (AGPase) in L1 and zoom of the homology modeled structure in the L1 mutant where the hydrogen bond between Gln 98 and Thr 351 is shown. **c** Superimposed wild-type (in purple) and mutated protein in L1 (in orange). The mutated *APL2* sequences were translated into polypeptides (http://web.expasy.org/translate/) and automated homology modeling was carried out using SWISS-MODEL (https://swissmodel.expasy.org/interactive/wrGgpx/models) with the potato tuber AGPase SS (Protein Databank: 1yp4) as the template. The model of the mutant protein was superimposed on the wild-type version using DS Visualizer (http:/accelrys.com/products/collaborative-science/biovia-discovery-studio/visualization.html). (Color figure online)

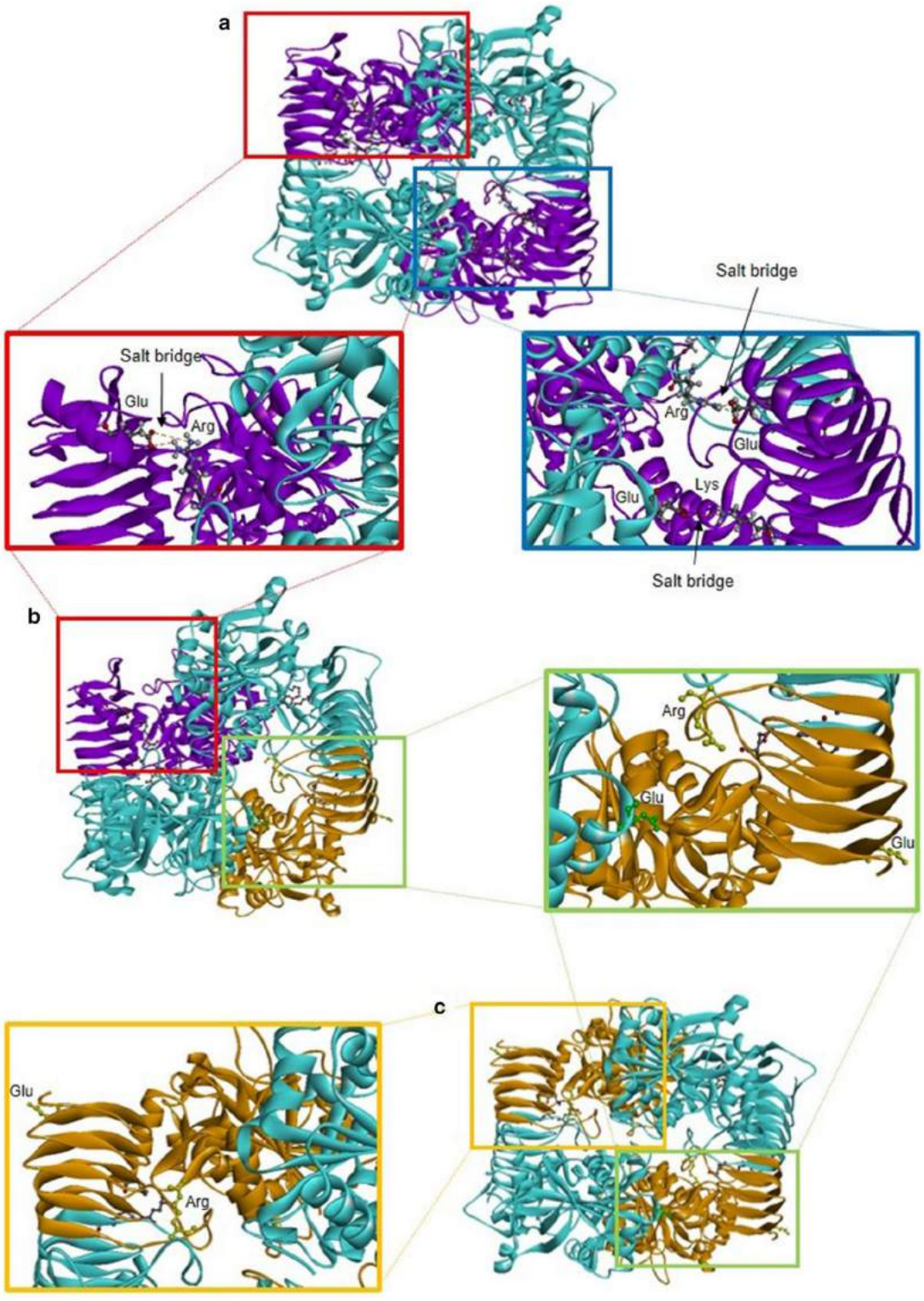
Materials and methods

Target sites and sgRNA design

Target sites for wild-type Cas9 (single sgRNA) and Cas9D10A (two sgRNAs targeting adjacent sites) were selected within the *OsAPL2* coding sequence (GenBank AK071497.1) using E-CRISP (Heigwer et al. 2014) with the following parameters: only NGG PAM, only G as 5′ base, off-target tolerates many mismatches, non-seed region ignored, introns ignored. The sgRNAs (Fig S1a) catalytic efficiencies were designed to minimize off-targets. The catalytic efficiency of the sgRNAs was predicted using gRNA scorer (Chari et al. 2015).

Vector construction

The wild-type Cas9 vector pJIT163-2NSCas9 and the sgRNA vector pU3-gRNA were obtained from Dr. C. Gao, Chinese Academy of Sciences, Beijing, China (Shan et al. 2014). The nickase vector pJIT163-


a
Salt bridge
Glu
Arg
Salt bridge
Arg
Glu
Lys
Glu
Salt bridge
b
Arg
Glu
Glu
c
Glu
Arg

◀**Fig. 3** Comparisons of heterotetrameric structure. **a** Heterotetrameric structure of wild type AGPase that consists of two large subunits (APL2, in purple) and two small subunits (APS2b, in blue). Salt bridge interaction (between oppositely charged residues that are sufficiently close to each other for electrostatic attraction) (shown in blue square) between Lys 508 (Chain C) and Glu 161(Chain B) and between Glu 444 (Chain C) and Arg 335 (Chain D). Salt bridge interaction (shown in red squared) between Glu 444 (Chain A) and Arg 335 (Chain B). **b** Heterotetrameric structure of AGPase that consists of a wild type large subunit (APL2, in purple), a mutated large subunit (L1, in orange) and two small subunits (APS2b, in blue). Salt bridge interaction (shown in red squared) between Glu 444 (Chain A) and Arg 335 (Chain B) and (shown in green squared) between Arg 335 (Chain D, in yellow), Glu 444 (Chain C, in yellow) and Glu 161 (Chain B, in green). **c** Heterotetrameric structure of AGPase consisting of two mutated large subunits (L1, in orange) and two small subunits (APS2b, in blue). Salt bridge interaction (shown in orange squared) between Arg 335 (Chain B, in yellow) and Glu 444 (Chain A, in yellow) and (shown in green squared) between Arg 335 (Chain D, in yellow), Glu 444 (Chain C, in yellow) and the Glu 161 (Chain B, in green). (Color figure online)

2NSCas9D10A was constructed in-house by mutating the *cas9* gene in vector pJIT163-2NSCas9 to produce Cas9D10A and combining this with the maize *ubiquitin*-1 promoter and Cauliflower mosaic virus 35*S* terminator. The three sgRNAs described above were prepared as synthetic double-stranded oligonucleotides and introduced separately into pU3-gRNA at the AarI restriction site; thus all genomic sites with the form 5′-(20)-NGG-3′ can be targeted. The *hpt* selectable marker gene was provided on a separate vector as previously described (Christou et al. 1991).

Rice transformation and recovery of transgenic plants

Seven-day-old mature zygotic embryos (*Oryza sativa* cv. EYI) were transferred to osmotic medium (MS medium supplemented with 0.3 g/L casein hydrolysate, 0.5 g/L proline, 72.8 g/L mannitol and 30 g/L sucrose) 4 h before bombardment with 10 mg gold particles coated with the transformation vectors. The Cas9 vector (wild type or nickase), the corresponding sgRNA vector(s) and the selectable marker *hpt* were introduced at a 3:3:1 ratio for wild-type Cas9 and a 3:3:3:1 ratio for the nickase with two sgRNAs (Christou et al. 1991; Sudhakar al. 1998; Valdez et al. 1998). The embryos were returned to osmotic medium for 12 h before selection on MS medium (MS medium supplemented with 0.3 g/L casein, 0.5 g/L proline and 30 g/L sucrose) with 50 mg/L hygromycin and 2.5 mg/l 2,4-dichlorophenoxyacetic acid in the dark for 2–3 weeks. Callus was maintained on selective medium for 6 weeks with sub culturing every 2 weeks as described (Farré et al. 2012). Transgenic plantlets were regenerated and hardened off in soil. Negative controls were regenerated plants from the same experiment which were no transformed (i.e. they did not contain Cas9WT/D10A, hpt and sgRNA). Negative controls behaved exactly like wild type plants.

Confirmation of the presence of *Cas9* and gRNA

Genomic DNA was isolated from callus. leaves and panicles of regenerated plants by phenol extraction and ethanol precipitation (Bassie et al. 2008; Kang and Yang 2004). The presence of the wild-type Cas9 sequence was confirmed by PCR using primers 5′-GTC CGA TAA TGT GCC CAG CGA-3′ and 5′-GAA ATC CCT CCC CTT GTC CCA-3′; the presence of the Cas9D10A sequence was determined using primers 5′-GCA AAG AAC TTT CGA TAA CGG CAG CAT CCC TCA CC-3′ and 5′-CCT TCA CTT CCC GGA TCA GCT TGT CAT TCT CAT CGT-3′; and the presence of the pU3-gRNA vectors was confirmed using the conserved primers 5′-TTG GGT AAC GCC AGG GTT TT-3′ and 5′-TGT GGA TAG CCG AGG TGG TA-3′.

Analysis of induced mutations

The *OsAPL2* mutation induced by the wild-type Cas9 was detected by PCR using primers 5′-CGT TAG CAT CGG GTG TGA ACT-3′ and 5′-GGA CCC CCT ATC ATA CGC AGT-3′. The *OsAPL2* mutation induced by Cas9D10A was similarly detected using primers 5′-CTT GTT GTT CAG GAT GGA TGC–3′ and 5′–GTG CAT TGT GCC TGT GGA A-3′. The PCR products were sequenced using an ABI 3730xl DNA analyzer by Stabvida (http://www.stabvida.com/es/). To confirm the mutations, PCR products generated using the primers listed above were purified using the Geneclean® II Kit (MP Biomedicals), transferred to the pGEM-T Easy vector (Promega) and introduced into competent *Escherichia coli* cells. From *E. coli*, PCR fragments of ~ 400 bp were purified, cloned and sequenced using an ABI 3730xl DNA analyzer by

Stabvida. At least five clones per PCR product were sequenced using primerM13Fwd (Table S1).

Protein structural modelling

The mutated *OsAPL2* sequences were translated into polypeptides (http://web.expasy.org/translate/) and automated homology modeling was carried out using SWISS-MODEL (https://swissmodel.expasy.org/interactive/wrGgpx/models) with the potato tuber AGPase SS (Protein Databank: 1yp4) as the template. The model of the mutant protein was superimposed on the wild-type version using DS Visualizer (http:/accelrys.com/products/collaborative-science/biovia-discovery-studio/visualization.html). Heterotetrameric AGPase structures were energy minimized using DS Visualizer.

Enzymatic activity and carbohydrate levels

Leaf extracts were prepared as previously reported (Tang et al. 2016) for the measurement of AGPase activity (ADP-glucose pyrophosphorylase, E.C. 2.7.7.27) in the forward direction (Nishi et al. 2001), and sucrose synthase activity (UDP-Glc:D-fructose 2-glucosyltransferase, EC 2.4.1.13) as previously described (Doehlert et al. 1988). Flag leaf samples harvested at 19:00 h were homogenized under liquid nitrogen and extracted in perchloric acid to measure the starch content, or in ethanol to measure the soluble sugar content. The quantity of each carbohydrate was determined by spectrophotometry (Yoshida et al. 1976).

RNA extraction and real-time qRT-PCR analysis

Total leaf RNA was isolated using the RNeasy Plant Mini Kit (Qiagen) and DNA was digested with DNase I (RNase-free DNase Set, Qiagen). Total RNA was quantified using a Nanodrop 1000 spectrophotometer (Thermo Fisher Scientific) and 2 µg total RNA was used as template for first strand cDNA synthesis with Quantitech® reverse transcriptase (Qiagen) in a 20-µL total reaction volume, following the manufacturer's recommendations. Real-time qRT-PCR was performed on a BioRad CFX96™ system using 20-µL mixtures containing 5 ng synthesized cDNA, 1 × iQ SYBR green supermix and 0.5 µM forward and reverse primers. The *OsAPL1*, *OsAPL3*, *OsAPL4*, *OsAPS1*, *OsAPS2a/b*, *OsAPL2* and *OsPho1* transcripts were amplified using the primers listed in Table S1, as described by Tang et al. (2016) Primers at the end of *OsAPL2* were designed to amplify the non-common region between the WT and mutants E1 and L1 (Table S1). Serial dilutions of cDNA (80–0.0256 ng) were used to generate standard curves for each gene. PCR was performed in triplicate using 96-well optical reaction plates. Values represent the mean of three biological replicates ± SE. Amplification efficiencies were compared by plotting the ΔCt values of different primer combinations of serial dilutions against the log of starting template concentrations using the CFX96™ software. The rice housekeeping *OsUBQ5* (LOC_Os01g22490) was used as an internal control.

Results

Design of a CRISPR/Cas9 mutation strategy

We designed three sgRNAs targeting the *OsAPL2* and selected the target sequences using E–CRISP to minimize the likelihood of off-targets. For the wild-type Cas9, we selected a single target site in exon 13 (hereafter named T1), whereas for Cas9D10A nickase we selected two adjacent targets in the first exon (hereafter named T2 and T3). The locations of each target are shown in Fig S1a. The sgRNA cassettes were separately transferred to the pU3-gRNA vector and introduced into rice embryos along with the rice codon-optimized wild-type Cas9 or Cas9D10A sequence, and the selectable marker *hpt* conferring hygromycin resistance.

Recovery and analysis of mutant lines

We regenerated transgenic plants representing each transformation strategy. Sequencing *E. coli* colonies revealed that the mutation generated by wild–type Cas9 in an early transformant (mutant L1) was an insertion of a single nucleotide at site T1 (Fig S1b) whereas a mutation generated by the Cas9D10A nickase (mutant E1) in another transformant was a deletion of 11 nucleotides at sites T2/T3 (Fig S1c). We focused on heterozygous mutations in order to determine whether the expression of the corresponding wild-type allele was affected in regenerated plants (see below). As well as testing for on-target mutations,

E-CRISP identified potential off-target cleavage sites at three loci based on the number of mismatches allowed in the target sequence and 2 bp upstream of the DSB. A single potential off-target was identified for mutant L1 in the *OsAGPlar* (EU267957.1) whereas two potential off-targets were identified for mutant E1 in the *OsXPO7* (LOC4334606) and the *OsNAT6* (LOC4330689). Sequencing these loci revealed no evidence of off-target mutations.

Structural comparisons

In order to investigate potential changes at the protein level, we translated the L1 and E1 mutant *OsAPL2* sequences and generated three-dimensional models using the SWISS-MODEL program. Compared to the wild-type sequence (Fig S2a), the E1 mutation resulted in an early stop codon such that the residual product was only 21 amino acids in length (Fig S2c). In contrast, the insertion of a single nucleotide in mutant L1 generated a change in the protein sequence downstream of the catalytic site (Fig S2b). Clearly, the E1 mutation generated a non-functional product. However, much of the L1 protein sequence was preserved and by superimposing the mutant sequence over that of the wild-type protein we found that the L1 mutant also contained a variant loop structure in the same vicinity. The resulting changes to the active site are shown in Fig. 2. The key difference is the change in orientation of the Gln 98 side chain relative to the active site, which affects the topology of the substrate-binding cleft and influences substrate accommodation (Tuncel et al. 2014; Tang et al. 2016) (Fig. 2b). The change occurs because hydrogen bonding between residue Arg 500 and Tyr 498 is eliminated (Fig. 2a) and a new hydrogen bond is formed with Thr 351 (Fig. 2b). The heterotetrameric structure of the AGPase comprises two *APS2b* and two *APL2* subunits (Fig. 3a). The four subunits of the wild type heterotetramer are stabilized by three salt bridges between Lys 508 (Chain C)-Glu161 (Chain B), Glu 444 (Chain C)-Arg 335 (Chain D) and Glu 444 (Chain A)-Arg 335(Chain B), that are eliminated in the mutated forms.

Analysis of AGPase and sucrose synthase activity

We determined the direct impact of each mutation on AGPase activity by comparing enzyme activity in the flag leaves of the mutant and wild-type plants as described by Nishi et al. (2001), using three biological replicates and two technical replicates for each biological sample (Fig. 4a). No analysis was possible in seed as both mutants were infertile. AGPase activity in mutant E1 was low (10.6 ± 1.8 mmol NADPH/min/mg FW) whereas activity in mutant L1 was much higher (30.7 ± 1.4 nmol NADPH/min/mg FW) than wild-type plants (15.5 ± 1.8 nmol NADPH/min/mg FW). AGPase activity in the corresponding negative controls was similar to wild-type levels, as expected. We also measured the sucrose synthase activity in each mutant because this provides an alternative route for the synthesis of ADP-glucose, which might be induced when the primary pathway is blocked (Fig. 4b). Sucrose synthase activity in the E1 mutant was higher (3.5 ± 0.27 nmol sucrose/min/mg FW) whereas the L1 mutant had similar activity to the wild-type plants and negative controls (1.4–2.1 nmol sucrose/min/mg FW).

Analysis of AGPase family gene expression

Next, we measured the expression of Os*APL2* and transcripts of other starch biosynthetic genes (*OsAPL1*, *OsAPL3*, *OsAPL4*, *OsAPS1*, *OsAPS2a*, *OsAPS2b* and *OsPhoI*) in the flag leaves to determine whether mutating the *OsAPL2* gene had an indirect regulatory impact on genes encoding other subunits. Our results indicated that the relative expression levels of *OsAPS2b* and *OsAPL2* increased significantly in both mutants compared to wild-type plants. This was surprising given that *OsAPS2b* expresses mostly in the endosperm, *OsAPL2* is expressed at low levels in leaves and the tetramer between *APL2* and *APS2b* is not usually expressed at significant levels in WT leaves. Mutating *OsAPL2* therefore appears to promote the ectopic expression of a cytosolic AGPase which is not typically present in leaves. In contrast, *OsAPS2a* expression levels declined significantly in both mutants compared to wild-type plants, whereas the expression of *OsAPL1* and *OsPhoI* was significantly higher than wild-type levels in both mutants (Fig. 5). The expression of *OsAPS1* was 9 times higher than wild-type levels in mutant E1, and 2 times higher in mutant L1 (Fig. 5b). *OsAPL3* and *OsAPL4* expression remained similar to wild-type levels in both mutants.

