



Universitat de Lleida

## Genetic engineering and synthetic biology-based approaches for the production of novel valuable materials

(International mention)

Erika Lorena Soto Chavarro

<http://hdl.handle.net/10803/676007>

**ADVERTIMENT.** L'accés als continguts d'aquesta tesi doctoral i la seva utilització ha de respectar els drets de la persona autora. Pot ser utilitzada per a consulta o estudi personal, així com en activitats o materials d'investigació i docència en els termes establerts a l'art. 32 del Text Refós de la Llei de Propietat Intel·lectual (RDL 1/1996). Per altres utilitzacions es requereix l'autorització prèvia i expressa de la persona autora. En qualsevol cas, en la utilització dels seus continguts caldrà indicar de forma clara el nom i cognoms de la persona autora i el títol de la tesi doctoral. No s'autoritza la seva reproducció o altres formes d'explotació efectuades amb finalitats de lucre ni la seva comunicació pública des d'un lloc aliè al servei TDX. Tampoc s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX (framing). Aquesta reserva de drets afecta tant als continguts de la tesi com als seus resums i índexs.

**ADVERTENCIA.** El acceso a los contenidos de esta tesis doctoral y su utilización debe respetar los derechos de la persona autora. Puede ser utilizada para consulta o estudio personal, así como en actividades o materiales de investigación y docencia en los términos establecidos en el art. 32 del Texto Refundido de la Ley de Propiedad Intelectual (RDL 1/1996). Para otros usos se requiere la autorización previa y expresa de la persona autora. En cualquier caso, en la utilización de sus contenidos se deberá indicar de forma clara el nombre y apellidos de la persona autora y el título de la tesis doctoral. No se autoriza su reproducción u otras formas de explotación efectuadas con fines lucrativos ni su comunicación pública desde un sitio ajeno al servicio TDR. Tampoco se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR (framing). Esta reserva de derechos afecta tanto al contenido de la tesis como a sus resúmenes e índices.

**WARNING.** Access to the contents of this doctoral thesis and its use must respect the rights of the author. It can be used for reference or private study, as well as research and learning activities or materials in the terms established by the 32nd article of the Spanish Consolidated Copyright Act (RDL 1/1996). Express and previous authorization of the author is required for any other uses. In any case, when using its content, full name of the author and title of the thesis must be clearly indicated. Reproduction or other forms of for profit use or public communication from outside TDX service is not allowed. Presentation of its content in a window or frame external to TDX (framing) is not authorized either. These rights affect both the content of the thesis and its abstracts and indexes.

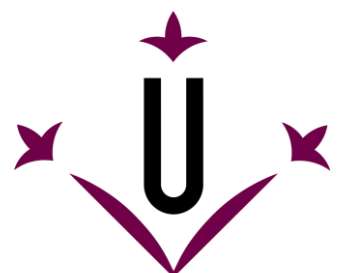


DOCTORAL THESIS

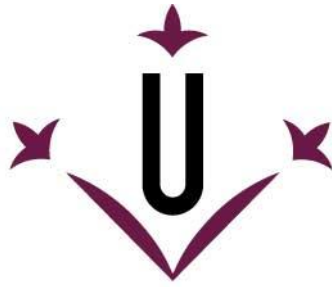
# Genetic engineering and synthetic biology based approaches for the production of novel valuable materials



**ERIKA LORENA  
SOTO CHAVARRO**



**Universitat de Lleida**



**Universitat de Lleida**

**TESI DOCTORAL**

**Genetic engineering and synthetic biology-based approaches for the production of novel valuable materials**

**(International mention)**

**Erika Lorena Soto Chavarro**

Memòria presentada per optar al grau de Doctor per la Universitat de Lleida

Programa de Doctorat en Ciència y Tecnología Agraria i Alimentaria

Directors

**Gemma Villorbina Noguera**

**Ludovic Bassié**

2022







*“Convierte tu sed  
en mares de intenso  
amor por tus sueños”.*

*Sin miedo al dolor  
Elkin Ramírez, Kraken*

## **Dedication**

*To God, for giving me strength in the most  
difficult moments of my life.*

*To my husband Wilmar and my two  
children Emmanuel and Joan David  
who are the most beautiful engine  
that drives my life and fills it with joy and love.*

*And to my parents for their infinite love.*





## Acknowledgements

I would like to express my sincere gratitude to my supervisor Dr. Gemma Villorbina for supporting me over the years with great enthusiasm and patience and co-supervisor Dr. Ludovic Bassie for his useful discussions, for his constant help and his kind advice. It has been an absolute pleasure to have your guidance in all the time of research and writing of this thesis. Thanks to your human quality helped me develop and make this project a reality. Without your guidance and encouragement this PhD would not have been achievable.