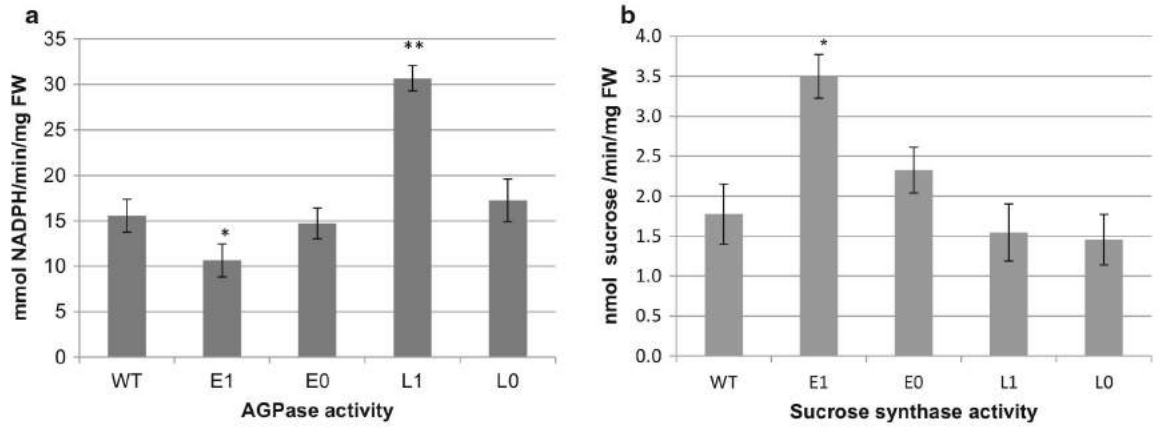


Fig. 4 **a** AGPase activity and **b** sucrose synthase activity in the flag leaves of wild-type (WT), mutants, and negative controls. E0 is the negative control (negative transformants regenerated at the same time as the mutated lines) for E1; L0 is the corresponding negative control for L1. Values are means ± SDs (n = 3 biological replicates, 2 technical replicates for each biological replicate). The asterisk indicates a statistically significant difference between WT and mutant, as determined by a Student's *t* test (*$P < 0.05$, **$P < 0.001$)

OsAPL2 has three alleles in the genome and the E1 and L1 mutants are heterozygous (at least one of the alleles is mutated while the remaining is/are WT). To determine which allele (WT or mutated) is responsible for the increase in *OsAPL2* expression in leaves, we designed primers at the end of the gene in the non-common region between the WT and mutants E1 and L1 to only amplify WT allele(s). We measured expression levels in the flag leaves. *APL2* expression in E1 remained constant irrespective of using non-common or common region primers, suggesting that the WT allele was responsible for the increased expression of the gene. L1 retained 90% of the expression of *OsAPL2* using non- common region primers suggesting that the increase in *OsAPL2* expression was also due to the WT allele (Fig. 5c).

Analysis of starch and sugar levels

A significant decline in starch content was observed in the leaves of both mutants. E1 showed a remarkable decrease in starch content to 2% by weight, ~ 15% of the normal levels found in wild-type leaves. L1 accumulated 7% starch by weight, approximately half the normal level. The negative controls had similar levels of starch as the wild-type plants (Fig. 6a). In contrast, the soluble sugar content of both mutants was higher than wild-type levels: ~ 40% higher in E1 and ~ 20% higher in L1. The negative controls contained similar amounts of soluble sugars to the wild-type plants (Fig. 6b). These results are surprising given that the *OsAPL2* mutation appeared to promote the ectopic expression of a cytosolic AGPase in leaves that is normally expressed at very low levels in wild-type plants, indicating that starch synthesis in leaves remains dependent on the plastidial AGPase even when cytosolic large and small subunits are expressed.

Discussion

Amylose is synthesized by AGPase and GBSS whereas amylopectin also requires SBE and DBE to introduce and refine its branching structure (Ohdan et al. 2005). The heterotetrameric AGPase catalyzes the first committed step in starch synthesis. In rice there are two genes encoding small catalytic subunits (*APS1* and *APS2*) and four encoding larger regulatory subunits (*APL1*, *APL2*, *APL3* and *APL4*). Furthermore, *OsAPS2* produces two distinct polypeptides through alternative splicing: *APS2a* and *APS2b*, the former including a transit peptide for import into the plastids and the latter lacking this sequence causing it to remain in the cytosol (Tang et al. 2016). AGPases are key enzymes in the starch biosynthesis pathway and are regulated by the ratio 3-PGA/Pi (Preiss et al. 1982). Among the larger regulatory subunits, only APL2 lacks a transit peptide and remains in the

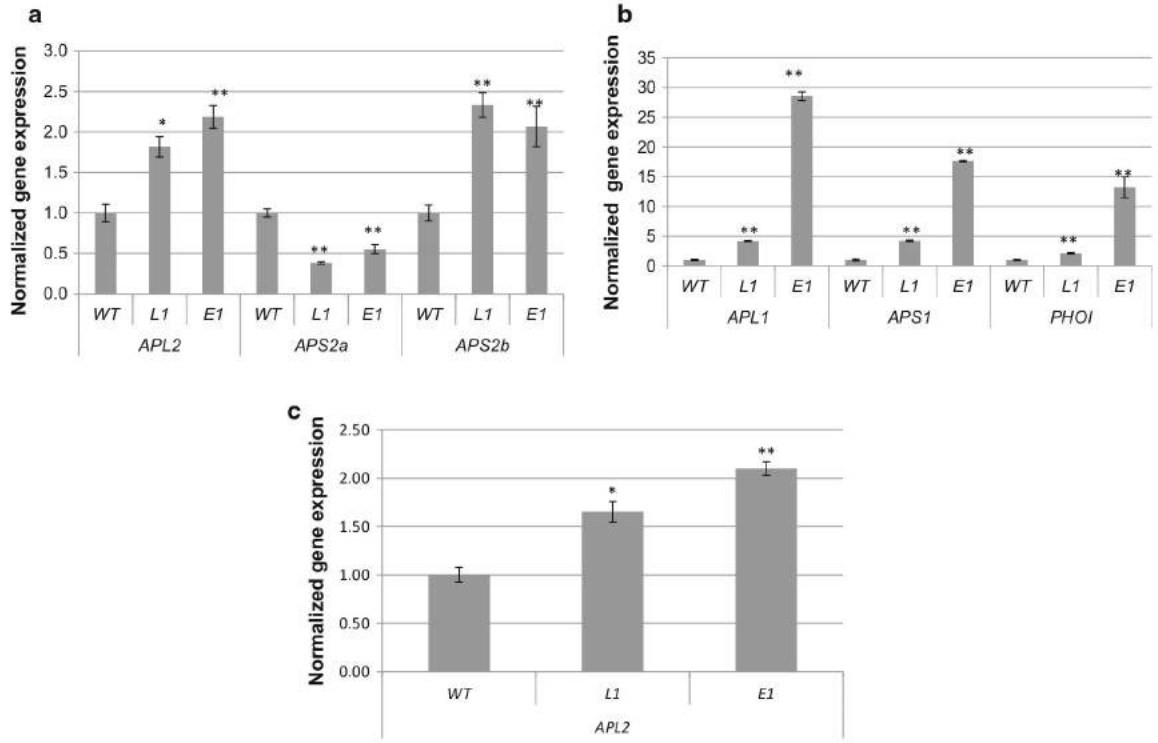


Fig. 5 Relative expression levels of rice AGPase family genes in the flag leaves of wild-type and mutants. **a** *OsAPL2*, *OsAPS2a* and *OsAPS2b*. **b** *OsAPL1*, *OsAPS1* and *OsPHOI*. **c** *OsAPL2* amplifying C-terminal end of the protein (non-common between WT and mutants). Values are means ± SDs (n = 3 technical replicates). The asterisks indicate a statistically significant difference between wild-type (WT) and mutants, as determined by a Student's *t*-test (*$P < 0.05$; **$P < 0.01$)

cytosol, whereas the other three subunits are imported into the plastid. Thus, only one AGPase assembles in the cytosol, comprising subunits APL2 and APS2b.

APS1 and *APL1* are expressed strongly in the early endosperm, whereas expression of *APS2b* and *APL2* begins 3 days after fertilization and remains at high levels thereafter. *APL3* is expressed at low levels throughout seed development, and the *APL4* and *APS2a* transcripts are barely detected at all (Ohdan et al. 2005). This suggests that the endosperm contains one predominant plastid AGPase, comprising APL1 and APS1, which is important during early development, and one predominant cytosolic AGPase, comprising *APL2* and *APS2b*, which is important during the middle and late stages of development, when starch accumulates (Ohdan et al. 2005). During seed development, starch is accumulated in amyloplasts which serve as a reservoir for the germinating seed. The major tetramer (*APL2–APS2b*) is in the cytosol and acts to generate stable (i.e. not transitory) starch in this compartment (Rychter and Rao 2005). In leaf, starch accumulation is transitory in chloroplasts to generate photosynthetic precursors. The starch deposited in chloroplasts is degraded during the night and resulting G-1-P is converted to triose-P and exchanged for Pi from the cytosol (Heldt et al. 1977; Stitt et al. 1981; Dennis et al. 1982; Lee et al. 2016a, b). Pi is an activator of photosynthetic enzymes, including Rubisco (Heldt et al. 1978; Bhagwat 1981). Thus, in leaves the main AGPases are plastidial. *APS2a* and *APL3* are strongly expressed in young leaves, whereas *APS1* is expressed at low levels, *APS2b* is not detected at all, and the remaining subunits (*APL1*, *APL2* and *APL4*) are minimally expressed. Later in development, *APS2a* and *APL3* expression declines (although they are still the most abundant subunits) and *APL1* expression increases to parity with *APL3*. This suggests that the major leaf AGPase initially

comprises subunits *APS2a* and *APL3*, with *APS1* and *APL3* combining to form a less abundant subunit, but the increase in *APS1* expression may result in a progressive accumulation of the *APS1/APL3* heterotetramer later in development. Importantly, all the AGPases in the leaf appear to be located in the plastid given that no cytosolic forms are expressed at any developmental stage [*APS2b* is not detected in leaves (Ohdan et al. 2005)].

In addition to the enzymes discussed above, phosphorylase is involved in starch degradation, but earlier studies suggest it may also play a role in starch biosynthesis (Satoh et al. 2008; Steup 1990; Yu et al. 2001). There are two types of phosphorylase that differ in terms of structure, kinetics, expression patterns and subcellular location (Satoh et al. 2008). PhoI is located in plastids and appears to facilitate starch biosynthesis in storage tissues, given that it is expressed at a higher level in the endosperm than in leaves (Colleoni et al. 1999; Ball and Morell 2003; Tetlow et al. 2004; Ohdan et al. 2005) and is essential for starch synthesis and accumulation (Shimomura et al. 1982; Steup 1990; Satoh et al. 2008). In contrast, PhoH is located in the cytosol and does not play a role in starch biosynthesis. Sucrose synthase (SuSy) catalyzes the conversion of UDP and sucrose into fructose and UDP-glucose in the cytosol (Li et al. 2013). The carbon atoms in starch are derived from fructose-6-phosphate in photosynthetic tissues (Fig. 1), but from cytosolic sucrose converted to UDP-glucose in other tissues, followed by translocation to the amyloplasts (Nakamura 2005).

Based on the expression profiles of the rice AGPase subunits, we hypothesized that mutating *OsAPL2* (and thus removing the large regulatory subunit that is mainly expressed in the endosperm) should have no impact on starch biosynthesis in leaves. Interestingly, the rice *shrunken* mutant is deficient in AGPase due to the loss of subunit APS2b (the endosperm cytosolic small subunit that assembles with APL2), resulting in the loss of ~ 80% of wild-type AGPase activity in endosperm. Remarkably, the loss of *APS2b* enhances the expression of several alternative subunits in the endosperm and, in some cases, also in the leaves even though *OsAPS2b* itself is not expressed in vegetative tissues. Specifically, the expression of *OsAPL2* is strongly induced in the endosperm and at very low levels in the leaves. The plastid subunits *APS1*, *APS2a*, *APL1* and *APL3* accumulate to higher levels in the endosperm, and *APL1* and *APS1* also accumulate to higher levels in the leaves (Ohdan et al. 2005).

Other AGPase mutants have been described, but the reports focused mostly on the biochemical phenotype rather than the impact on other subunits. The rice *sugary* mutant is deficient in AGPase activity due to the loss of *OsAPS2* expression, resulting in the absence of both *APS2a* and *APS2b*. This mutant not only accumulates much less starch than wild-type plants in both the endosperm and leaves, but also features larger amyloplasts lacking visible starch granules (Kawagoe et al. 2005). The rice *apl1* mutant lacks an active *APL1* subunit due to the deletion of an essential conserved domain, and although there was no

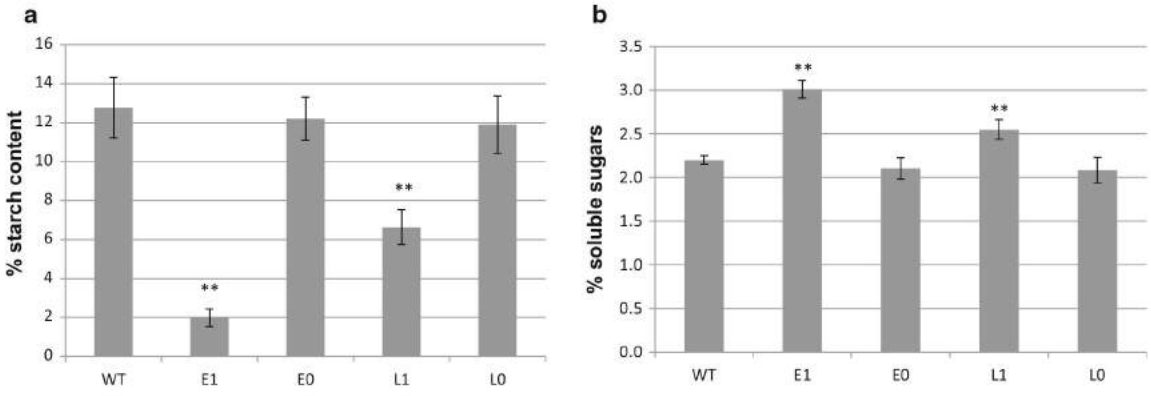


Fig. 6 Starch and soluble sugar content in flag leaves. **a** Starch and **b** soluble sugar content in mutants (E1 and L1), corresponding negative controls (E0 and L0) and wild type. Values are expressed as means ± SDs (n = 3 biological replicates, 2 technical replicates for each biological). The asterisks indicate a statistically significant difference between wild-type (WT) and mutant, as determined by a Student's *t*-test (*$P < 0.05$; **$P < 0.01$)

change in AGPase activity in the endosperm, no AGPase activity was detected in the leaves and the leaves contained < 5% of normal starch levels but normal levels of soluble sugars (Rosti et al. 2007). These data suggest that the APL1 subunit is necessary for starch synthesis in leaves but not in non-photosynthetic organs. The rice *apl3* mutant lacks an active *APL3* subunit, and AGPase activity was reduced by 67% in the embryos (resulting in 35% less starch than normal) and by 23% in the endosperm, suggesting that *APL3* makes a major contribution in the embryo rather than to the endosperm (Cook et al. 2012).

The rice *phoI* mutant is characterized by the substantial loss of starch and altered amylopectin structure. There were no significant differences in the activity of AGPase, DBE, SBE or GBSS compared to wild-type plants, suggesting that PhoI influences two steps during starch biosynthesis, i.e. starch initiation and starch elongation (Satoh et al. 2008). The overexpression of *SuSy* in potato increased ADP-glucose and starch levels compared to wild-type plants, and the mutant tubers contained 55–85% more starch (Baroja-Fernández et al. 2009). In maize overexpressing *SuSy*, there was likewise an increase in ADP-glucose levels and a 10–15% more starch in the endosperm (Li et al. 2013). However, SuSy is not the major determinant of ADP-glucose production in cereals. An alternative explanation is that SuSy is responsible for the accumulation of UDP-glucose, resulting in higher levels of downstream metabolites such as glucose-1-phosphate, ADP-glucose and starch (Howard et al. 2012).

We knocked out one allele of the *OsAPL2* in two separate experiments to obtain two different mutants. We targeted an upstream exon and produced a truncated protein with no catalytic activity (E1). In a second set of experiments we targeted an exon downstream of the active site, to maintain some activity but perturb the tertiary structure (L1). We hypothesized that knocking out *OsAPL2* would have no effect on leaves. We specifically investigated heterozygous mutations in order to determine whether the mutation caused any feedback effects on the wild-type allele to restore normal starch biosynthesis. Remarkably, we found that both mutations caused a reduction in starch levels and higher levels of soluble sugars in the leaves, even though the corresponding subunit *APS2b* is not expressed in photosynthetic tissues. Mutant E1 contained 85% less starch and 36% more sugar in the leaves than wild-type plants, whereas mutant L1 contained 50% less starch and 18% more sugar. Mutants were infertile because the lower starch levels resulted in male sterility (Lee et al. 2016a, b). Recently, Tang et al. (2016) reported similar results for the starch *w24* mutant which contained 21% less starch and 13–43% more sugar. The *w24* mutant carries a homozygous single nucleotide substitution in exon 4 of the *OsAPL2*. In the *w24* mutant Glc-6-P is converted to Glc-1-P which is then converted by AGPases to ADP-Glc in amyloplasts in pollen.

AGPase activity in E1 was reduced by 31 % correlating with the mutation of one allele, whereas AGPase activity in L1 increased by 2-fold because the mutated *APL2* allele retained some residual activity, simultaneously with an increase in the activity of the other subunits (please see below). We also measured the activity of SuSy because this enzyme is thought to provide an alternative route for starch biosynthesis, by converting sucrose and ADP directly to ADP-glucose, with fructose as a by-product (Li et al. 2013). We found that SuSy activity increased in E1 possibly as a direct compensation for the loss of AGPase activity, whereas SuSy activity in L1 was similar to wild-type plants, because L1 maintained activity in all alleles (WT and mutated). These data support a hypothesis in which Susy uses ADP which accumulates due to the lack of AGPase activity (Li et al. 2013).

Translation in the E1 mutant was terminated within exon 1. In contrast, the L1 mutant had an insertion of an adenosine residue within exon 13, causing a frameshift that generated a shorter protein with a different C-terminal domain. Sequence has shown that the C-terminal of the protein was conserved within the AGPase large subunit family and forms a loop structure located near the putative substrate and effector binding sites (Tuncel et al. 2014; Tang et al. 2016). Several *OsAPL2* point mutants in this region have been described, including substitutions T139I, A171V and L155P which replaced the original residues with bulkier side chains that altered the structure and modified the topology of the substrate and/or effector binding sites (Tuncel et al. 2014; Tang et al. 2016). The L1 mutant potentially has a similar impact by abolishing the hydrogen bond between residues Gln 98 with Arg 500 and Tyr 498 in the loop structure and forming a new hydrogen bond with Thr 351. Furthermore, 19 amino acids at the C-terminus of

the large and small subunits of potato AGPase are essential for correct folding and/or assembly into multimeric complexes (Laughlin et al. 1998). In agreement with these results, the C-termini of *OsAPL2* and *OsAPS2b* were more important for the efficient multimerization of the AGPase subunits than the corresponding N-terminal regions (Tang et al. 2016). Our tetrameric modelling simulations show clearly that the salt bridge that stabilizes the binding between two *APL2* and two *APS2b* subunits changes but the tetramer structure can still form due to a new salt bridge. L1 lacks the C-terminal β-sheets required for efficient subunit interactions. This explains the increase in subunit mRNA abundance without a corresponding increase in starch levels.

APS2b and *APL2* are the only cytosolic AGPase subunits. They are preferentially expressed in the endosperm, so mutating *OsAPL2* is not expected to impact in leaves, where neither subunit is present at significant levels and only plastidial AGPase activity is relevant. We compared the biochemical and structural parameters of AGPase in wild-type and mutant leaves. Surprisingly we observed changes in starch and soluble sugar levels, AGPase activity and SuSy activity. Expression analysis in the mutants showed that the *OsAPL2* transcript was more abundant than in wild-type plants, as was the transcript for the small subunit counterpart *OsAPS2b*. It therefore appears that in both mutants the *OsAPL2* locus is induced to compensate for the mutation of one allele, resulting in a counter-intuitive increase in expression from the wild-type allele, and a trans-acting effect on the expression of *OsAPS2b*. The expression levels of *OsAPL1* and *OsAPS1* also increased compared to wild-type plants. The expression level of *OsAPS2a* was reduced to half the level observed in wild-type plants whereas the levels of *OsAPL3* and *OsAPL4* mRNA remained similar to wild-type plants.

Increases in the expression of functional alleles to compensate for loss of expression in one allele at the same locus is not uncommon and has been reported earlier. Examples include dosage compensation to reflect the differential distribution of sex-chromosomes in animals and in a small number of plants, such as *Silene latifolia* (Muyle et al. 2012); compensation in the case of aneuploidy or polyploidy (Makarevitch and Harris 2010); compensation for monoallelic silencing as observed in paramutation and imprinting (Guidi et al. 2004); and buffering gene expression levels in the case of autosomal gene or allele copy number variation (Trieu et al. 2003). The compensatory increase in *OsAPL2* expression in L1 and E1 is caused by a mutation that changes the number of functional alleles, consistently with buffering. Little is known about inter-allelic buffering at the mRNA level in plants, but studies in yeast and insects have indicated that this may be a common mechanism that acts to stabilize the transcriptome (Lundberg et al. 2012; Bader et al. 2015). Such mechanisms may also partly explain the stability of heterozygous populations and hybrid vigor in plants (Verta et al. 2016).

Transcriptional compensation has also been observed among functionally redundant paralogous genes when one copy is mutated, e.g. among the nicotianamine synthase genes (Klatte et al. 2009). This may help to explain why the feedback mechanism that induced *OsAPL2* expression when one allele was mutated also induced the functionally related *OsAPS2*, although only the *OsAPS2b* mRNA was expressed at high levels. The major small subunit in leaves, APS2a, was significantly downregulated in E1 and L1 mutants, suggesting that the level of *OsAPS2a* mRNA may be independently suppressed at the post-transcriptional level. In turn, *OsAPS1* may be enhanced to compensate for the suppression of *OsAPS2a* in plastids. *OsAPL3* (the major large subunit isoform in leaves) and *OsAPL4* expression were similar to the wild type. The suppression of *OsAPS2a* explains why the increase in AGPase expression did not normalize starch levels in the two mutants. *OsAPS2a* is located in the plastids and the main impact of the two mutations (E1 and L1) was to induce the expression of an ectopic cytosolic form of AGPase, which is decoupled from the rest of the pathway because the latter is located within the plastids. The suppression of the major cytosolic small subunit may therefore be sufficient to limit starch synthesis. Ectopic cytosolic AGPase expression, even at high levels is unable to compensate for the loss of activity due to the absence of metabolic shuttling between the plastids and cytosol. Similar expression patterns were reported in other studies when *OsAPS2b* expression was blocked, i.e. an increase in *OsAPL2* expression, dramatically increases *OsAPL1* and *OsAPS1* expression, but no changes in *OsAPL3* and *OsAPL4* (Ohdan et al. 2005). *OsPhoI* was also induced in L1 and E1 perhaps reflecting its key role in the initiation of starch biosynthesis (Satoh et al. 2008).

An alternative explanation for the lack of compensation of starch levels even when transcription of other subunits increases may be changes in Pi levels. In E1 the increase in SuSy activity responsible for the production of Pi in the cytosol promotes the exchange of Triose-P from the plastids reducing C3 intermediates required for starch synthesis (Preiss et al. 1994). In L1 even though SuSy activity did not increase, the concentration of soluble sugars were elevated and this may prevent Pi from being recycled for photophosphorylation and thereby very likely reducing C3 intermediates required for starch biosynthesis (Flugge et al. 1999).

In summary, mutating one allele of *OsAPL2*, encoding the major endosperm large subunit of AGPase and the only subunit expressed in the cytosol, resulted in the unexpected expression of both *OsAPL2* and *OsAPS2b* in the leaves, the latter encoding the only cytosolic small subunit. However, the formation of a complete ectopic AGPase in the leaf cytosol did not increase overall starch synthesis. Instead, the leaves contained less starch than wild-type plants most likely reflecting the lower levels of plastidial *OsAPS2a*, increased SuSy activity (at least in E1), the increased in soluble sugars and/or the inability of *OsAPS1* to replace *OsAPS2a* function completely. Our findings indicate that the new cytosolic AGPase was not sufficient to compensate for the loss of plastidial AGPase, probably because there is no wider starch biosynthesis pathway in the leaf cytosol and thus no pathway intermediates are shuttled between the two compartments. The principal differences between mutants E1 and L1 reflect the impact of changes in AGPase activity: in L1 overall AGPase activity has been changed whereas the alternative SuSy pathway was activated in E1.

Acknowledgements We thank Dr. Caixia Gao (Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China) for providing pJIT163-2NLSCas9 containing the cas9 gene codon-optimized for rice, and the empty pU3-gRNA vector for the introduction of sgRNAs into rice. This work was supported by funding from the Spanish Ministry of Economy and Competitiveness (MINECO) (BIO2014-54426), and a Juan de la Cierva fellowship to GF (IJCI- 2014-19528). LP is the recipient of MINECO fellowship. ES is the recipient of a PhD fellowship from the University of Lleida (BIO2014-54441-P).

References

Bader DM, Wilkening S, Lin G, Tekkedil MM, Dietrich K, Steinmetz LM, Gagneur J (2015) Negative feedback buffers effects of regulatory variants. Mol Syst Biol 11:785. https://doi.org/10.15252/msb.20145844

Ball SG, Morell MK (2003) From bacterial glycogen to starch: understanding the biogenesis of the plant starch granule. Annu Rev Plant Bio 54:207–233

Ballicora MA, Iglesias AA, Preiss J (2003) ADP-glucose pyrophosphorylase: a regulatory enzyme for plant starch synthesis. Microbiol Mol Biol Rev 67:213–225

Baroja-Fernandez E, Muñoz FJ, Montero M et al (2009) Enhancing sucrose synthase activity in transgenic potato (*Solanum tuberosum L.*) tubers results in increased levels of starch, ADPglucose and UDPglucose and total yield. Plant Cell Physiol 50:1651–1662

Bassie L, Zhu C, Romagosa I, Christou P, Capell T (2008) Transgenic wheat plants expressing an oat arginine decarboxylase cDNA exhibit increases in polyamine content in vegetative tissue and seeds. Mol Breed 22:39–50

Baysal C, Bortesi L, Zhu C, Farré 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. Mol Breed 36:108. https://doi.org/10.1007/s11032-016-0533-4

Bhagwat AS (1981) Activation of spinach ribulose 1,5-bisphosphate carboxylase by inorganic phosphate. Plant Sci Lett 23:197–206

Bortesi L, Zhu C, Zischewski J et al (2016) Patterns of CRISPR/Cas9 activity in plants, animals and microbes. Plant Biotechnol J 14(12):2203–2216

Chari R, Mali P, Moosburner M, Church GM (2015) Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach. Nat Methods 12:823–826

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 immature zygotic embryos. Nat Biotechnol 9:957–962

Colleoni C, Dauvillèe D, Moulle G et al (1999) Genetic and biochemical evidence for the involvement of α-1,4 glucanotransferases in amylopectin synthesis. Plant Physiol 120:993–1003

Cook FR, Fahy B, Trafford K (2012) A rice mutant lacking a large subunit of ADP-glucose pyrophosphorylase has drastically reduced starch content in the culm but normal plant morphology and yield. Funct Plant Biol 39:1068–1078

Dennis DT, Miernyk JA (1982) Compartmentation of nonphotosynthetic carbohydrate metabolism. Annu Rev Plant Physiol 33:27–50

Doehlert DC, Kuo TM, Felker FC (1988) Enzymes of sucrose and hexose metabolism in developing kernels of two inbreds of maize. Plant Physiol 86:1013–1019

Doudna JA, Charpentier E (2014) The new frontier of genome engineering with CRISPR-Cas9. Science 346:1258096

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 Res 21:1093–1097

Fauser F, Schiml S, Puchta H (2014) Both CRISPR/Cas-based nucleases and nickases can be used efficiently for genome engineering in Arabidopsis thaliana. Plant J 79:348–359

Flugge UI (1999) Phosphate translocators in plastids. Annu Rev Plant Physiol Plant Mol Biol 50:27–45

Giroux MJ, Hannah L (1994) ADP-glucose pyrophosphorylase in shrunken-2 and brittle-2 mutants of maize. Mol Gen Genet 243:400–408

Guidi CJ, Veal TM, Jones SN, Imbalzano AN (2004) Transcriptional compensation for loss of an allele of the Ini1 tumor suppressor. J Biol Chem 279:4180–4185

Heigwer F, Kerr G, Boutros M (2014) E-CRISP: fast CRISPR target site identification. Nat Methods 11:122–123

Heldt HW, Chon CH, Maronde D, Herold A, Stankovic AZ, Walker DA, Kraminer A, Kirk MR, Heber U (1977) Role of orthophosphate and other factors in the regulation of starch formation in leaves and isolated chloroplasts. Plant Physiol 59:1146–1155

Heldt HW, Chon CJ, Lorimer H (1978) Phosphate requirement for the light activation of ribulose-1,5-biphosphate carboxylase in intact spinach chloroplasts. FEBS Lett 92:234–240

Howard TP, Fahy B, Craggs A et al (2012) Barley mutants with low rates of endosperm starch synthesis have low grain dormancy and high susceptibility to preharvest sprouting. New Phytol 194:158–167

Jobling S (2004) Improving starch for food and industrial applications. Curr Opin Plant Biol 7:210–218

Johnson PE, Patron NJ, Bottrill AR et al (2003) A low-starch barley mutant, risø 16, lacking the cytosolic small subunit of ADP-glucose pyrophosphorylase, reveals the importance of the cytosolic isoform and the identity of the plastidial small subunit. Plant Physiol 131:684–696

Kang TJ, Yang MS (2004) Rapid and reliable extraction of genomic DNA from various wild-type and transgenic plants. BMC Biotechnol 4:20. https://doi.org/10.1186/1472-6750-4-20

Kawagoe Y, Kubo A, Satoh H, TakaiwaF Nakamura Y (2005) Roles of isoamylase and ADP-glucose pyrophosphorylase in starch granule synthesis in rice endosperm. Plant J 42:164–174

Klatte M, Schuler M, Wirtz M, Fink-Straube C, Hell R, Bauer P (2009) The analysis of Arabidopsis nicotianamine synthase mutants reveals functions for nicotianamine in seed iron loading and iron deficiency responses. Plant Physiol 150:257–271

Laughlin MJ, Chantler SE, Okita TW (1998) N- and C- terminal peptide sequences are essential for enzyme assembly allosteric, and/or catalytic properties of ADP-glucose pyrophosphorylase. Plant J 14:159–168

Lee SK, Hwang SK, Han M, Eom JS, Kang HG, Han Y et al (2007) Identification of the ADP-glucose pyrophosphorylase isoforms essential for starch synthesis in the leaf and seed endosperm of rice (*Oryza sativa* L.). Plant Mol Biol 65:531–546

Lee J, Chung JH, Kim HM, Kim DW, Kim H (2016a) Designed nucleases for targeted genome editing. Plant Biotechnol J 14:448–462

Lee SK, Eom JS, Hwang SK, Shin D, An G, Okita TW, Jeon JS (2016b) Plastidic phosphoglucomutase and ADP-glucose pyrophosphorylase mutants impair starch synthesis in rice pollen grains and cause male sterility. J Exp Bot 67:5557–5569

Li J, Baroja-Fernández E, Bahaji A, Muñoz FJ, Ovecka M, Montero M et al (2013) Enhancing sucrose synthase activity results in an increased levels of starch and ADP-Glucose in maize (*Zea mays* L.) seed endosperms. Plant Cell Physiol 54:282–294

Lundberg LE, Figueiredo ML, Stenberg P, Larsson J (2012) Buffering and proteolysis are induced by segmental monosomy in Drosophila melanogaster. Nucleic Acids Res 40:5926–5937

Ma X, Zhang Q, Zhu Q, Liu W, Chen Y, Qiu R et al (2015) A robust CRISPR/Cas9 system for convenient, high-efficiency multiplex genome editing in monocot and dicot plants. Mol Plant 8:1274–1284. https://doi.org/10.1016/j.molp.2015.04.007

Makarevitch I, Harris C (2010) Aneuploidy causes tissue-specific qualitative changes in global gene expression patterns in maize. Plant Physiol 152:927–938

Martin C, Smith AM (1995) Starch biosynthesis. Plant Cell 7:971–985

Mikami M, Toki S, Endo M (2016) Precision targeted mutagenesis via Cas9 paired nickases in rice. Plant Cell Physiol 57:1058–1068

Müller-Röber B, Sonnewald U, Willmitzer L (1992) Inhibition of the ADP-glucose pyrophosphorylase in transgenic potatoes leads to sugar-storing tubers and influences tuber formation and expression of tuber storage protein genes. EMBO J 11:1229–1238

Muyle A, Zemp N, Deschamps C, Mousset S, Widmer A, Marais GA (2012) Rapid de novo evolution of x chromosome dosage compensation in *Silene latifolia*, a plant with young sex chromosomes. PLoS Biol 10:e1001308

Nakamura Y, Francisco PB Jr, Hosaka Y, Sato A, Sawada T, Kubo A, Fujita N (2005) Essential amino acids of starch synthase IIa differentiate amylopectin structure and starch quality between *japonica* and *indica* rice varieties. Plant Mol Biol 58:213–227

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

Ohdan T, Francisco PB 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. J Exp Bot 56:3229–3244

Pandey MK, Rani NS, Madhav MS, Sundaram RM, Varaprasad GS, Sivaranjani AK et al (2012) Different isoforms of starch-synthesizing enzymes controlling amylose and amylopectin content in rice (*Oryza sativa* L.). Biotechnol Adv 30:1697

Preiss J (1982) Regulation of the biosynthesis and degradation of starch. Ann Rev Plant Physiol 33:431–454

Preiss J (1994) Regulation of the C_3 reductive cycle and carbohydrate synthesis. In: Tolbert NE, Preiss J (eds) Regulation of atmospheric CO_2 and O_2 by photosynthetic carbon metabolism, 1st edn. Oxford University Press, New York, pp 93–102

Ran FA, Hsu PD, Lin CY, Gootenberg JS, Konermann S, Trevino AE et al (2013) Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell 154:1380–1389

Rösti S, Fahy B, Denyer K (2007) A mutant of rice lacking the leaf large subunit of ADP-glucose pyrophosphorylase has drastically reduced leaf starch content but grows normally. Funct Plant Biol 34:480–489

Rychter AM, Rao IM (2005) Role of phosphorus in photosynthetic carbon metabolism. In: Pessarakli M (ed) Handbook of photosynthesis, 2nd edn. Taylor y Francis group, Tucson, pp 123–148

Satoh H, Shibahara K, Tokunaga T, Nishi A, Tasaki M, Hwang SK et al (2008) Mutation of the plastidial alpha-glucan phosphorylase gene in rice affects the synthesis and structure of starch in the endosperm. Plant Cell 20:1833–1849

Shan Q, Wang Y, Li J, Gao C (2014) Genome editing in rice and wheat using the CRISPR/Cas system. Nat Protoc 9:2395–2410

Shimomura S, Nagai M, Fukui T (1982) Comparative glucan specificities of two types of spinach leaf phosphorylase. J Biochem 91:703–717

Steup M (1990) Starch degrading enzymes. In: Dey PM, Harborne JB (eds) Methods in plant biochemistry. Academic Press, London, pp 103–128

Stitt M, Heldt HW (1981) Physiological rates of starch breakdown in isolated intact spinach chloroplasts. Plant Physiol 68:755–761

Sudhakar D, Duc LT, Bong BB, Tinjuangjun P, Maqbool SB, Valdez M et al (1998) An efficient rice transformation system utilizing mature seed-derived explants and a portable, inexpensive particle bombardment device. Transgenic Res 7:289–294

Sun Y, Jiao G, Liu Z et al (2017) Generation of high-amylose rice through CRISPR/Cas9-mediated targeted mutagenesis of starch branching enzymes. Front Plant Sci 8:298. https://doi.org/10.3389/fpls.2017.00298

Tang XJ, Peng C, Zhang J, Cai Y, You XM, Kong F et al (2016) ADP-glucose pyrophosphorylase large subunit 2 is essential for storage substance accumulation and subunit interactions in rice endosperm. Plant Sci 249:70–83

Tester RF, Morrison WR, Schulman AH (1993) Swelling and gelatinization of cereal starches. V. Risø mutants of bomi and carlsberg II barley cultivars. J Cereal Sci 17:1–9

Tetlow IJ, Wait R, Lu Z, Akkasaeng R, Bowsher CG, Esposito S et al (2004) Protein phosphorylation in amyloplasts regulates starch branching enzyme activity and protein-protein interactions. Plant Cell 16:694–708

Trieu M, Ma A, Eng SR, Fedtsova N, Turner EE (2003) Direct autoregulation and gene dosage compensation by POU-domain transcription factor Brn3a. Development 130:111–121

Tsai CY, Nelson OE (1966) Starch-deficient maize mutant lacking adenosine diphosphate glucose pyrophosphorylase activity. Science 151:341–343

Tuncel A, Kawaguchi J, Ihara Y, Matsusaka H, Nishi A, Nakamura T et al (2014) The rice endosperm ADP-Glucose pyrophosphorylase large subunits essential for optimal catalysis and allosteric regulation of the heterotetrameric enzyme. Plant Cell Physiol 55:1169–1183

Valdez M, Cabrera-Ponce JL, Sudhakar D, Herrera-Estrella L, Christou P (1998) Transgenic central american, west african and asian elite rice varieties resulting from particle bombardment of foreign DNA into mature seed-derived explants utilizing three different bombardment devices. Annu Bot 82:795–801

Verta JP, Landry CR, MacKay J (2016) Dissection of expression-quantitative trait locus and allele specificity using a haploid/diploid plant system-insights into compensatory evolution of transcriptional regulation within populations. New Phytol 211:159–171

Yoshida S, Forno DA, Cock JH, Gomez KA (1976) Determination of sugar and starch in plant tissue, 3erd edition. Laboratory manual for physiological studies of rice. The international rice research institute, Laguna Philippines, pp 46–49

Yu Y, Mu HH, Wasserman BP, Carman GM (2001) Identification of the maize amyloplast stromal 112-kD protein as a plastidic starch phosphorylase. Plant Physiol 125:351–359

Yuan D, Bassie L, Sabalza M, Miralpeix B, Dashevskaya S, Farre G et al (2011) The potential impact of plant biotechnology on the Millennium Development Goals. Plant Cell Rep 30:249–265

Zhang D, Wu J, Zhang Y, Shi C (2012) Phenotypic and candidate gene analysis of a new floury endosperm mutant (*osagpl2-3*) in rice. Plant Mol Biol Report 30:1303–1312

Zhu C, Sanahuja G, Yuan D, Farré G, Arjó G, Berman J et al (2013) Biofortification of plants with altered antioxidant content and composition: genetic engineering strategies. Plant Biotechnol J 11:129–141

Zhu C, Bortesi L, Baysal C, Twyman RM, Fischer R, Capell T et al (2017) Characteristics of genome editing mutations in cereal crops. Trends Plant Sci 22:38–52

Plant Biotechnology Journal

Plant Biotechnology Journal (2016) 14, pp. 2203–2216 doi: 10.1111/pbi.12634

Review article

Patterns of CRISPR/Cas9 activity in plants, animals and microbes

Luisa Bortesi[1,†], Changfu Zhu[2,†], Julia Zischewski[1], Lucia Perez[2], Ludovic Bassié[2], Riad Nadi[2], Giobbe Forni[2], Sarah Boyd Lade[2], Erika Soto[2], Xin Jin[2], Vicente Medina[2], Gemma Villorbina[2], Pilar Muñoz[2], Gemma Farré[2], Rainer Fischer[1,3], Richard M. Twyman[4], Teresa Capell[2], Paul Christou[2,5] and Stefan Schillberg[3,*]

[1]Institute for Molecular Biotechnology, RWTH Aachen University, Aachen, Germany

[2]Department of Plant Production and Forestry Science, School of Agrifood and Forestry Science and Engineering (ETSEA), University of Lleida-Agrotecnio Center, Lleida, Spain

[3]Fraunhofer Institute for Molecular Biology and Applied Ecology IME, Aachen, Germany

[4]TRM Ltd, York, UK

[5]ICREA, Catalan Institute for Research and Advanced Studies, Barcelona, Spain

Received 11 July 2016;
revised 5 September 2016;
accepted 7 September 2016.
**Correspondence* (Tel +49 241 6085 11050; fax +49 241 6085 50050; email stefan.schillberg@ime.fraunhofer.de)
†These authors contributed equally to the work.