I would like to offer my special thanks to Dr. Paul Christou who with his helpful feedback and in depth knowledge was invaluable to the project, I appreciate his patient guidance, encouragement and excellent advice throughout my PhD thesis.

A very special thanks to Dr. Teresa Capell, who was always ready to help, for supporting me throughout all these years, for her constant feedback and her energetic personality and enthusiasm on science. To the professors Dr. Vicente Medina, for his kind help with the microscopy work and Pilar Muñoz for her kind guidance.

I gratefully acknowledge the Universitat de Lleida (UDL, Spain) for providing the PhD fellowship. And a special thank you to the secretary Nuria Gabernet for assisting me with many official documents. I would like to thank Jaume Capell for his practical training in growing plants and taking care of my rice plants in the greenhouse. I am thankful to Núria Gabernet for dealing with administrative work and for always being kind and helpful.

My PhD was immeasurable aided by the very friendly atmosphere and continuous technical and emotional support during the last years, both in the plant biotechnology laboratory and the organic chemistry laboratory where I had the opportunity to learn from each and every one of its members. It is a pleasure to convey my gratitude to all people who helped me in achieving this important purpose. 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 Lucía Perez 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 being incomparable lab colleague and friend.

To Riad Nadi, Giobbe Forni, Gemma Masip, Daniela Zanga, Eduard Molinero and Mohamed Herma, from Plant Biotech lab and Alberto Millán, Diana Cosovanu, Paulo Torres, Johana Aguilera and Edinson Yara from Organic Chemistry lab we shared many unforgettable moments during my PhD, thank you for all for their friendship, their support and their encouragement during these years.

Also, I would like to thank to professor Michael Krogh Jensen and colleagues Duschisca, Emil, Lea, David, Konrad and Jie in Biosustain DTU during my PhD stay in Denmark.

I want to express my gratitude to Nancy Ortega and Anna Espart for her unconditional friendship, her company both good and bad moments and for becoming my family even being far from my home, Colombia.

Finally, I want to thank my entire family, especially my husband Wilmar, my two children Emmanuel and Joan David, my brother Eduardo Alfonso and my parents Inés and Luis Eduardo for giving me their support, encouragement, and for believing in me.

## Summary

This research work was conceived as part of a larger and more complex scientific project that includes the application of metabolic engineering and genome editing tools to introduce the squalene metabolic pathway into rice (*Oryza sativa*) and into *E. coli* as a platform for the production of other compounds of interest such as carotenoids and comparatively analyse the two biological models in economic and fundamental science terms.

Specifically, the big project covers two parts in rice, of which the first one was the objective of this thesis. The first part consisted of achieving the knockout of the key endosperm genes for starch biosynthesis and observing the effects on the metabolic machinery of the plant; the second part aimed to achieve the insertion of the genes involved in the biosynthesis of squalene in the same target where the knock-out was produced, by means of CRISPR/Cas9.

In chapter two of this thesis (manuscript in preparation) we explored three methodological approaches for the microbial metabolic engineering leading to the production of squalene by the application of different synthetic biology tools. We designed and tested three genetic engineering strategies to be applied to *E. coli*. 1) Insertion of heterologous MEP pathway genes and squalene synthase genes in *E. coli*; 2) Overexpression of *E. coli* MEP pathway genes and heterologous squalene synthase genes in *E. coli*; 3) Expression of heterologous squalene synthase genes in the *E. coli* strain engineered with the dual MEP/MVA pathway. With strategy 1, the highest instant productivity of squalene (3.8 mg/L/h) was reached, which is in the third position of the highest productivity reported so far in the literature. This engineered strain could have broad potential to be used for industrial-scale squalene production.

Chapter three resulted in the manuscript *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*. In this study, 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*.