Keywords: genome editing, mutational signature, off-target mutations, on-target mutations, sgRNA design, site-directed mutagenesis, species-dependent effects.

Summary

The CRISPR/Cas9 system and related RNA-guided endonucleases can introduce double-strand breaks (DSBs) at specific sites in the genome, allowing the generation of targeted mutations in one or more genes as well as more complex genomic rearrangements. Modifications of the canonical CRISPR/Cas9 system from *Streptococcus pyogenes* and the introduction of related systems from other bacteria have increased the diversity of genomic sites that can be targeted, providing greater control over the resolution of DSBs, the targeting efficiency (frequency of on-target mutations), the targeting accuracy (likelihood of off-target mutations) and the type of mutations that are induced. Although much is now known about the principles of CRISPR/Cas9 genome editing, the likelihood of different outcomes is species-dependent and there have been few comparative studies looking at the basis of such diversity. Here we critically analyse the activity of CRISPR/Cas9 and related systems in different plant species and compare the outcomes in animals and microbes to draw broad conclusions about the design principles required for effective genome editing in different organisms. These principles will be important for the commercial development of crops, farm animals, animal disease models and novel microbial strains using CRISPR/Cas9 and other genome-editing tools.

Introduction

Clustered regularly interspaced short palindromic repeats (CRISPRs) are repetitive sequences found in bacterial and archaeal genomes interrupted by spacers captured from previously encountered virus genomes and other invasive DNA. Their function is to provide a form of adaptive immunity via CRISPR-associated (Cas) proteins that act as RNA-directed endonucleases to degrade the same type of invasive DNA if it is encountered again (Lee *et al.*, 2015). Three major CRISPR/Cas systems have been described (Kumar and Jain, 2015) although several additional systems have been reported more recently (Makarova *et al.*, 2011, 2015). In type II systems (Jinek *et al.*, 2012), fragments of invasive DNA (protospacers) approximately 20 bp in length are captured due to their proximity to a short and highly degenerate sequence known as a protospacer adjacent motif (PAM) and these fragments become the spacers in the genomic CRISPR array. Transcription of the array yields a long transcript which is processed into shorter CRISPR RNAs (crRNAs), each representing a single spacer. The crRNA forms a complex with endonuclease Cas9 and a transactivating crRNA (tracrRNA) which mediates the interaction. The Cas9 ribonucleoprotein (RNP) complex then binds to DNA containing a PAM and a protospacer matching the crRNA. Cleavage occurs three nucleotides upstream of the PAM on both strands, mediated by the Cas9 endonuclease domains RuvC and HNH, respectively, introducing a precise double-strand break (DSB) with blunt ends that causes target DNA degradation (Chen and Gao, 2014; Doudna and Charpentier, 2014; Osakabe and Osakabe, 2015).

The ability of the type II CRISPR/Cas9 system to recognize specific DNA targets has been exploited to develop an RNA-guided genome-editing platform that is more versatile than equivalent platforms involving protein-based DNA-binding modules such as zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs; Hsu *et al.*, 2014). The natural system has been converted into a universal genome-editing platform by reducing it to two convenient components (Figure 1a). The first is the Cas9 endonuclease, typically from *Streptococcus pyogenes* (*Sp*Cas9), which for eukaryotic targets is equipped with a nuclear localization signal (Belhaj *et al.*, 2013, 2015). The second is a synthetic guide RNA (sgRNA) combining the tracrRNA and crRNA functions of the natural system into a single molecule (Belhaj *et al.*, 2015; Sander and Joung, 2014). The sgRNA targets a unique 20-bp sequence in the genome of the host organism, and any sequence can be chosen as long as it is adjacent to a PAM. Codon-optimized versions of the *cas9* gene offer maximum activity in different host species (Bortesi and Fischer, 2015) although the wild-type Cas9 is also active in

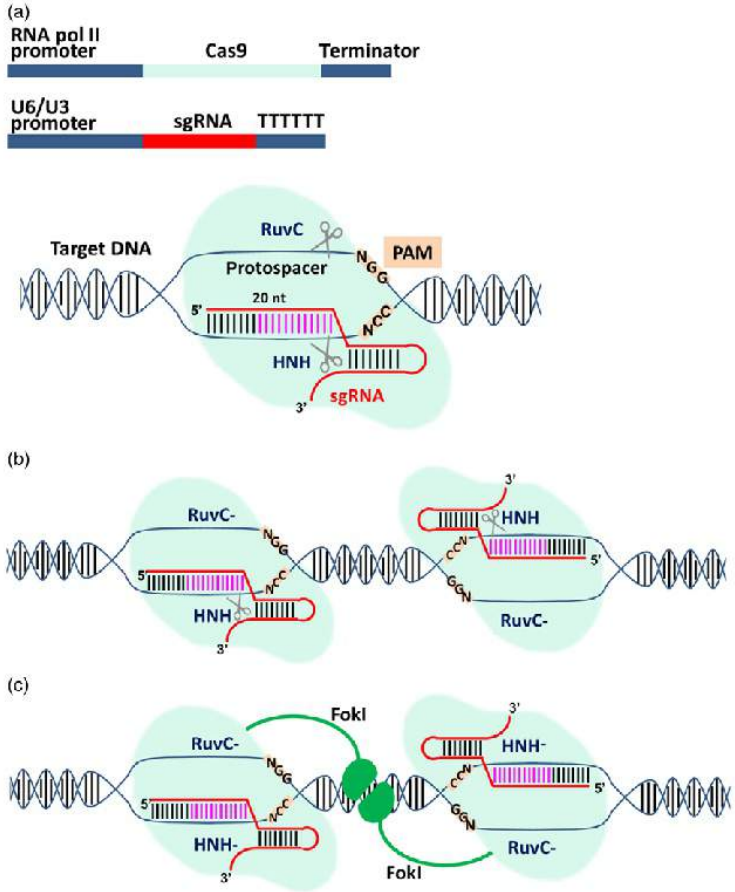


Figure 1 The engineered CRISPR/Cas9 system for genome editing. (a) Outline of the two components required for targeted cleavage: delivery of DNA constructs for transcription of the Cas9 nuclease by RNA polymerase II and the synthetic guide RNA (sgRNA) by RNA polymerase III (usually the U6 or U3 promoter) is the most common procedure, especially in plants. Alternatively, the two components can be provided as RNA or directly as a ribonucleoprotein complex (RNP; not shown). The sgRNA contains a 20-nt-long sequence complementary to the genomic target (protospacer). When the Cas9/sgRNA complex finds a matching target in the genome followed by an NGG stretch called protospacer adjacent motif (PAM), the two endonuclease domains in Cas9 (RuvC and HNH) cleave the noncomplementary and complementary strands in the target, respectively, generating a blunt double-strand break (DSB) 3-bp upstream of the PAM. The part of sgRNA proximal to the PAM (in pink) is called the seed region, and base pairing with the protospacer in this region is strictly required for recognition and cleavage of the target. Mismatches in the PAM-distal region are tolerated to some extent. (b) One endonuclease domain can be mutated (e.g. RuvC with the D10A mutation as shown in the figure or HNH with the H840A mutation, not shown) generating a Cas9 nickase. Using two sgRNAs matching adjacent genomic regions, a staggered DSB can be generated by two paired nickases. (c) If both endonuclease domains of Cas9 are mutated, the enzyme becomes catalytically inactive and is called dead Cas9 (dCas9). The dCas9 protein can still bind at its target and if fused to a nonspecific endonuclease such as FokI can generate staggered DSBs. In both (b) and (c), two precisely disposed protospacers have to be found in the genome for cleavage to occur, greatly reducing the number of possible off-target effects.

heterologous systems such as rice protoplasts (Jiang *et al.*, 2013b).

If Cas9 retains its normal catalytic activity, a blunt DSB is generated at the genomic target site as would be the case in the natural bacterial environment. In higher eukaryotes, the DSB is usually repaired by nonhomologous end joining (NHEJ), an error-prone pathway that tends to introduce small insertions and deletions, collectively known as indels (Belhaj *et al.*, 2013; Quétier, 2016). If no donor DNA is present, these indels are the only footprints of editing and they are often used to trigger frameshift mutations by targeting an exon near the 5′ end of the gene, but if donor DNA is present the same NHEJ events can facilitate the neat insertion of a DNA cassette. These events are shown in the upper left panel of Figure 2. If one of the endonuclease domains is mutated (e.g. RuvC with the D10A mutation or HNH with the H840A mutation), then Cas9 becomes a nickase, and two sgRNAs matching adjacent genomic targets can generate a staggered DSB (Jinek *et al.*, 2012; Figure 1b). This increases the accuracy of targeting because specificity relies on two target sites with a total unique sequence length of ~40 bp, and any off-target nicks are repaired by endogenous repair systems without introducing the errors

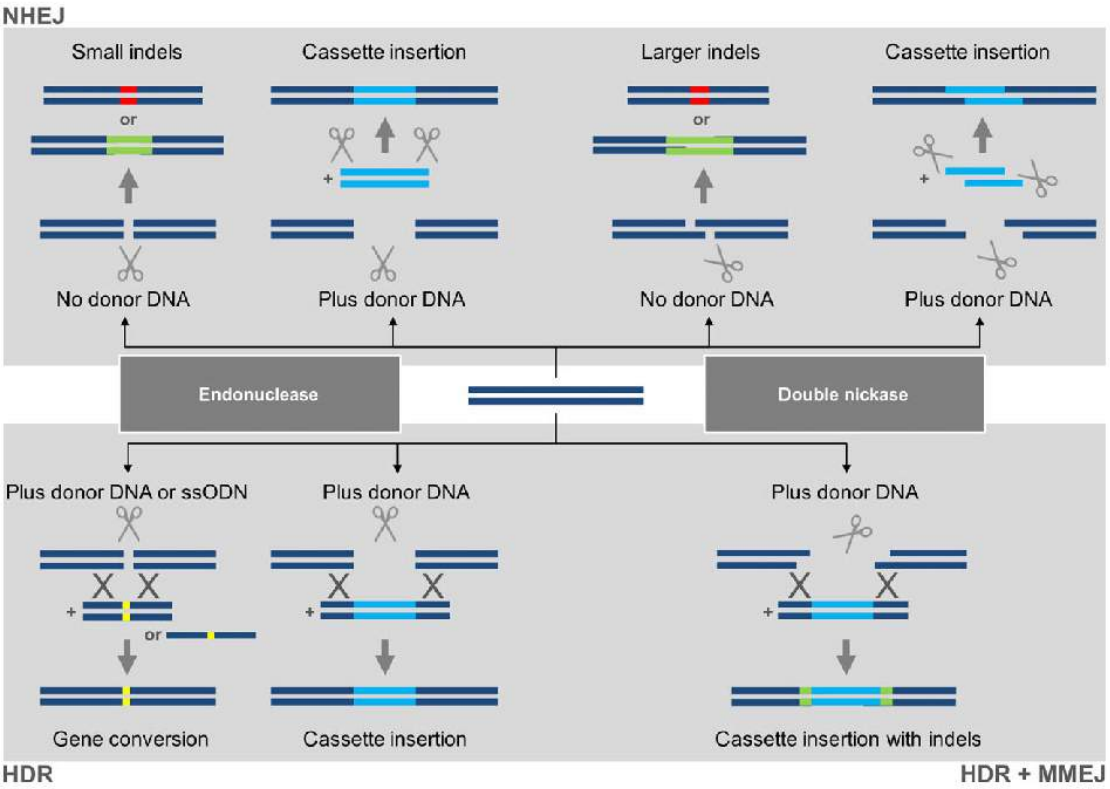


Figure 2 Genome editing with CRISPR/Cas9 can have multiple outcomes depending on the nature of the double-strand break (DSB), the prevalent repair pathway and the presence of donor DNA. The upper panel shows the major outcomes of the nonhomologous end-joining (NHEJ) pathway. In the absence of donor DNA, Cas9 endonuclease generates a blunt DSB (indicated by vertical scissors) which is repaired yielding small indels. Alternatively, the double nickase strategy generates a staggered DSB (indicated by diagonal scissors) and these tend to produce larger indels because the single-stranded tails are often involved in the repair. The indels are shown as insertions (green) or deletions (red). If donor DNA is added to the cell and is flanked by the same target sites present in the genomic locus, then compatible ends are produced which can result in a clean cassette insertion (blue). The lower panel shows the major outcomes of the homology-dependent repair (HDR) pathway if a donor DNA template is available carrying the desired modification. Donor DNA carrying a subtle change such as a nucleotide substitution (yellow) can be provided as either a duplex molecule or a single-stranded oligodeoxyribonucleotide (ssODN), and both will lead to allele replacement (gene conversion). Alternatively, the homology region may be used to flank a new sequence which will lead to cassette insertion. If the double nickase approach is used, the single strand overhangs may promote microhomology-mediated end joining (MMEJ) which can lead to imperfect cassette insertions with indels at the flanks (green).

inherent in DSB repair (Fauser *et al.*, 2014; Mikami *et al.*, 2016; Ran *et al.*, 2013). The other advantage of staggered DSBs is that donor DNA with matching sticky ends can be introduced into the cell or organism, yielding targeting events that facilitate the insertion of a donor DNA cassette (Maresca *et al.*, 2013). These events are shown in the upper right panel of Figure 2. As an alternative to dual nickases, staggered breaks can also be introduced with the *Francisella novicida* endonuclease (*Fn*Cpf1), which even in its natural state is a two-component system (a single gRNA combines the crRNA and tracrRNA functions). It also introduces the DSB (and the resulting indel) at the far end of the target site, thus preserving the original target for subsequent rounds of editing if necessary (Haeussler and Concordet, 2016; Zetsche *et al.*, 2015). If both endonuclease domains of Cas9 are mutated, the enzyme becomes catalytically inactive but can still bind at its target site. These dead Cas9 (dCas9) proteins can be used for more diverse applications including epigenetic modification, transcriptional regulation and imaging at the chromosomal level (Hsu *et al.*, 2013) or can be fused to a nonspecific endonuclease such as Fok1 as another strategy to generate staggered DSBs (Guilinger *et al.*, 2014; Tsai *et al.*, 2014; Figure 1c).

The provision of donor DNA can also be used to select for homology-dependent repair (HDR) events that are common in many microbes but occur once for every 10^5–10^6 NHEJ events in higher eukaryotes. This approach can be used either for the insertion of a donor cassette or the replacement of one allele with another, allowing the knockin of entire genes or the replacement of single nucleotides (Belhaj *et al.*, 2015). HDR is sometimes resolved cleanly, but in other cases, indels may be formed at the borders of the inserted cassette due to microhomology-mediated end joining (MMEJ). This is particularly evident when the Cas9 double nickase is used because it leaves single-stranded tails (McVey and Lee, 2008; Schiml *et al.*, 2016). These events are shown in the lower panel of Figure 2. In diploid organisms, targeted mutations can be homozygous, heterozygous or

biallelic, the latter resulting from the creation of two different mutant alleles at the target.

Relevant differences between host species

Many articles have been published describing the use of CRISPR/Cas9 in different species without considering the broader implications and species-dependent effects caused by target site preference, DSB structure and the characteristics of the target genome. The size of a genome determines the overall number of potential CRISPR/Cas9 targets because the larger the genome, the greater the number of PAMs. However, the larger the number of potential targets the greater the likelihood that some of them will be repeated, so the number of unique targets does not necessarily increase in proportion to genome size. In general terms, monocotyledonous plants (monocots) tend to have larger genomes than dicotyledonous plants (dicots), and vertebrates tend to have larger genomes than invertebrates, although there are many exceptions at the level of individual species (Gregory *et al.*, 2007; Li and Du, 2014; Michael and Jackson, 2013). Metazoans tend to have larger genomes than unicellular organisms, and eukaryotes tend to have larger genomes than bacteria and archaea (Li and Du, 2014). However, large variations in genome size within clades of eukaryotes of similar biological complexity indicate variations in the amount of repetitive DNA. Because the CRISPR/Cas9 system is typically used for the editing of genes, the exome is a more relevant comparator than the whole genome. Accordingly, mutation frequencies are generally similar in all animals and plants, suggesting that genome size does not have a significant influence on the efficiency of targeted genome editing mediated by the CRISPR/Cas9 system (Xie and Yang, 2013; Zhang *et al.*, 2016).

The GC content of the genome is known to correlate with genome size in bacteria, but in eukaryotes, the relationship is more complex due to the presence of isochores and regions containing highly repetitive DNA (Li and Du, 2014). The GC content varies greatly among different microbes but tends to fall within a relatively narrow range in animals and plants, which is higher in monocots than dicots, and higher in vertebrates than invertebrates. There is also significant variation in GC content among the chromosomes of individual animal and plant species, which is more prevalent in animals than plants, but there is no general correlation between GC content and chromosome size (Li and Du, 2014). The importance of GC content is that it has a significant impact on sgRNA efficiency (Doench *et al.*, 2014; Gagnon *et al.*, 2014; Ma *et al.*, 2015). In plants, animals and microbes, sgRNAs with a GC content greater than 50% are often reported to be more efficient (Feng *et al.*, 2014; Jiang *et al.*, 2013a,b; Pan *et al.*, 2016; Wang *et al.*, 2014a; Zhang *et al.*, 2014). By analysing sgRNAs that have been experimentally validated in plants, the GC content of most was found to lie between 30% and 80% (Liang *et al.*, 2016). Similarly, sgRNAs with an unusually high or low GC content tend to be less effective than those with an average GC content in animals, and sgRNAs targeting the transcribed strand are less effective than those targeting the nontranscribed strand (Wang *et al.*, 2014a). The same authors also found that Cas9 preferentially binds to sgRNAs containing purine residues at the last four positions of the spacer sequence and that the efficiency of cleavage is influenced by the affinity between the sgRNA and Cas9. Similar conclusions were drawn during the development of a bioinformatics tool to design sgRNAs for the effective targeting of mouse and human genes (Doench *et al.*, 2014). The analysis of sgRNA nucleotide composition in animals has revealed other nucleotide preferences (Doench *et al.*, 2014; Wang *et al.*, 2014a; Xu *et al.*, 2015a), but the same preferences have not been observed in plants (Liang *et al.*, 2016). Indeed, no statistically significant preferences in nucleotide composition were observed at any of the 20 positions of the spacer region in plants, suggesting this is a key difference in design principles for sgRNAs used in animals and plants (Liang *et al.*, 2016).

Selection of target sites in different species

The role of the PAM

The major PAM recognized by *Sp*Cas9 (5′-NGG-3′, typically denoted as NGG-PAM) occurs once every ~10 bp in a random DNA sequence and is found every 8–12 bp in the exons of all host species investigated thus far (Anders *et al.*, 2016). Nevertheless, the requirement for this PAM restricts targetable genomic sites to sequences immediately adjacent to this motif. This limitation has been overcome by random mutagenesis followed by screening for mutants with changes in PAM specificity; for example, VRQR-*Sp*Cas9 and VRER-*Sp*Cas9 recognize NGA-PAM and NGCG-PAM, respectively (Kleinstiver *et al.*, 2015). Endonucleases from other sources have also been investigated, such as KKH-*Sa*Cas9 from *Staphylococcus aureus*, which recognizes N_3RRT-PAM (Kleinstiver *et al.*, 2015), *Brevibacillus laterosporus* Cas9 (*Blat*Cas9), which recognizes N_4CND-PAM (Karvelis *et al.*, 2015), and the above-mentioned *Fn*Cpf1, which recognizes the unusually AT-rich TTN-PAM (Fagerlund *et al.*, 2015; Zetsche *et al.*, 2015).

The availability of complete genome sequences and oligonucleotide synthesis techniques allows the rapid design and synthesis of sgRNA libraries that can potentially target any gene in the genome (Wang *et al.*, 2014a). Many bioinformatics tools are available for the design of sgRNAs (Hendel *et al.*, 2015; Mohr *et al.*, 2016), and these can often highlight the presence of potential off-target sites in the genome (Varshney *et al.*, 2015). The number of target sequences identified by *in silico* genome analysis is influenced by the stringency of selection; that is, the number and position of any mismatches allowed between the sgRNA spacer and potential off-target sites. However, indels in the alignment between the target and the sgRNA are not taken into account by most online tools (Lin *et al.*, 2014). Although *Sp*Cas9 primarily recognizes NGG-PAM, it also binds to NAG-PAM with a much lower affinity, so prediction software can be used to include potential off-target sites adjacent to either of these motifs. Most tools additionally provide several sgRNA sequences for each gene because the efficiency of targeting depends on many factors, including the uptake/expression of the Cas9 protein and sgRNA, the accessibility of the target and the catalytic efficiency of the enzyme. Recently, Horlbeck *et al.* demonstrated that nucleosomes impede Cas9 binding and cleavage both *in vitro* and *in vivo* in human cells, and they and developed an algorithm to predict highly active sgRNAs taking into account the information on nucleosome occupancy (Horlbeck *et al.*, 2016). Several targets may need to be tested to get the best empirical balance between efficient on-target mutagenesis and the absence of off-target activity (Bassett *et al.*, 2015).

Selection of targets in plants

An extensive comparative analysis of potential *Sp*Cas9 target sites in plants was carried out by *in silico* prediction to determine the impact of parameters such as genome size and GC content in the

dicots Arabidopsis (*Arabidopsis thaliana*), *Medicago truncatula*, soya bean (*Glycine max*) and tomato (*Solanum lycopersicum*) and the monocots *Brachypodium distachyon*, rice (*Oryza sativa*), sorghum (*Sorghum bicolor*) and maize (*Zea mays*; Xie and Yang, 2013; Xie *et al.*, 2014). The genome sizes of the eight species ranged from 120 to 2065 Mb which is representative of most land plants, and the GC content ranged from 34% to 47%. One of the outputs of the study was the online database CRISPR-PLANT (http://www.genome.arizona.edu/crispr/). Potential 20-bp target sequences were extracted from the current genome sequences and sorted into five different categories according to their specificity and potential for off-target activity, based on the number and position of mismatches and the presence of NGG-PAMs and NAG-PAMs. In all eight species, 5–12 NGG-PAMs were identified in every 100 bp of the genome and the total number was positively correlated with the genome size. However, the number of specific targets ranged from 4 to 11 million and correlated positively with genome size in dicots but not in monocots. The number of specific targets did not correlate with the number of transcripts or with the number of NGG-PAMs. These results indicated that although larger genomes contain more PAMs and therefore more potential targets, the new targets were more likely to align with others and would therefore lack specificity. For seven of the eight species, it was possible to design specific sgRNAs to target 83%–99% of annotated transcripts (94.3% for Arabidopsis, 83.4% for *M. truncatula*, 89.5% for tomato, 96.4% for soya bean, 98.6% for *B. distachyon*, 87.3% for rice and 92.6% for sorghum) and 67.9%–96% of these transcripts contained at least 10 different targetable NGG-PAM sites. This indicated that off-target effects are unlikely to present a constraint in most plant species. The exception was maize, where only 29.5% of annotated transcripts matched a specific sgRNA. Among the eight species, maize had the largest genome, the highest GC content and the greatest number of annotated transcripts, reflecting the abundance of highly repetitive DNA and dispersed repeats. These features are shared with other cereals such as wheat and barley, and it may be challenging to develop unique target sites for the majority of genes in these species, although genome editing using CRISPR/Cas9 has been successful in all three cereals (Lawrenson *et al.*, 2015; Svitashev *et al.*, 2015; Upadhyay *et al.*, 2013). Cas9 variants, such as the abovementioned *Sp*Cas9 VQR and VRER mutants that recognize noncanonical PAMs, can broaden the range of genome editing; that is, they approximately double the number of accessible sites in rice compared to wild-type *Sp*Cas9 and may therefore facilitate the editing of more complex plant genomes (Hu *et al.*, 2016).

Selection of targets in other eukaryotes

Two independent studies have considered the creation of genome-scale sgRNA libraries for human cells, which can be considered a model for mammals and perhaps vertebrates in general due to the overall similarity in broad genomic characteristics. A library comprising 73 151 sgRNAs targeting 7114 genes (preferentially constitutive 5′ exons) and 100 nontargeting controls was designed by Wang *et al.* (2014a), and the sgRNAs were filtered for potential off-target effects based on sequence similarity, the presence of at least two mismatches compared to any off-target site, a GC content of 20%–80% (to reduce the likelihood of secondary structures) and fewer than four consecutive identical nucleotides. The library included 10 sgRNAs for each of 7033 protein-coding genes as well as all possible sgRNAs for each of the 84 genes encoding ribosomal proteins. Shalem *et al.* (2014) tested the feasibility of genome-scale CRISPR/Cas9 knockout screening by designing and testing a library of 64 751 sgRNAs targeting the constitutive 5′ exons of 18 080 genes with an average coverage of 3–4 sgRNAs per gene. Each target site was selected with a refined heuristic approach to minimize off-target activity. A zebrafish sgRNA library was developed based on 18 367 469 target sequences predicted in the reference genome, allowing the rapid selection of sgRNAs with embedded information concerning the predicted off-target sites for each 12-bp seed region (Varshney *et al.*, 2015). Each 20-bp target site adjacent to an NGG-PAM site differs by at least three mismatches from any other target sequence adjacent to an NGG-PAM or NAG-PAM. A genome-wide sgRNA library for Drosophila (*Drosophila melanogaster*) has been prepared containing 40 279 sgRNAs targeting 13 501 genes (78% of all genes) including 8989 targeted by three or more independent sgRNAs (Bassett *et al.*, 2015) and the *Saccharomyces cerevisiae* genome contains 645 392 specific targets (unique seed region followed by a NGG-PAM) and 108 493 less specific targets (DiCarlo *et al.*, 2013).

Comparison of targeting parameters

Resolution and targeting efficiency in plants

The resolution of gene editing refers to the nature of the repair pathway (NHEJ, MMEJ and/or HDR) and the broad definition of the resulting mutation (insertion, deletion, replacement, inversion translocation) as shown in Figure 2. NHEJ is the preferred DNA repair pathway in somatic plant cells (Puchta, 2005); therefore, most CRISPR/Cas9 events are resolved by this mechanism, resulting in error-prone repair and the introduction of indels. The reported efficiency of indels induced by CRISPR/Cas9 in both dicots and monocots can vary significantly even within the same species (e.g. 1.1%–90.4% in Arabidopsis), but most species offer examples of efficiency approaching 100%, including Arabidopsis (Yan *et al.*, 2015), soya bean (Cai *et al.*, 2015), potato (Wang *et al.*, 2015a), tomato (Pan *et al.*, 2016), petunia (Zhang *et al.*, 2016), tobacco (Gao *et al.*, 2015), poplar (Zhou *et al.*, 2015), grapefruit (Jia *et al.*, 2016), maize (Svitashev *et al.*, 2015) and rice (Ma *et al.*, 2015). Notably, the CRISPR/Cas9 system can efficiently induce mutations in different tissues and cell types, including embryogenic callus (e.g. rice and maize), hairy roots (e.g. soya bean), protoplasts (e.g. *Nicotiana benthamiana*, lettuce, maize and rice), cotyledons (e.g. tomato) and leaves (e.g. *N. benthamiana*, petunia and poplar). Studies reporting the efficiency of different types of targeting events in plants are summarized in Table S1. This intraspecific variability partly reflects how our knowledge of the CRISPR/Cas9 system has increased since the system was first used for genome editing, resulting in more efficient experimental designs. Although direct comparisons between experiments carried out under different conditions are not possible, two major common principles have emerged.

The first principle is that the efficiency of genome editing is strongly influenced by the expression of the components. In Arabidopsis, initial experiments based on *in planta* transformation and Cas9 expression controlled by the constitutive *Cauliflower mosaic virus* (CaMV) 35S promoter resulted in low editing efficiencies and mostly somatic mutations that were not transmitted to the progeny. But the frequency of heritable mutations increased to 90.4% when the constitutive promoter was replaced with an egg-cell-specific promoter (Wang *et al.*, 2015b), a cell-division-specific promoter (Hyun *et al.*, 2015; Yan *et al.*, 2015) or a germ-line-specific promoter (Mao *et al.*, 2016). In all other plant

systems, constitutive promoters have achieved high mutation frequencies, with biallelic and homozygous mutants readily produced in the first generation of monocots such as maize (Svitashev *et al.*, 2015) and rice (Lowder *et al.*, 2015; Ma *et al.*, 2015; Miao *et al.*, 2013; Wang *et al.*, 2016) and dicots such as tomato (Brooks *et al.*, 2014; Pan *et al.*, 2016), poplar (Fan *et al.*, 2015; Zhou *et al.*, 2015), potato (Wang *et al.*, 2015a), petunia (Zhang *et al.*, 2016) and tobacco (Gao *et al.*, 2015). These observations also suggest that mutagenesis often occurs at an early stage during the transformation process, before the first cell division. High levels of sgRNA limit the efficiency of genome editing at least in tomato (Pan *et al.*, 2016) and Arabidopsis (Ma *et al.*, 2015). In comparison with protoplasts, the efficiency of targeted chromosomal fragment deletion between paired sgRNA/Cas9 sites is lower in transgenic plants. This may reflect the relatively low levels of sgRNAs and Cas9 in callus tissue and regenerated plants (Xie *et al.*, 2015).

The second principle is that the nature of the sgRNA is also an important determinant of targeting efficiency. Although the rules are not completely understood, there is little doubt that some sgRNAs are more mutagenic than others and this is a key factor in determining the outcome of each editing experiment. As stated above, sgRNAs with a GC content greater than 50% often have a high editing efficiency, possibly because of a stronger binding to their target site (Feng *et al.*, 2014; Jiang *et al.*, 2013a; Pan *et al.*, 2016; Wang *et al.*, 2014a; Zhang *et al.*, 2014). In addition to the GC content of the spacer, the secondary structures of sgRNAs also affect the efficiency of editing (Makarova *et al.*, 2011). The formation of a stem-loop structure in the protospacer region can inhibit the binding of the sgRNA to the target strand, reducing the likelihood of a DSB (Ma *et al.*, 2015). Targeting one gene with multiple sgRNAs has been shown to greatly increase the mutation frequency and the recovery of homozygous mutants in rice (Wang *et al.*, 2016; Xie *et al.*, 2015; Zhang *et al.*, 2014) and T0 tomato plants (Brooks *et al.*, 2014). An extension of the culture period increased the proportion of mutated cells in *Agrobacterium tumefaciens*-infected rice callus (Mikami *et al.*, 2015a,b) and in soya bean somatic embryo cultures (Jacobs *et al.*, 2015), probably reflecting the proliferation of existing mutant cells as well as new mutations. However, this method can also reduce the regeneration capacity of the cells and increase the risk of obtaining chimeric plants (Xu *et al.*, 2015b).

The efficiency of genome editing by HDR is generally lower than NHEJ because homologous recombination occurs 10^5–10^6 times less frequently than repair by end ligation in plants (Figure 2). Only a handful of reports describe successful genome editing by HDR in higher plants, but they represent an interesting variety of different approaches. Gene conversion was achieved with an efficiency of 5% using short single-stranded oligodeoxyribonucleotides (ssODNs) as the repair template in Arabidopsis (Sauer *et al.*, 2016) and with an efficiency of 9% using dsDNA as the repair template in *N. benthamiana* (Li *et al.*, 2013). When directly compared in maize, a short ssODN template (127 nt) was twice as efficient as a plasmid donor (800 bp), achieving mutation frequencies of 0.4% and 0.2%, respectively (Svitashev *et al.*, 2015). Expression cassettes flanked by 1-kb homology arms were inserted with a frequency of 4.6% in soya bean (Li *et al.*, 2015) and 4% in maize (Svitashev *et al.*, 2015). Gene replacement using a combinatorial dual-sgRNA/Cas9 vector to remove 255 bp of endogenous sequences and insert a ~1.9-kb cassette with homology arms of 733 and 825 bp was achieved with a frequency of 0.8% in Arabidopsis (Zhao *et al.*, 2016). In one of the first examples of CRISPR/Cas9 genome editing, Li *et al.* (2013) failed to induce HDR in Arabidopsis protoplasts and attributed this to the intrinsically low efficiency of HDR in these cells compared to *N. benthamiana*, but more recent data suggest that the low efficiency probably reflected the relatively small number of DSBs, given that a 1.1%–5.6% mutation frequency was achieved by NHEJ with the same sgRNAs. An outstanding efficiency of 100% HDR-mediated conversion of the *ALS* gene was achieved in rice by Sun *et al.* (2016) using two sgRNAs for cleavage, flanking the homology arms on the donor with CRISPR sites to release the repair template *in vivo* and increasing the amount of donor DNA by introducing both the vector donor and free donor fragments (476 bp). This is by far the highest HDR frequency observed in higher eukaryotes but has yet to be tested in other species, and it will be necessary to determine whether undesirable random integration events also occur in this system. Particle bombardment appears to be up to fivefold more effective than *Agrobacterium*-mediated transformation for the promotion of HDR induced by CRISPR/Cas9 in maize (Svitashev *et al.*, 2015) and rice (Sun *et al.*, 2016). Lower levels of Cas9 but higher levels of sgRNA and repair template can increase the likelihood of resolution by HDR in yeast (Stovicek *et al.*, 2015), and this has been achieved by particle bombardment in rice (Sun *et al.*, 2016) and viral replicons in tobacco (Baltes *et al.*, 2014). Interference with the NHEJ pathway can also promote HDR as demonstrated in rice using a *lig4* mutant background to abolish end-joining ligase activity, although the impairment of NHEJ may also increase the frequency of spontaneous mutations (Endo *et al.*, 2016).

Resolution and targeting efficiency in animals

The CRISPR/Cas9 system has been used successfully in many animals, including invertebrates and vertebrates (Table S2). Unlike plants, the CRISPR/Cas9 system can be used for genome editing in particular tissues by hydrodynamic injection or by introducing the components using *Adeno-associated virus* (AAV) or *Adenovirus* vectors (Rodriguez *et al.*, 2014; Senis *et al.*, 2014; Swiech *et al.*, 2015; Xue *et al.*, 2014). Although there are differences in DNA repair pathways across species, the general preponderance of NHEJ over HDR observed in plants is also observed in animals (Figure 3). Indels generated by NHEJ have therefore been introduced with an efficiency of up to ~90% in the nematode *Caenorhabditis elegans* (Friedland *et al.*, 2013), Drosophila (Bassett *et al.*, 2015), rabbit (Lv *et al.*, 2016), chicken (Oishi *et al.*, 2016), mouse (Yang *et al.*, 2014) and human cells (Liang *et al.*, 2015), whereas the insertion of donor DNA by HDR has been reported with a frequency of ~5%–20% in *C. elegans* (Dickinson *et al.*, 2013), rat (Shao *et al.*, 2014) and mouse (Platt *et al.*, 2014). Site-specific indels were induced in zebrafish embryos by the *in vivo* microinjection of a sgRNA/Cas9 complex incorporating an additional tracrRNA sequence, causing mutagenesis at two sites that were impossible to edit with TALENs (Hwang *et al.*, 2013). In mouse, the use of ssODNs instead of a plasmid donor increased the efficiency of HDR from 10%–30% to 10%–80% (Yang *et al.*, 2014).

In human cell lines, reported mutation frequencies are often significantly lower than 5%, especially when induced pluripotent stem cells are used as the host (Miyaoka *et al.*, 2016; Yang *et al.*, 2013; Zhu *et al.*, 2015). However, much higher mutation rates of ~60–90% have also been reported in these cells, which may reflect differences in the activity of specific sgRNAs (Liang *et al.*, 2015; Veres *et al.*, 2014). Gene knockin by HDR in human cells is

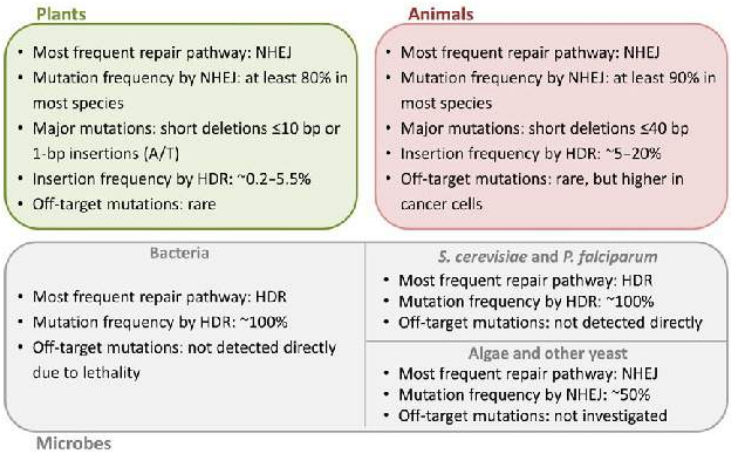


Figure 3 The outcome of genome editing with CRISPR/Cas9 is subject to species-dependent effects determined by the prevalent DNA repair pathways. Nonhomologous end joining (NHEJ) is prevalent in plants and animals, but the resulting indels tend to be smaller in plants than animals, and 1-bp insertions of A/T pairs are moderately frequent in plants but unusual in animals. Animal cells are also more efficient at HDR than plant cells although the frequency depends on the species and cell type. The range of HDR-based insertion frequencies in plants represents a broad analysis of eight articles reporting such data, but exceptional HDR frequencies as high as 9% (Li *et al.*, 2013) and 100% (Sun *et al.*, 2016) have been reported, the latter achieved by including additional free donor fragments. In contrast to animals and plants, NHEJ is much less prevalent in microbes generally, but particularly in bacteria and certain eukaryotes (including the yeast *Saccharomyces cerevisiae* and the malaria parasite *Plasmodium falciparum*). In these species, NHEJ products are so rare that CRISPR/Cas9 without a donor template is often lethal and can be used to select for HDR events without using marker genes. In other yeast and in algae, the NHEJ pathway is prevalent and the behaviour is likely to be similar to that observed in animals and plants. These principles were derived by the authors from the data in Tables S1–S3.

usually successful in fewer than 10% of cells, although this can be increased to 50%–66% if NHEJ is suppressed (Chu *et al.*, 2015). The transfection of human stem cells expressing a doxycycline-inducible Cas9 gene (iCas9) with sgRNAs enabled the drug-free selection of precise HDR-mediated modifications with ssODN donors (Zhu *et al.*, 2015). The efficiency of HDR also increased sevenfold when short hairpin RNAs were introduced expressing the *Adenovirus* Ad4 protein responsible for suppressing Ligase 4, which is required for NHEJ in mice and humans (Chu *et al.*, 2015). Biallelic double-gene mutants were generated by transfecting cells with three separate plasmids and injecting ssODNs matching two of the target genes to achieve HDR-mediated mutation (Wang *et al.*, 2013). Mice were engineered to constitutively express Cas9, and when injected with a combination of sgRNAs and HDR donors carried by AAV vectors, NHEJ and HDR events occurred at frequencies that increased over time (Platt *et al.*, 2014).