Chapter four derived from the manuscript *CRISPR/Cas9 mutations in the rice Waxy/GBSSI gene induce allele-specific and zygoty-dependent feedback effects on endosperm starch biosynthesis*. In this work we used CRISPR/Cas9 to introduce mutations affecting the *Waxy* (*Wx*) locus encoding granule-bound starch synthase one



(GBSSI) in rice endosperm. We found that the mutations reduced but did not abolish GBSS activity in seeds due to partial compensation caused by the 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.

With this research and findings, it is expected to increase the knowledge in biotechnology using metabolic engineering and synthetic biology tools, as they are key to facing the challenges posed by the Sustainable Development Goals (SDG). Responsible consumption and production (SDG 12), life under water (SDG 14), and zero hunger (SDG 2) can be addressed by the creation of alternative sources for sustainable squalene production to avoid shark's uncontrolled slaughter and improving rice crops by genome editing.

## Resumen

Este trabajo de investigación se concibió como parte de un proyecto científico más grande y complejo que incluye la aplicación de herramientas de ingeniería metabólica y edición del genoma para introducir la vía metabólica del escualeno en el arroz (*Oryza sativa*) y en *E. coli* como plataforma para la producción de otros compuestos de interés como los carotenoides y analizar comparativamente los dos modelos biológicos en términos económicos y ciencia fundamental.

Concretamente en arroz, el gran proyecto consta de dos partes, de las cuales la primera era el objetivo de esta tesis. La primera parte consistió en lograr el bloqueo de genes clave del endospermo para la biosíntesis del almidón y observar los efectos sobre la maquinaria metabólica de la planta; la segunda parte pretende conseguir la inserción de los genes implicados en la síntesis de escualeno en el mismo target donde se produjo el bloqueo de genes, mediante CRISPR/Cas9.

En el capítulo dos de esta tesis (manuscrito en preparación) se exploraron tres enfoques metodológicos para la ingeniería metabólica microbiana que conducen a la producción de escualeno mediante la aplicación de diferentes herramientas de biología sintética. Se diseñaron y probaron tres estrategias de ingeniería genética para ser aplicadas a *E. coli*. 1) Inserción de genes heterólogos de la vía MEP y genes de escualeno sintasa en *E. coli*; 2) Sobreexpresión de genes de la vía MEP de *E. coli* y genes heterólogos de escualeno sintasa en *E. coli*; 3) Expresión de genes heterólogos de escualeno sintasa en la cepa de *E. coli* diseñada con la vía dual MEP/MVA. Con la estrategia 1 se alcanzó la mayor productividad instantánea de escualeno (3, 8 mg/L/h), la cual se encuentra en la tercera posición de mayor productividad reportada hasta el momento en la literatura. Esta cepa diseñada podría tener un amplio potencial para usarse en la producción de escualeno a escala industrial.

El capítulo tres dió como resultado el manuscrito *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*. En este estudio, se utilizó CRISPR/Cas9 para crear dos mutantes heterocigóticos, uno con una AGPasa severamente truncada y no funcional y el otro con una modificación estructural C-terminal que causa una pérdida parcial de actividad. Inesperadamente, se observó el agotamiento del almidón en las hojas de ambos mutantes y el correspondiente aumento en el nivel de azúcares solubles. Esto reflejó la expresión no anticipada de *OsAPL2* y *OsAPS2b* en las hojas, generando una AGPasa ectópica completa en el citosol de la hoja, y una disminución correspondiente en la expresión de la pequeña subunidad plastidial *OsAPS2a* que solo se complementó parcialmente con un aumento en la expresión de *OsAPS1*.

El capítulo cuatro derivó en el manuscrito *CRISPR/Cas9 mutations in the rice Waxy/GBSSI gene induce allele-specific and zygoty-dependent feedback effects on endosperm starch biosynthesis*. En este trabajo, se utilizó CRISPR/Cas9 para introducir mutaciones que afectan al locus *Waxy* (*Wx*) que codifica la enzima granule-bound starch synthase I

(GBSSI) en el endospermo de arroz. Se encontró que las mutaciones disminuyeron pero no eliminaron la actividad de GBSSI en las semillas debido a la compensación parcial causada por la regulación positiva de GBSSII. La actividad de GBSS en los mutantes fue del 61 al 71 % de los niveles de wild type, de manera similar a dos mutantes creados con radiación, pero el contenido de amilosa se redujo al 8 al 12 % en semillas heterocigotas y al 5 % en semillas homocigotas, acompañado de organización celular anormal en la capa de aleurona y estructuras amorfas de granos de almidón.