Resolution and targeting efficiency in microbes

The CRISPR/Cas9 system is a double-edged sword in bacteria because DSBs are inefficiently repaired by NHEJ, thus targeting the bacterial genome with CRISPR/Cas9 tends to be lethal (Selle and Barrangou, 2015). However, most bacterial strains are either recombinogenic or have established functional recombineering systems, so the provision of donor DNA allows the CRISPR/Cas9 system to be used as a powerful tool for the selection of HDR events and the simultaneous counterselection of background cells, resulting in highly efficient HDR without the need for positive and negative selectable markers. Using this strategy, the wild-type *Sp*Cas9 system has been applied in both Gram-positive and Gram-negative bacteria with an HDR efficiency of 65%–100% (Table S3). In *Clostridium cellulolyticum*, the Cas9 nickase variant together with a repair template achieved HDR efficiencies greater than 95% (Xu *et al.*, 2015c).

Eukaryotic microbes show diverse behaviours in response to CRISPR/Cas9, and the system must be tailored for use in different species. In contrast to higher plants, the constitutive expression of Cas9 is toxic in *Chlamydomonas reinhardtii*, preventing the recovery of transformants even in the absence of sgRNA. The basis of this phenomenon is unclear, but the transient expression of Cas9 and sgRNA has proven sufficient to generate indels by NHEJ (Jiang *et al.*, 2014) as has the delivery of Cas9 RNPs (Shin *et al.*, 2016). In contrast, Cas9 is not toxic in the marine diatom *Phaeodactylum tricornutum* and mutation frequencies of up to 63% have been achieved by NHEJ following stable transformation with a *cas9* transgene (Nymark *et al.*, 2016). Similarly, different yeast species show diverse behaviours in response to genome editing. *S. cerevisiae* is unusually amenable to HDR, so gene editing at the single nucleotide level can be achieved using short donor templates (Bao *et al.*, 2015; Biot-Pelletier and Martin, 2016; DiCarlo *et al.*, 2013) and the simultaneous deletion and insertion of several genes has been reported at frequencies of up to 100% (Stovicek *et al.*, 2015; Tsai *et al.*, 2015a). Although targeted nicks and DSBs increase the HDR frequency by up to 4000-fold in *S. cerevisiae*, only a 10-fold increase was observed in *Pichia pastoris*, where the NHEJ pathway is more prevalent (Weninger *et al.*, 2016). The malaria parasite *Plasmodium falciparum* is an interesting target because it appears to be naturally deficient in the canonical NHEJ pathway, and only HDR occurs when a donor template is provided (Ghorbal *et al.*, 2014). Genome editing using the CRISPR/Cas9 systems therefore relies on DSB-induced HDR with an external donor template, and this has been achieved with an efficiency of 50%–100% when using ssODNs.

Mutation signatures

The mutations induced by CRISPR/Cas9 in plants are mainly short deletions of 10 bp or less and single-base insertions, typically A/T in all species (Figure 3). Single-base substitutions are rare, with the exception of soya bean protoplasts where they were the most frequent mutation (Sun *et al.*, 2015). Less frequent longer deletions may represent the results of MMEJ, indicating that gene-specific factors can influence the outcome of DSB repair (Xu *et al.*, 2015b). In rice, mutation signatures vary according to the target (Miao *et al.*, 2013; Xu *et al.*, 2015b; Zhou *et al.*, 2014). The consistent mutations observed at the same target in several independent soya bean hairy-root cultures and somatic embryos also suggest that there may be as yet undiscovered rules governing the types of mutations that are favoured at a given target (Jacobs *et al.*, 2015). In the same study, the seven most effective sgRNAs exclusively generated short deletions, whereas those with lower efficiency were associated with more insertions and substitutions (Jacobs *et al.*, 2015). Interestingly, all the off-target mutations found in rice by Li *et al.* (2016) were 1-bp insertions, indicating that the pairing of sgRNA with the target sequence may also influence the mutation type.

The picture emerging from studies in animal systems is similar to that in plants: most on-target and off-target mutations reported in animals are short deletions of up to 40 bp (An *et al.*, 2016; Friedland *et al.*, 2013; Kim *et al.*, 2015). Although less common, deletions as large as 250 bp have been reported in mice (Heckl *et al.*, 2014) and human cells (Liang *et al.*, 2015), but insertions are usually shorter, typically 1–15 bp (Cheng *et al.*, 2014; Friedland *et al.*, 2013; Kim *et al.*, 2015; Liang *et al.*, 2015). CRISPR/Cas9 mutation signatures are not widely reported in microbes due to the lethality of the NHEJ pathway in bacteria and some microbial eukaryotes, and the relatively small number of studies.

The Cas9 double nickase generates mutations with different signatures compared to the intact enzyme because of the nature of the staggered DSB. For example, a Cas9 double nickase was used to generate a DSB with 52-nt overhangs in Arabidopsis. The average size of the resulting insertions was considerably larger (>80 nt) than for the fully functional Cas9 nuclease, and in most cases the insertions were copies of the sequence immediately upstream or downstream from the insertion site, which is indicative of MMEJ (Schiml *et al.*, 2014, 2016). The double nickase approach also generated longer deletions in mice (39–56 bp) as well as insertions of up to 67 bp (Cheng *et al.*, 2014). Interestingly, the two alternative Cas9 variants from *S. thermophilus* and *S. aureus* generate different mutation signatures associated with different PAMs in Arabidopsis. For *St*Cas9, sgRNAs with a NNAGAA-PAM generated 8.6% insertions (mostly 1-bp) and 3.8% deletions, whereas those with a NNGGAA-PAM generated 11.6% insertions (mostly 1-bp) and 3.8% deletions. For *Sa*Cas9, sgRNAs with a NNGAA-PAM generated 52.1% insertions (mostly 1-bp), whereas those with a NNGGGT-PAM generated 46.7% deletions and 21.6% insertions, both of which were generally larger than 1 bp (Steinert *et al.*, 2015).

Off-target mutations and methods to increase efficacy and accuracy

Off-target mutations in plants

Off-target activity is generally rare in higher plants (Table S1). Where reported, it tends to involve a minority of sgRNAs even when mutations are investigated by the thorough method of whole-genome sequencing (Feng *et al.*, 2014). A low frequency of unwanted mutations has been reported when the sgRNA features mismatches outside the seed sequence, for example in rice (Xu *et al.*, 2015b; Zhang *et al.*, 2014) and wheat (Upadhyay *et al.*, 2013), indicating that such events could be avoided by designing more specific sgRNAs. Careful sgRNA design can ensure specific targeting even when the genome contains closely related paralogous genes (Baysal *et al.*, 2016). However, in a small number of cases, unexpected cleavage has been observed at sites with one or more mismatches within the seed region, for example in Arabidopsis (Sauer *et al.*, 2016), barley (Lawrenson *et al.*, 2015), soya bean (Jacobs *et al.*, 2015) and rice (Xie and Yang, 2013). Target sequences with a GC content higher than 70% may increase the likelihood of off-target effects (Tsai *et al.*, 2015b), which might explain the unexpected mutations observed by Li *et al.* (2016) (GC = 65%–80%) but not those reported by Jacobs *et al.* (2015) (GC = 57%) or Sauer *et al.* (2016) (GC = 50%). Even so, the frequency of off-target mutations is much lower than that of on-target mutations, allowing the recovery of solely on-target mutations in all experiments. Interestingly, Xu *et al.* (2015b) detected off-target mutations only in T1 rice plants carrying the *cas9* and *sgRNA* transgenes, but not in those where the CRISPR components had segregated, suggesting that off-target effects might be reduced or avoided by selecting appropriate T1 progeny. The frequency of unwanted mutations depends on the abundance of the Cas9/sgRNA RNP complex so the likelihood can be reduced by transient expression of the components rather than stable transgene integration, although this could reduce on-target efficiency too (Tsai *et al.*, 2015b). The Cas9 double nickase resulted in efficient genome engineering in Arabidopsis, without off-target effects in homologous genomic regions (Fauser *et al.*, 2014). On the whole, off-target mutations in plants are generally less frequent than the somatic mutations that arise during tissue culture (Li *et al.*, 2016).

Off-target mutations in animals and microbes

Off-target mutations in human cells were initially reported to be up to 50% more common than mutations at the on-target site, raising concerns about the intrinsic fidelity of the CRISPR/Cas9 system (Fu *et al.*, 2013; Hsu *et al.*, 2013; Kim *et al.*, 2015; Mali *et al.*, 2013; Pattanayak *et al.*, 2013). However, those experiments were conducted on cancer cell lines, which are often characterized by dysfunctional DNA repair mechanisms, or used sgRNAs that are known to be promiscuous (Kim *et al.*, 2015). When stem cells were used as the host, whole-genome sequencing revealed the absence of off-target mutations (Smith *et al.*, 2014) or only a few off-target events (Veres *et al.*, 2014). Similarly, minimal off-target activity has been reported in zebrafish (Hruscha *et al.*, 2013), mice (Heckl *et al.*, 2014), chicken (Oishi *et al.*, 2016) and rabbit (Lv *et al.*, 2016). These studies are summarized in Table S2. However, for both animals and plants, most studies have sought off-target activity at preselected sites rather than by unbiased whole-genome sequencing, which means that off-target activity cannot be ruled out at unpredicted sites. Off-target mutations have not been directly observed in microbes but may be inferred due to their indirect toxicity (Table S3).

Methods to increase targeting efficiency and accuracy

Several approaches have been described to increase the efficiency and accuracy of CRISPR/Cas9 in plants, animals and microbes regardless of the host species and tissue, based on sgRNA design, nuclease choice and the delivery strategy. The ideal sgRNA should maximize on-target activity while minimizing off-target activity. In addition to following the general principles of sgRNA design discussed above, fidelity can be improved using a truncated spacer (<20 nt) and by adding two guanidine residues to the 5′ end of the sequence (Cho *et al.*, 2014; Fu *et al.*, 2013; Kim *et al.*, 2015). The choice of nuclease can also increase targeting accuracy. Cas9 nickases must act as dimers and therefore double the length of the target site, reducing off-target activity by 50- to 1500-fold (Fauser *et al.*, 2014; Ran *et al.*, 2013; Schiml *et al.*, 2014). This is also true for the hybrid endonuclease dCas9-FokI which is generated by fusing the enzymatically inactive dCas9 to the nonspecific endonuclease domain of the restriction enzyme FokI (Guilinger *et al.*, 2014; Tsai *et al.*, 2014). The propensity of Cas9 to tolerate mismatches between the sgRNA and target site is attributed to a high binding energy, so the rational engineering of amino acid residues involved in DNA binding has produced two high-fidelity mismatch-sensitive variants of Cas9 that achieved promising results in human cells (Kleinstiver *et al.*, 2016; Slaymaker *et al.*, 2016). The delivery of the sgRNA, nuclease and (where appropriate) the HDR donor molecule also affects the outcome of the experiment. Typically, the components are delivered as plasmid DNA and must be expressed in the cell, but they can also be introduced as RNA or as preformed RNP complexes that can be delivered efficiently by electroporation or transfection to mammalian cells (Burger *et al.*, 2016; Chu *et al.*, 2016; Liang *et al.*, 2015), plant protoplasts (Subburaj *et al.*, 2016; Woo *et al.*, 2015) or microalgae (Shin *et al.*, 2016). Because the RNP complexes are cleared rapidly, they are less likely to cleave at off-target sites (Liang *et al.*, 2015). For HDR, the use of short ssODN donor molecules seems to result in higher insertion frequencies than donor delivery by plasmid, but the size is limited to ~200 nt (Sauer *et al.*, 2016; Zhao *et al.*, 2016). Like Cas9, endonucleases from the Argonaute family have recently been found also to use oligonucleotide guides to target invasive genomes (Gao *et al.*, 2016). The DNA-guided nuclease NgAgo binds a 5′-phosphorylated, single-stranded, ~24-nt guide DNA (gDNA) that creates site-specific DSBs without needing a PAM. Preliminary characterization suggests a low tolerance of gDNA/target mismatches and highly efficient editing of GC-rich genomic targets. However, the efficiency of the system and reproducibility of the results obtained with NgAgo are being questioned (Cyranoski, 2016), so it remains to be seen whether it can be effectively used for genome editing.

More ambitious genome editing

Multiplex targeting

One of the features of the CRISPR/Cas9 system that sets it apart from ZFNs and TALENs is that multiple sgRNAs can be introduced simultaneously into a cell with little additional effort, allowing more ambitious genome-editing strategies such as the simultaneous mutation of different genes and the creation of more extensive mutations. Multiple genes can be targeted with one sgRNA by deliberately designing a promiscuous sequence, and this has been used to mutate multiple targets in related rice genes due to the tolerance of mismatches (Endo *et al.*, 2015). Multiple sgRNAs can be introduced either as separate constructs or in the form of a polycistronic cassette, which allows any number of targets to be edited simultaneously regardless of their relationship (Xie *et al.*, 2015). The latter strategy has been used extensively in rice, for example to generate null alleles at the *SlAGO7* locus (Brooks *et al.*, 2014), to mutate multiple genes at the *YSA* locus (Lowder *et al.*, 2015), to simultaneously mutate multiple genes in the *MPK* family (Xie *et al.*, 2015), to investigate the frequency of mutations in the *ERF922* gene (Wang *et al.*, 2016) and to mutate the *GSTU* and *MRP15* genes in rice and Arabidopsis (Ma *et al.*, 2015). Gao *et al.* (2015) simultaneously mutated the tobacco *PDS* and *PDR6* genes which generate an albino phenotype. Multiple genes or multiple targets within genes have also been mutated in animals. For example, Dickinson *et al.* (2013) simultaneously mutated four targets in the *C. elegans lin-31* gene to modify the MAP kinase phosphorylation sites at the C-terminus. Platt *et al.* (2014) simultaneously mutated the mouse *TP53*, *LKB1* and *KRAS* genes, which are the three most prevalent oncogenes in the lung, and a similar approach was used to study functional redundancy within the *TET* gene family (Wang *et al.*, 2013).

Large-scale mutations

Multiple sgRNAs not only allow simultaneous targeting at different sites but they can also be combined to induce large deletions and other rearrangements. For example, chromosomal segments of up to 245 kb have been deleted in rice plants by the introduction of two sgRNAs that create DSBs on the same chromosome (Zhou *et al.*, 2014), and similar strategies have been used to achieve deletions ranging from 65 kb to 30 Mb in mammalian cells (Essletzbichler *et al.*, 2014; Zhang *et al.*, 2015). Whereas most tandem DSBs are resolved by deleting the intervening DNA, another possible outcome is the creation of a chromosomal inversion. Li *et al.* (2015) reported that precise inversions of DNA fragments ranging in size from ~50 bp to hundreds of kb could be generated efficiently in mice and human cells, as well as deletions and duplications resulting from transallelic recombination between DSBs on sister chromatids. Similarly, large-scale rearrangements have been generated in human cell lines to model the chromosomal hallmarks of cancer. For example, Choi and Meyerson (2014) used pairs of sgRNAs to introduce paracentric and pericentric chromosomal inversions as well as the *CD74-ROS1* chromosomal translocation event often seen in lung cancer, and Maddalo *et al.* (2014) used a similar strategy to induce the *Eml4-Alk* inversion which is a hallmark of non-small-cell lung carcinoma. Chromosomal translocations resembling those associated with acute myeloid leukaemia and Ewing's sarcoma were induced at a high frequency using pairs of sgRNAs targeting different chromosomes by Torres *et al.* (2014). In contrast, translocations have not yet been reported in plants, and chromosomal inversions are rare: Liang *et al.* (2016) found that only one of nine expression constructs producing sgRNA pairs was able to generate an inversion and such events occurred in only two of the 23 transgenic rice plants expressing this sgRNA pair. Interestingly, Li *et al.* (2015) reported that most chromosomal inversions in their human and mouse cell lines were accompanied by small terminal indels, suggesting that repair was promoted by MMEJ. The difference in the frequency of inversions and other chromosomal rearrangements between

mammals and plants may therefore reflect the relative activity of this repair pathway.

Editing targets currently in commercial development

In addition to proof-of-concept and optimization studies in model systems, there is now great interest in the commercial applications of CRISPR/Cas9, particularly in the pharmaceutical industry where it can be used to develop accurate disease models and platforms for drug screening (Jang and Ye, 2016; Tschaharganeh *et al.*, 2016), and in the agricultural industry where it can be used to produce new crop and farm animal varieties with enhanced traits (Sovová *et al.*, 2016). The need to characterize the outcome of genome editing in detail as a way to facilitate commercial development is therefore clear, not only to streamline the development of innovative products but also in the light of the different regulatory pathways for genome-edited crops in the USA (based on product, considered an output of plant breeding) and the EU (based on process, considered GMO technology at the moment but still under debate; Sprink *et al.*, 2016). In the field of plant biotechnology, research is now focusing on genome editing in a broad range of crop species including citrus fruits, maize, poplar, potato, rice, sorghum, soya bean, tomato and wheat (Khatodia *et al.*, 2016; Song *et al.*, 2016). The purpose of such studies is the improvement of agronomic traits, but most publications describe the ability to edit the relevant target genes rather than the traits themselves (Sovová *et al.*, 2016). For example, CRISPR/Cas9 has been used to knock out the rice *sweet* genes that confer sensitivity to bacterial blight (Jiang *et al.*, 2013b), the wheat *MLO* genes that confer sensitivity to powdery mildew (Shan *et al.*, 2013; Wang *et al.*, 2014b), and the rice *mpk5* gene that regulates defence responses (Xie and Yang, 2013). Although all these studies were successful, some were conducted on protoplasts, and even when plants were regenerated they were not tested directly for pathogen resistance. In contrast, cucumber plants in which the *eIF4E* gene encoding eukaryotic translation initiation factor 4E was mutated using CRISPR/Cas9 have been tested for resistance against *Cucumber vein yellowing virus*, *Zucchini yellow mosaic virus* and *Papaya ring spot mosaic virus* (Chandrasekaran *et al.*, 2016). Similarly, Wang *et al.* (2016) knocked out the rice *OsERF922* gene encoding the transcription factor ERF and showed that the homozygous T2 plants were more resistant to rice blast but were otherwise identical to wild-type plants in terms of growth and yield traits. Sun *et al.* (2016) used CRISPR/Cas9 to induce HDR in rice, resulting in a single nucleotide substitution in the *ALS* gene that conferred herbicide resistance. Li *et al.* (2016) individually mutated four rice genes affecting yield traits (*Gn1a*, *DEP1*, *GS3* and *IPA1*) and achieved a variety of promising mutant phenotypes in the T2 generation, including more grains (*gn1a*), denser panicles and semi-dwarf culms (*dep1*), larger grains and long awns (*gs3*), and a change in tiller number, either more or less depending on the precise target site (*ipa1*). Finally, CRISPR/Cas9 has also been used to target genes encoding polyphenol oxidases (PPOs) in mushroom. The enzyme causes browning of the fungal tissue, and knocking out one of six genes in the PPO gene family reduced overall PPO activity by 30% thus extending the shelf life (Waltz, 2016).