Con los hallazgos de esta investigación, se espera aumentar el conocimiento en biotecnología utilizando herramientas de ingeniería metabólica y biología sintética, ya que son claves para afrontar los retos que plantean los Objetivos de Desarrollo Sostenible (ODS). El consumo y la producción responsables (ODS 12), la vida submarina (ODS 14) y el hambre cero (ODS 2) pueden abordarse mediante la creación de fuentes alternativas para la producción sostenible de escualeno para evitar la matanza incontrolada de tiburones y la mejora de los cultivos de arroz mediante la edición del genoma.

## Resum

Aquest treball de recerca es va concebre com a part d'un projecte científic més gran i complex que inclou l'aplicació d'eines d'enginyeria metabòlica i edició del genoma per introduir la via metabòlica de l'esqualè a l'arròs (*Oryza sativa*) i *E. coli* com a plataforma per a la producció d'altres compostos d'interès com ara els carotenoides i analitzar comparativament els dos models biològics en termes econòmics i ciència fonamental.

Concretament en arròs, el gran projecte consta de dues parts, la primera de les quals l'objectiu d'aquesta tesi. La primera part va consistir aconseguir el bloqueig dels gens clau de l'endosperm per a la biosíntesi del midó i observar-ne els efectes sobre la maquinària metabòlica de la planta; la segona part pretén aconseguir la inserció dels gens implicats en la síntesi d'esqualè al mateix target on es va produir el bloqueig, mitjançant CRISPR/Cas9.

Al capítol dos d'aquesta tesi (manuscrit en preparació) es van explorar tres enfocaments metodològics per a l'enginyeria metabòlica microbiana que condueixen a la producció d'esqualè mitjançant l'aplicació de diferents eines de biologia sintètica. Es va dissenyar i provar tres estratègies d'enginyeria genètica per ser aplicades a *E. coli*. 1) Inserció de gens heteròlegs de la via MEP i gens d'esqualè sintasa a *E. coli*; 2) Sobreexpressió de gens de la via MEP d'*E. coli* i gens heteròlegs d'esqualè sintasa a *E. coli*; 3) Expressió de gens heteròlegs d'esqualè sintasa *E. coli* dissenyat amb la via dual MEP/MVA. Amb l'estratègia 1 es va assolir la major productivitat instantània d'esqualè (3.8 mg/L/h), la qual es troba en la tercera posició de més productivitat reportada fins ara a la literatura. Aquest cep dissenyat podria tenir un ampli potencial per utilitzar-se en la producció d'esqualè a escala industrial.

El capítol tres va donar com a resultat el manuscrit *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*. En aquest estudi, utilitzem CRISPR/Cas9 per crear dos mutants heterozigòtics, un amb una AGPasa severament truncada i no funcional i l'altre amb una modificació estructural C-terminal que causa una pèrdua parcial d'activitat. Inesperadament, observem l'esgotament del midó a les fulles dels dos mutants i el corresponent augment en el nivell de sucres solubles. Això va reflectir l'expressió no anticipada de *OsAPL2* i *OsAPS2b* a les fulles, generant una AGPasa ectòpica completa al citosol del full, i una disminució corresponent a l'expressió de la petita subunitat plastidial *OsAPS2a* que només es va complementar parcialment amb un augment en l'expressió d'*OsAPS1*.

El capítol quatre va derivar en el manuscrit *CRISPR/Cas9 mutations in the rice Waxy/GBSSI gene induce allele-specific and zygosity-dependent feedback effects on endosperm starch biosynthesis*. En aquest treball, utilitzem CRISPR/Cas9 per introduir mutacions que afecten el locus *Waxy (Wx)* que codifica l'enzim granule-bound starch synthase I (GBSSI) a l'endosperm d'arròs. Trobem que les mutacions van disminuir però no van eliminar l'activitat de GBSS a les llavors a causa de la compensació parcial causada per la regulació positiva de GBSSII. L'activitat de GBSS als mutants va ser del 61 al 71 %

dels nivells de wild type, de manera similar a dos mutants creats amb radiació, però el contingut d'amilosa es va reduir al 8 al 12 % en llavors heterozigots i al 5 % en llavors homozigots, acompanyat d'organització cel·lular anormal a la capa d'aleurona i estructures amorfes de grans de midó.