In academic research, the choice of genome-editing technique depends mainly on the simplicity and cost of the approach and the availability of tools and expertise, but applied research and commercial crop development must also take into account the associated intellectual property (IP) and licensing issues. Each of today's genome-editing tools is protected by patents or patent applications (Schinkel and Schillberg, 2016), and navigation of the IP landscape is straightforward in the case of oligonucleotide-directed mutagenesis, ZFNs and TALENs (Table S4). In contrast, the IP situation for the CRISPR/Cas9 technology is strongly contested by at least three major players: Massachusetts Institute of Technology/Broad Institute, UC Berkeley and Vilnius University (Schinkel and Schillberg, 2016). The ongoing legal dispute has delayed the commercial development of crops produced using CRISPR/Cas9 technology, although DuPont Pioneer has recently received an exclusive licence (Grushkin, 2016). DuPont Pioneer has exploited CRISPR/Cas9 technology for the development of drought-resistant maize and waxy maize with an improved starch composition. In the latter case, CRISPR/Cas9 was used to knock out the *Wx1* gene resulting in maize kernels that only accumulate amylopectin. The company recently announced that they will bring the genome-edited maize to the market within the next 5 years.

Summary and outlook

The CRISPR/Cas9 system has been used for genome editing in a wide range of different organisms but the outcome in terms of resolution, efficiency, accuracy and mutation structure depends on various factors including target site choice, sgRNA design, the properties of the endonuclease, the type of DSB introduced, whether or not the DSB is unique, the quantity of endonuclease and sgRNA, and the intrinsic differences in DNA repair pathways in different species, tissues and cells. Species-dependent effects include the preponderance of NHEJ compared to HDR in higher eukaryotes, contrasting with the preference for HDR in bacteria and some unicellular eukaryotes, and subtle differences in the mutation signatures generated in animals and plants (Figure 3). Whereas canonical CRISPR/Cas9 predominantly introduces small deletions (<10 bp) and single-base insertions in plants, both types of indel tend to be larger in animals (deletions <40 bp and insertions of 1–15 bp) and there is a greater frequency of larger deletions. In both animals and plants, Cas9 double nickase introduces staggered DSBs and this results in even larger indels (typically <100 bp). Another difference is the relative efficiency of larger genome rearrangements in animals compared to plants. These differences are likely to reflect species-dependent aspects of the competing NHEJ, MMEJ and HDR repair pathways, suggesting that the outcome of genome editing could be influenced by modulating the activity of particular repair enzymes, as shown by the increased prevalence of HDR in cells lacking normal levels of Ligase 4 in both animals and plants. Further investigations and detailed comparisons of genome-editing outcomes in different species will provide insight into interaction between component-specific effects (nuclease activity, sgRNA design) and host-specific effects (genome structure and content, DNA repair pathways) to enable the refinement of genome-editing strategies in a context-dependent manner.

Acknowledgements

Work on gene targeting at the RWTH Aachen University is funded by the European Research Council Advanced Grant 'Future-Pharma', Grant Number 269110. Fraunhofer IME has received funding from Dow AgroSciences for research on zinc finger nucleases. Synthetic biology and genome-editing work at the UdL is supported by grants from the Spanish Ministry of Economy and

Competitiveness (BIO2014-54426-P and BIO2014-54441-P) and the Catalan Government 2014 SGR 1296 Agricultural Biotechnology Research Group.

Conflict of Interest

Fraunhofer IME has received funding from Dow AgroSciences for research on zinc finger nucleases.

References

An, L., Hu, Y., Chang, S., Zhu, X., Ling, P., Zhang, F., Liu, J. *et al.* (2016) Efficient generation of FVII gene knockout mice using CRISPR/Cas9 nuclease and truncated guided RNAs. *Sci. Rep.* **6**, 25199.

Anders, C., Bargsten, K. and Jinek, M. (2016) Structural plasticity of PAM recognition by engineered variants of the RNA-guided endonuclease Cas9. *Mol. Cell*, **61**, 895–902.

Baltes, N.J., Gil-Humanes, J., Cermak, T., Atkins, P.A. and Voytas, D.F. (2014) DNA replicons for plant genome engineering. *Plant Cell*, **26**, 151–163.

Bao, Z., Xiao, H., Liang, J., Zhang, L., Xiong, X., Sun, N., Si, T. *et al.* (2015) Homology-integrated CRISPR–Cas (HI-CRISPR) system for one-step multigene disruption in Saccharomyces cerevisiae. *ACS Synth. Biol.* **4**, 585–594.

Bassett, A.R., Kong, L. and Liu, J. (2015) A genome-wide CRISPR library for high-throughput genetic screening in Drosophila cells. *J. Genet. Genom.* **42**, 301–309.

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

Belhaj, K., Chaparro-Garcia, A., Kamoun, S. and Nekrasov, V. (2013) Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR/Cas system. *Plant Methods*, **9**, 39.

Belhaj, K., Chaparro-Garcia, A., Kamoun, S., Patron, N.J. and Nekrasov, V. (2015) Editing plant genomes with CRISPR/Cas9. *Curr. Opin. Biotechnol.* **32**, 76–84.

Biot-Pelletier, D. and Martin, V.J.J. (2016) Seamless site-directed mutagenesis of the Saccharomyces cerevisiae genome using CRISPR-Cas9. *J. Biol. Eng.* **10**, 6.

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

Brooks, C., Nekrasov, V., Lippman, Z.B. and Van Eck, J. (2014) Efficient gene editing in tomato in the first generation using the clustered regularly interspaced short palindromic repeats/CRISPR-associated 9 System. *Plant Physiol.* **166**, 1292–1297.

Burger, A., Lindsay, H., Felker, A., Hess, C., Anders, C., Chiavacci, E., Zaugg, J. *et al.* (2016) Maximizing mutagenesis with solubilized CRISPR-Cas9 ribonucleoprotein complexes. *Development*, **143**, 2025–2037.

Cai, Y., Chen, L., Liu, X., Sun, S., Wu, C., Jiang, B., Han, T. *et al.* (2015) CRISPR/Cas9-mediated genome editing in soybean hairy roots. *PLoS ONE*, **10**, e0136064.

Chandrasekaran, J., Brumin, M., Wolf, D., Leibman, D., Klap, C., Pearlsman, M., Sherman, A. *et al.* (2016) Development of broad virus resistance in non-transgenic cucumber using CRISPR/Cas9 technology. *Mol. Plant Pathol.* **17**, 1140–1153.

Chen, K. and Gao, C. (2014) Targeted genome modification technologies and their applications in crop improvements. *Plant Cell Rep.* **33**, 575–583.

Cheng, R., Peng, J., Yan, Y., Cao, P., Wang, J., Qiu, C., Tang, L. *et al.* (2014) Efficient gene editing in adult mouse livers via adenoviral delivery of CRISPR/Cas9. *FEBS Lett.* **588**, 3954–3958.

Cho, S.W., Kim, S., Kim, Y., Kweon, J., Kim, H.S., Bae, S. and Kim, J.S. (2014) Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res.* **24**, 132–141.

Choi, P.S. and Meyerson, M. (2014) Targeted genomic rearrangements using CRISPR/Cas technology. *Nat. Commun.* **5**, 3728.

Chu, V.T., Weber, T., Wefers, B., Wurst, W., Sander, S., Rajewsky, K. and Khun, R. (2015) Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat. Biotechnol.* **33**, 543–550.

Chu, V.T., Weber, T., Graf, R., Sommermann, T., Petsch, K., Sack, U., Volchkov, P. *et al.* (2016) Efficient generation of Rosa26 knock-in mice using CRISPR/Cas9 in C57BL/6 zygotes. *BMC Biotechnol.* **16**, 4.

Cyranoski, D. (2016) Replications, ridicule and a recluse: the controversy over NgAgo gene-editing intensifies. *Nature*, **536**, 136–137.

DiCarlo, J.E., Norville, J.E., Rios, X., Aach, J. and Church, G.M. (2013) Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. *Nucleic Acids Res.* **41**, 4336–4343.

Dickinson, D.J., Ward, J.D., Reiner, D.J. and Goldstein, B. (2013) Engineering the Caenorhabditis elegans genome using Cas9-triggered homologous recombination. *Nat. Methods*, **10**, 1028–1036.

Doench, J.G., Hartenian, E., Graham, D.B., Tothova, Z., Hegde, M., Smith, I., Sullender, M. *et al.* (2014) Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat. Biotechnol.* **32**, 1262–1267.

Doudna, J.A. and Charpentier, E. (2014) The new frontier of genome engineering with CRISPR-Cas9. *Science*, **346**, 1258096.

Endo, M., Mikami, M. and Toki, S. (2015) Multigene knockout utilizing off-target mutations of the CRISPR/Cas8 system in rice. *Plant Cell Physiol.* **56**, 41–47.

Endo, M., Mikami, M. and Toki, S. (2016) Biallelic gene targeting in rice. *Plant Physiol.* **170**, 667–677.

Essletzbichler, P., Konopka, T., Santoro, F., Chen, D., Gapp, B.V., Kralovics, R., Brummelkamp, T.R. *et al.* (2014) Megabase-scale deletion using CRISPR/Cas9 to generate a fully haploid human cell line. *Genome Res.* **24**, 2059–2065.

Fagerlund, R.D., Staals, R.H. and Fineran, P.C. (2015) The Cpf1 CRISPR-Cas protein expands genome-editing tools. *Genome Biol.* **16**, 251.

Fan, D., Liu, T., Li, C., Jiao, B., Li, S., Hou, Y. and Luo, K. (2015) Efficient CRISPR/Cas9-mediated targeted mutagenesis in Populus in the first generation. *Sci. Rep.* **5**, 12217.

Fauser, F., Schiml, S. and Puchta, H. (2014) Both CRISPR/Cas-based nucleases and nickases can be used efficiently for genome engineering in *Arabidopsis thaliana*. *Plant J.* **79**, 348–359.

Feng, Z., Mao, Y., Xu, N., Zhang, B., Wei, P., Yang, D.L., Wang, Z. *et al.* (2014) Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in Arabidopsis. *Proc. Natl Acad. Sci. USA*, **111**, 4632–4637.

Friedland, A.E., Tzur, Y.B., Esvelt, K.M., Colaiácovo, M.P., Church, G.M. and Calarco, J.A. (2013) Heritable genome editing in *C. elegans* via a CRISPR-Cas9 system. *Nat. Methods*, **10**, 741–743.

Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K. and Sander, J.D. (2013) High-frequency off-target mutagenesis induced by CRISPR–Cas nucleases in human cells. *Nat. Biotechnol.* **31**, 822–826.

Gagnon, J.A., Valen, E., Thyme, S.B., Huang, P., Akhmetova, L., Pauli, A., Montague, T.G. *et al.* (2014) Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single guide RNAs. *PLoS ONE*, **9**, e98186.

Gao, J., Wang, G., Ma, S., Xie, X., Wu, X., Zhang, X., Wu, Y. *et al.* (2015) CRISPR/Cas9-mediated targeted mutagenesis in Nicotiana tabacum. *Plant Mol. Biol.* **87**, 99–110.

Gao, F., Shen, X.Z., Jiang, F., Wu, Y. and Han, C. (2016) DNA-guided genome editing using the Natronobacterium gregoryi Argonaute. *Nat. Biotechnol.* **34**, 768–773.

Gasiunas, G., Barrangou, R., Horvath, P. and Siksnys, V. (2012) Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl Acad. Sci. USA*, **109**, E2579–E2586.

Ghorbal, M., Gorman, M., Macpherson, C.R., Martins, R.M., Scherf, A. and Lopez-Rubio, J.-J. (2014) Genome editing in the human malaria parasite Plasmodium falciparum using the CRISPR-Cas9 system. *Nat. Biotechnol.* **32**, 819–821.

Gregory, T.R., Nicol, J.A., Tamm, H., Kullman, B., Kullman, K., Leitch, I.J., Murray, B.G. *et al.* (2007) Eukaryotic genome size databases. *Nucleic Acids Res.* **35**, D332–D338.

Grushkin, D. (2016) DuPont in CRISPR-Cas patent land grab. *Nat. Biotechnol.* **34**, 13.

Guilinger, J.P., Thompson, D.B. and Liu, D.R. (2014) Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat. Biotechnol.* **32**, 577–582.

Haeussler, M. and Concordet, J.P. (2016) Genome editing with CRISPR-Cas9: can it get any better? *J. Genet. Genom.* **43**, 239–250.

Heckl, D., Kowalczyk, M.S., Yudovich, D., Belizaire, R., Puram, R.V., McConkey, M.E., Thielke, A. *et al.* (2014) Generation of mouse models of myeloid malignancy with combinatorial genetic lesions using CRISPR-Cas9 genome editing. *Nat. Biotechnol.* **32**, 941–946.

Hendel, A., Fine, E.J., Bao, G. and Porteus, M. (2015) Quantifying on- and off-target genome editing. *Trends Biotechnol.* **33**, 132–140.

Horlbeck, M.A., Witkowsky, L.B., Guglielmi, B., Replogle, J.M., Gilbert, L.A., Villalta, J.E., Torigoe, S.E. *et al.* (2016) Nucleosomes impede Cas9 access to DNA in vivo and in vitro. *eLIFE*, **5**, e12677.

Hruscha, A., Krawitz, P., Rechenberg, A., Heinrich, V., Hecht, J., Haass, C. and Schmid, B. (2013) Efficient CRISPR/Cas9 genome editing with low off-target effects in zebrafish. *Development*, **140**, 4982–4987.

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y. *et al.* (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31**, 827–832.

Hsu, P.D., Lander, E.S. and Zhang, F. (2014) Development and applications of CRISPR-Cas9 for genome engineering. *Cell*, **157**, 1262–1278.

Hu, X., Wang, C., Fu, Y., Liu, Q., Jiao, X. and Wang, K. (2016) Expanding the range of CRISPR/Cas9 genome editing in rice. *Mol. Plant*, **9**, 943–945.

Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Tsai, S.Q., Sander, J.D., Peterson, R.T. *et al.* (2013) Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat. Biotechnol.* **31**, 227–229.

Hyun, Y., Kim, J., Cho, S.W., Choi, Y., Kim, J.S. and Coupland, G. (2015) Site-directed mutagenesis in *Arabidopsis thaliana* using dividing tissue-targeted RGEN of the CRISPR/Cas system to generate heritable null alleles. *Planta*, **241**, 271–284.

Jacobs, T.B., LaFayette, P.R., Schmitz, R.J. and Parrott, W.A. (2015) Targeted genome modifications in soybean with CRISPR/Cas9. *BMC Biotechnol.* **15**, 16.

Jang, Y.Y. and Ye, Z. (2016) Gene correction in patient-specific iPSCs for therapy development and disease modeling. *Human Genet.* **135**, 1041–1058.

Jia, H., Orbovic, V., Jones, J.B. and Wang, N. (2016) Modification of the PthA4 effector binding elements in Type I CsLOB1 promoter using Cas9/sgRNA to produce transgenic Duncan grapefruit alleviating XccΔpthA4:dCsLOB1.3 infection. *Plant Biotechnol. J.* **14**, 1291–1301.

Jiang, W., Bikard, D., Cox, D., Zhang, F. and Marraffini, L.A. (2013a) RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* **31**, 233–239.

Jiang, W., Zhou, H., Bi, H., Fromm, M., Yang, B. and Weeks, D.P. (2013b) Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. *Nucleic Acids Res.* **41**, e188.

Jiang, W., Brueggeman, A.J., Horken, K.M., Plucinak, T.M. and Weeks, D.P. (2014) Successful transient expression of Cas9/sgRNA genes in *Chlamydomonas reinhardtii*. *Eukaryot. Cell*, **13**, 1465–1469.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A. and Charpentier, E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, **337**, 816–821.

Karvelis, T., Gasiunas, G., Young, J., Bidelyte, G., Silanskas, A., Cigan, M. and Siksnys, V. (2015) Rapid characterization of CRISPR-Cas9 protospacer adjacent motif sequence elements. *Genome Biol.* **16**, 253.

Khatodia, S., Bhatotia, K., Passricha, N., Khurana, S.M.P. and Tuteja, N. (2016) The CRISPR/Cas genome-editing tool: application in improvement of crops. *Front. Plant Sci.* **7**, 506.

Kim, D., Bae, S., Park, J., Kim, E., Kim, S., Yu, H.R., Hwang, J. *et al.* (2015) Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat. Methods*, **12**, 237–243.

Kleinstiver, B.P., Prew, M.S., Tsai, S.Q., Topkar, V.V., Nguyen, N.T., Zheng, Z., Gonzales, A.P. *et al.* (2015) Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature*, **523**, 481–485.

Kleinstiver, B.P., Pattanayak, V., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Zheng, Z. and Joung, J.K. (2016) High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*, **529**, 490–495.

Kumar, V. and Jain, M. (2015) The CRISPR–Cas system for plant genome editing: advances and opportunities. *J. Exp. Bot.* **66**, 47–57.

Lawrenson, T., Shorinola, O., Stacey, N., Li, C., Østergaard, L., Patron, N., Uauy, C. *et al.* (2015) Induction of targeted, heritable mutations in barley and *Brassica oleracea* using RNA-guided Cas9 nuclease. *Genome Biol.* **16**, 258.

Lee, J., Chung, J.H., Kim, H.M., Kim, D.W. and Kim, H. (2015) Designed nucleases for targeted genome editing. *Plant Biotechnol. J.* **14**, 448–462.

Li, X.Q. and Du, D. (2014) Variation, evolution, and correlation analysis of C+G content and genome or chromosome size in different kingdoms and phyla. *PLoS ONE*, **9**, e88339.

Li, J.F., Norville, J.E., Aach, J., McCormack, M., Zhang, D., Bush, J., Church, G.M. *et al.* (2013) Multiplex and homologous recombination–mediated genome editing in Arabidopsis and *Nicotiana benthamiana* using guide RNA and Cas9. *Nat. Biotechnol.* **31**, 688–691.

Li, J., Shou, J., Guo, Y., Tang, Y., Wu, Y., Jia, Z., Zhai, Y. *et al.* (2015) Efficient inversions and duplications of mammalian regulatory DNA elements and gene clusters by CRISPR/Cas9. *J. Mol. Cell Biol.* **7**, 284–298.

Li, M., Li, X., Zhou, Z., Wu, P., Fang, M., Pan, X., Lin, Q. *et al.* (2016) Reassessment of the four yield-related genes Gn1a, DEP1, GS3, and IPA1 in rice using a CRISPR/Cas9 system. *Front. Plant Sci.* **7**, 377.

Liang, X., Potter, J., Kumar, S., Zou, Y., Quintanilla, R., Sridharan, M., Carte, J. *et al.* (2015) Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. *J. Biotechnol.* **208**, 44–53.

Liang, G., Zhang, H., Lou, D. and Yu, D. (2016) Selection of highly efficient sgRNAs for CRISPR/Cas9-based plant genome editing. *Sci. Rep.* **6**, 21451.

Lin, Y., Cradick, T.J., Brown, M.T., Deshmukh, H., Ranjan, P., Sarode, N., Wile, B.M. *et al.* (2014) CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. *Nucleic Acid Res.* **42**, 7473–7485.

Lowder, L.G., Zhang, D., Baltes, N.J., Paul, J.W. III, Tang, X., Zheng, X., Voytas, D.F. *et al.* (2015) A CRISPR/Cas9 toolbox for multiplexed plant genome editing and transcriptional regulation. *Plant Physiol.* **169**, 971–985.

Lv, Q., Yuan, L., Deng, J., Chen, M., Wang, Y., Zeng, J., Li, Z. *et al.* (2016) Efficient generation of myostatin gene mutated rabbit by CRISPR/Cas9. *Sci. Rep.* **6**, 25029.

Ma, X., Zhang, Q., Zhu, Q., Liu, W., Chen, Y., Qiu, R., Wang, B. *et al.* (2015) A robust CRISPR/Cas9 system for convenient high-efficiency multiplex genome editing in monocot and dicot plants. *Mol. Plant*, **8**, 1274–1284.

Maddalo, D., Manchado, E., Concepcion, C.P., Bonetti, C., Vidigal, J.A., Han, Y.C., Ogrodowski, P. *et al.* (2014) In vivo engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system. *Nature*, **516**, 423–427.

Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J., Charpentier, E., Horvath, P., Moineau, S. *et al.* (2011) Evolution and classification of the CRISPR-Cas systems. *Nat. Rev. Microbiol.* **9**, 467–477.

Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J., Barrangou, R. *et al.* (2015) An updated evolutionary classification of CRISPR–Cas systems. *Nat. Rev. Microbiol.* **13**, 722–736.

Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L. *et al.* (2013) CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* **31**, 833–838.

Mao, Y., Zhang, Z., Feng, Z., Wie, P., Zhang, H., Botella, J.R. and Zhou, J.K. (2016) Development of germ-line-specific CRISPR-Cas9 systems to improve the production of heritable gene modifications in Arabidopsis. *Plant Biotechnol. J.* **14**, 519–532.

Maresca, M., Lin, V.G., Guo, N. and Yang, Y. (2013) Obligate ligation-gated recombination (ObLiGaRe): custom-designed nuclease-mediated targeted integration through nonhomologous end joining. *Genome Res.* **23**, 539–546.

McVey, M. and Lee, S.E. (2008) MMEJ repair of double-stranded breaks (director's cut): deleted sequences and alternative endings. *Trends Genet.* **24**, 529–538.

Miao, J., Guo, D., Zhang, J., Huang, Q., Qin, G., Zhang, X., Wan, J. *et al.* (2013) Targeted mutagenesis in rice using CRISPR-Cas system. *Cell Res.* **23**, 1233–1236.

Michael, T.P. and Jackson, S. (2013) The first 50 plant genomes. *Plant Genome*, **6**, 2.

Mikami, M., Toki, S. and Endo, M. (2015a) Comparison of CRISPR/Cas9 expression constructs for efficient targeted mutagenesis in rice. *Plant Mol. Biol.* **88**, 561–572.

Mikami, M., Toki, S. and Endo, M. (2015b) Parameters affecting frequency of CRISPR/Cas9 mediated targeted mutagenesis in rice. *Plant Cell Rep.* **34**, 1807–1815.

Mikami, M., Toki, S. and Endo, M. (2016) Precision targeted mutagenesis via Cas9 paired nickases in rice. *Plant Cell Physiol.* **57**, 1058–1068.

Miyaoka, Y., Berman, J.R., Cooper, S.B., Mayert, S.J., Chan, A.H., Zhang, B., Karlin-Neumann, G.A. *et al.* (2016) Systematic quantification of HDR and

NHEJ reveals effects of locus, nuclease, and cell type on genome editing. *Sci. Rep.* **6**, 23549.
Mohr, S.E., Hu, Y., Ewen-Campen, B., Housden, B.E., Viswanatha, R. and Perrimon, N. (2016) CRISPR guide RNA design for research applications. *FEBS J.* **283**, 3232–3238.
Nymark, M., Sharma, A.K., Sparstad, T., Bones, A.M. and Winge, P. (2016) A CRISPR/Cas9 system adapted for gene editing in marine algae. *Sci. Rep.* **6**, 24951.
Oishi, I., Yoshii, K., Miyahara, D., Kagami, H. and Tagami, T. (2016) Targeted mutagenesis in chicken using CRISPR-Cas9 system. *Sci. Rep.* **6**, 23980.
Osakabe, Y. and Osakabe, K. (2015) Genome editing with engineered nucleases in plants. *Plant Cell Physiol.* **56**, 389–400.
Pan, C., Ye, L., Qin, L., Liu, X., He, Y., Wang, J., Chen, L. *et al.* (2016) CRISPR/Cas9-mediated efficient and heritable targeted mutagenesis in tomato plants in the first and later generations. *Sci. Rep.* **6**, 24765.
Pattanayak, V., Lin, S., Guilinger, J.P., Ma, E., Doudna, J.A. and Liu, D.R. (2013) High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat. Biotechnol.* **31**, 839–843.
Platt, R.J., Chen, S., Zhou, Y., Yim, M.J., Swiech, L., Kempton, H.R., Dahlman, J.E. *et al.* (2014) CRISPR-Cas9 knockin mice for genome editing and cancer modeling. *Cell*, **159**, 440–455.
Puchta, H. (2005) The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution. *J. Exp. Bot.* **56**, 1–14.
Quétier, F. (2016) The CRISPR-Cas9 technology: closer to the ultimate toolkit for targeted genome editing. *Plant Sci.* **242**, 65–76.
Ran, F.A., Hsu, P.D., Lin, C.Y., Gootenberg, J.S., Konermann, S., Trevino, A., Scott, D.A. *et al.* (2013) Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*, **154**, 1380–1389.
Rodriguez, E., Keiser, M., McLoughlin, H., Zhang, F. and Davidson, B.L. (2014) AAV-CRISPR: a new therapeutic approach to nucleotide repeat diseases. *Mol. Ther.* **22**, S94.
Sander, J.D. and Joung, J.K. (2014) CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* **32**, 347–355.
Sauer, N.J., Narvaez-Vasquez, J., Mozoruk, J., Miller, R.B., Warburg, Z.J., Woodward, M.J., Mihiret, J.A. *et al.* (2016) Oligonucleotide-mediated genome editing provides precision and function to engineered nucleases and antibiotics in plants. *Plant Physiol.* **170**, 1917–1928.
Schiml, S., Fauser, F. and Puchta, H. (2014) The CRISPR/Cas system can be used as nuclease for in planta gene targeting and as paired nickases for directed mutagenesis in Arabidopsis resulting in heritable progeny. *Plant J.* **80**, 1139–1150.
Schiml, S., Fauser, F. and Puchta, H. (2016) Repair of adjacent single-strand breaks is often accompanied by the formation of tandem sequence duplications in plant genomes. *Proc. Natl Acad. Sci. USA* **113**, 7266–7271.
Schinkel, H. and Schillberg, S. (2016) Genome editing: intellectual property and product development in plant biotechnology. *Plant Cell Rep.* **35**, 1487–1491.
Selle, K. and Barrangou, R. (2015) Harnessing CRISPR–Cas systems for bacterial genome editing. *Trends Microbiol.* **23**, 225–232.
Senis, E., Fatouros, C., Grosse, S., Wiedtke, E., Niopek, D., Mueller, A.K., Borner, K. *et al.* (2014) CRISPR/Cas9-mediated genome engineering: an adeno-associated viral (AAV) vector toolbox. *Biotechnol. J.* **9**, 1402–1412.
Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelson, T., Heckl, D. *et al.* (2014) Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science*, **343**, 84–87.
Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., Zhang, K. *et al.* (2013) Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat. Biotechnol.* **31**, 686–688.
Shao, Y., Guan, Y., Wang, L., Qiu, Z., Liu, M., Chen, Y., Wu, L. *et al.* (2014) CRISPR/Cas-mediated genome editing in the rat via direct injection of one-cell embryos. *Nat. Prot.* **9**, 2493–2512.
Shin, S.E., Lim, J.M., Koh, H.G., Kim, E.K., Kang, N.K., Jeon, S., Kwon, S. *et al.* (2016) CRISPR/Cas9-induced knockout and knock-in mutations in *Chlamydomonas reinhardtii*. *Sci. Rep.* **6**, 27810.
Slaymaker, I.M., Gao, L., Zetsche, B., Scott, D.A., Yan, W.X. and Zhang, F. (2016) Rationally engineered Cas9 nucleases with improved specificity. *Science*, **351**, 84–88.
Smith, C., Gore, A., Yan, W., Abalde-Atristain, L., Li, Z., He, C., Wang, Y. *et al.* (2014) Whole-genome sequencing analysis reveals high specificity of CRISPR/Cas9 and TALEN-based genome editing in human iPSCs. *Cell Stem Cell*, **15**, 12–13.
Song, G., Jia, M., Chen, K., Kong, X., Khattak, B., Xie, C., Li, A. *et al.* (2016) CRISPR/Cas9: a powerful tool for crop genome editing. *Crop J.* **4**, 75–82.
Sovová, T., Kerins, G., Demnerová, K. and Ovesná, J. (2016) Genome editing with engineered nucleases in economically important animals and plants: state of the art in the research pipeline. *Curr. Issues Mol. Biol.* **21**, 41–62.
Sprink, T., Eriksson, D., Schiemann, J. and Hartung, F. (2016) Regulatory hurdles for genome editing: process vs product-based approaches in different regulatory contexts. *Plant Cell Rep.* **35**, 1493–1506.
Steinert, J., Schiml, S., Fauser, F. and Puchta, H. (2015) Highly efficient heritable plant genome engineering using Cas9 orthologues from *Streptococcus thermophilus* and *Staphylococcus aureus*. *Plant J.* **84**, 1295–1305.
Stovicek, V., Borodina, I. and Forster, J. (2015) CRISPR–Cas system enables fast and simple genome editing of industrial *Saccharomyces cerevisiae* strains. *Metab. Eng. Commun.* **2**, 13–22.
Subburaj, S., Chung, S.J., Lee, C., Ryu, S.M., Kim, D.H., Kim, J.S., Bae, S. *et al.* (2016) Site-directed mutagenesis in Petunia × hybrida protoplast system using direct delivery of purified recombinant Cas9 ribonucleoproteins. *Plant Cell Rep.* **35**, 1535–1544.
Sun, X., Hu, Z., Chen, R., Jiang, Q., Song, G., Zhang, H. and Xi, Y. (2015) Targeted mutagenesis in soybean using the CRISPR-Cas9 system. *Sci. Rep.* **5**, 10342.
Sun, Y., Zhang, X., Wu, C., He, Y., Ma, Y., Hou, H., Guo, X. *et al.* (2016) Engineering herbicide-resistant rice plants through CRISPR/Cas9-mediated homologous recombination of acetolactate synthase. *Mol. Plant*, **9**, 628–632.
Svitashev, S., Young, J.K., Schwartz, C., Gao, H., Falco, S.C. and Cigan, A.M. (2015) Targeted mutagenesis, precise gene editing, and site-specific gene insertion in maize using Cas9 and guide RNA. *Plant Physiol.* **169**, 931–945.
Swiech, L., Heidenreich, M., Banerjee, A., Habib, N., Li, Y., Trombetta, J., Sur, M. *et al.* (2015) In vivo interrogation of gene function in the mammalian brain using CRISPR-Cas9. *Nat. Biotechnol.* **33**, 102–106.
Torres, R., Martin, M.C., Garcia, A., Cigudosa, J.C., Ramirez, J.C. and Rodriguez-Perales, S. (2014) Engineering human tumour-associated chromosomal translocations with the RNA-guided CRISPR-Cas9 system. *Nat. Commun.* **5**, 4964.
Tsai, S.Q., Wyvekens, N., Khayter, C., Foden, J.A., Thapar, V., Reyon, D., Goodwin, M.J. *et al.* (2014) Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat. Biotechnol.* **32**, 569–576.
Tsai, C.S., Kong, I.I., Lesmana, A., Million, G., Zhang, G.C., Kim, S.R. and Jin, Y.S. (2015a) Rapid and marker-free refactoring of xylose-fermenting yeast strains with Cas9/CRISPR. *Biotechnol. Bioeng.* **112**, 2406–2411.
Tsai, S.Q., Zheng, Z., Nguyen, N.T., Liebers, M., Topkar, V.V., Thapar, V., Wyvekens, N. *et al.* (2015b) GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33**, 187–197.
Tschaharganeh, D.F., Lowe, S.W., Garippa, R.J. and Livshits, G. (2016) Using CRISPR/Cas to study gene function and model disease in vivo. *FEBS J.* **283**, 3194–3203.
Upadhyay, S.K., Kumar, J., Alok, A. and Tuli, R. (2013) RNA-guided genome editing for target gene mutations in wheat. *G3 Genes Genomes Genet.* **3**, 2233–2238.
Varshney, G.K., Pei, W., Lafave, M.C., Idol, J., Xu, L., Gallardo, V., Carrington, B. *et al.* (2015) High-throughput gene targeting and phenotyping in zebrafish using CRISPR/Cas9. *Genome Res.* **25**, 1030–1042.
Veres, A., Gosis, B.S., Ding, Q., Collins, R., Ragavendran, A., Brand, H., Erdin, S. *et al.* (2014) Low incidence of off-target mutations in individual CRISPR-Cas9 and TALEN targeted human stem cell clones detected by whole-genome sequencing. *Cell Stem Cell*, **15**, 27–30.
Waltz, E. (2016) Gene-edited CRISPR mushroom escapes US regulation. *Nature*, **532**, 293.
Wang, H., Tang, H., Shivalila, C.S., Dawlaly, M.M., Cheng, A.W., Zhang, F. and Jaenisch, R. (2013) One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell*, **153**, 910–918.
Wang, T., Wei, J.J., Sabatini, D.M. and Lander, E.S. (2014a) Genetic screens in human cells using the CRISPR-Cas9 system. *Science*, **343**, 80–84.

Wang, Y., Cheng, X., Shan, Q., Zhang, Y., Liu, J., Gao, C. and Qiu, J.-L. (2014b) Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nat. Biotechnol.* **32**, 947–951.

Wang, S., Zhang, S., Wang, W., Xiong, X., Meng, F. and Cui, X. (2015a) Efficient targeted mutagenesis in potato by the CRISPR/Cas9 system. *Plant Cell Rep.* **34**, 1473–1476.

Wang, Z.P., Xing, H.L., Dong, L., Zhang, H.Y., Han, C.Y., Wang, X.C. and Chen, Q.J. (2015b) Egg cell-specific promoter-controlled CRISPR/Cas9 efficiently generates homozygous mutants for multiple target genes in Arabidopsis in a single generation. *Genome Biol.* **16**, 144.

Wang, F., Wang, C., Liu, P., Lei, C., Hao, W., Gao, Y., Liu, Y.G. *et al.* (2016) Enhanced rice blast resistance by CRISPR/Cas9-targeted mutagenesis of the ERF transcription factor gene OsERF922. *PLoS ONE*, **11**, e0154027.

Weninger, A., Hatzl, A.M., Schmid, C., Vogl, T. and Glieder, A. (2016) Combinatorial optimization of CRISPR/Cas9 expression enables precision genome engineering in the methylotrophic yeast Pichia pastoris. *J. Biotechnol.* **235**, 139–149.

Woo, J.W., Kim, J., Kwon, S.I., Corvalan, C., Cho, S.W., Kim, H., Kim, S.-G. *et al.* (2015) DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. *Nat. Biotechnol.* **33**, 1162–1164.

Xie, K. and Yang, Y. (2013) RNA-guided genome editing in plants using a CRISPR-Cas system. *Mol. Plant*, **6**, 1975–1983.

Xie, K., Zhang, J. and Yang, Y. (2014) Genome-wide prediction of highly specific guide RNA spacers for CRISPR-Cas9-mediated genome editing in model plants and major crops. *Mol. Plant*, **7**, 923–926.

Xie, K., Minkenberg, B. and Yang, Y. (2015) Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proc. Natl Acad. Sci. USA*, **112**, 3570–3575.

Xu, H., Xiao, T., Chen, C.H., Li, W., Meyer, C.A., Wu, Q., Wu, D. *et al.* (2015a) Sequence determinants of improved CRISPR sgRNA design. *Genome Res.* **25**, 1147–1157.

Xu, R.F., Li, H., Qin, R.Y., Li, J., Qiu, C.H., Yang, Y.C., Ma, H. *et al.* (2015b) Generation of inheritable and "transgene clean" targeted genome-modified rice in later generations using the CRISPR/Cas9 system. *Sci. Rep.* **5**, 11491.

Xu, T., Li, Y., Shi, Z., Hemme, C.L., Li, Y., Zhu, Y., Van Nostrand, J.D. *et al.* (2015c) Efficient genome editing in Clostridium cellulolyticum via CRISPR-Cas9 nickase. *Appl. Env. Microbiol.* **81**, 4423–4431.

Xue, W., Chen, S., Yin, H., Tammela, T., Papagiannakopoulos, T., Joshi, N.S., Cai, W. *et al.* (2014) CRISPR-mediated direct mutation of cancer genes in the mouse liver. *Nature*, **514**, 380–384.

Yan, L., Wei, S., Wu, Y., Hu, R., Li, H., Yang, W. and Xie, Q. (2015) High-efficiency genome editing in Arabidopsis using YAO promoter-driven CRISPR/Cas9 system. *Mol. Plant*, **8**, 1820–1823.

Yang, L., Guell, M., Byrne, S., Yang, J.L., De Los Angeles, A., Mali, P., Aach, J. *et al.* (2013) Optimization of scarless human stem cell genome editing. *Nucleic Acids Res.* **41**, 9049–9061.

Yang, H., Wang, H. and Jaenisch, R. (2014) Generating genetically modified mice using CRISPR/Cas-mediated genome engineering. *Nat. Prot.* **9**, 1956–1968.

Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E. *et al.* (2015) Cpf1 Is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*, **163**, 759–771.

Zhang, H., Zhang, J., Wei, P., Zhang, B., Gou, F., Feng, Z., Mao, Y. *et al.* (2014) The CRISPR/Cas9 system produces specific and homozygous targeted gene editing in rice in one generation. *Plant Biotechnol. J.* **12**, 797–807.

Zhang, L., Jia, R., Palange, N.J., Satheka, A.C., Togo, J., An, Y., Humphrey, M. *et al.* (2015) Large genomic fragment deletions and insertions in mouse using CRISPR/Cas9. *PLoS ONE*, **10**, e0120396.

Zhang, B., Yang, X., Yang, C., Li, M. and Guo, Y. (2016) Exploiting the CRISPR/Cas9 system for targeted genome mutagenesis in petunia. *Sci. Rep.* **6**, 20315.

Zhao, Y., Zhang, C., Liu, W., Gao, W., Liu, C., Song, G., Li, W.X. *et al.* (2016) An alternative strategy for targeted gene replacement in plants using a dual-sgRNA/Cas9 design. *Sci. Rep.* **6**, 23890.

Zhou, H., Liu, B., Weeks, D.P., Spalding, M.H. and Yang, B. (2014) Large chromosomal deletions and heritable small genetic changes induced by CRISPR/Cas9 in rice. *Nucleic Acids Res.* **42**, 10903–10914.

Zhou, X., Jacobs, T.B., Xue, L.J., Harding, S.A. and Tsai, C.J. (2015) Exploiting SNPs for biallelic CRISPR mutations in the outcrossing woody perennial Populus reveals 4-coumarate: CoA ligase specificity and redundancy. *New Phytol.* **208**, 298–301.

Zhu, Z., Verma, N., González, F., Shi, Z.D. and Huangfu, D. (2015) A CRISPR/Cas-mediated selection-free knockin strategy in human embryonic stem cells. *Stem Cell Rep.* **4**, 1103–1111.

Supporting information

Additional Supporting Information may be found online in the supporting information tab for this article:

Table S1 The efficiency, accuracy and structure of on/off-target mutations induced by CRISPR systems in different plant species.
Table S2 The efficiency, accuracy and structure of on/off-target mutations induced by CRISPR systems in different animal species.
Table S3 The efficiency, accuracy and structure of on/off-target mutations induced by CRISPR systems in different microbial species.
Table S4 Examples for the development of commercial plant products using different genome editing technologies and the involved IP.