Amb les troballes d'aquesta investigació, s'espera augmentar el coneixement en biotecnologia fent servir eines d'enginyeria metabòlica i biologia sintètica, ja que són claus per fer front als reptes que plantegen els Objectius de Desenvolupament Sostenible (ODS). El consum i la producció responsables (ODS 12), la vida submarina (ODS 14) i la fam zero (ODS 2) es poden abordar mitjançant la creació de fonts alternatives per a la producció sostenible d'esqualè per evitar la matança incontrolada de taurons i la millora dels cultius d'arròs mitjançant l'edició del genoma.

## Table of content

<b>Acknowledgements</b> .....	<b>i</b>
<b>Summary</b> .....	<b>iii</b>
<b>Resumen</b> .....	<b>v</b>
<b>Resum</b> .....	<b>vii</b>
<b>Index of figures</b> .....	<b>xiv</b>
<b>Index of tables</b> .....	<b>xviii</b>
<b>Abbreviations</b> .....	<b>xx</b>
<b>Outputs</b> .....	<b>xxvi</b>
<b>Chapter I. General Introduction</b> .....	<b>1</b>
1.1 General introduction.....	3
1.2 References .....	10
<b>Thesis aims and objectives</b> .....	<b>17</b>
<b>Chapter II: Metabolic engineering in microbes for squalene production</b> .....	<b>21</b>
2.0 Abstract.....	21
2.1 Introduction .....	22
2.2 Scope of this work.....	37
2.3 Materials and methods.....	39

2.3.1 Approach 1: Insertion of heterologous MEP pathway genes and squalene synthase genes in <i>E. coli</i> .....	40
2.3.1.1 Bacterial strains and reagents.....	40
2.3.1.2 Remotion of transit peptide signal .....	40
2.3.1.3 Plasmid preparation for bacterial transformation.....	42
2.3.1.4 PCR screening and clone sequencing.....	42
2.3.1.5 Determination of <i>E. coli</i> culture parameters .....	44
2.3.1.5 Shake flask culture .....	44
2.3.1.6 Quantitative Determination of Squalene.....	44
2.3.2 Approach 2: Overexpression of <i>E. coli</i> MEP pathway genes and heterologous squalene synthase genes in <i>E. Coli</i> . .....	45
2.3.2.1 Bacterial strains and reagents.....	45
2.3.2.2 Plasmid construction.....	45
2.3.2.3 PCR Primer design .....	47
2.3.2.4 Confirmation of Gene presence.....	48
2.3.2.5 Bacterial transformation and clone screening.....	49
2.3.2.6 Shake flask culture .....	49
2.3.3 Approach 3: Expression of heterologous squalene synthase genes in the <i>E. coli</i> strain engineered with the dual MEP/MVA pathway.....	49
2.3.3.1 Recovery of bacterial strain from a transferred biological sample material .....	49
2.3.3.2 Confirmation by PCR of the presence of the key genes .....	50
2.3.3.3 Transformation of the dual pathway strain with two squalene synthase genes .....	50
2.3.3.4 Shake flask culture .....	50
2.3.3.5 Quantitative Determination of Squalene.....	50
2.4 RESULTS.....	51
2.4.1 Approach 1: Heterologous expression of MEP pathway genes from plants and two squalene synthase genes in <i>E. coli</i> .....	51
2.4.1.1 Selection of genes .....	51
2.4.1.1 Truncation of <i>OsDXS</i> 1, 2, 3 and <i>IDI2</i> .....	51
2.4.1.2 Strain building .....	52
2.4.1.3 Characterization of squalene production in the engineered strains.....	60

2.4.2 Approach 2: Generation of <i>E. coli</i> strains overexpressing endogenous MEP-biosynthetic genes and heterologous SQS genes.....	62
2.4.2.1 Plasmid constructs .....	62
2.4.2.2 Squalene production in <i>E. coli</i> strain overexpressing MEP genes and expressing SQS .....	63
2.4.3 Approach 3: Expression of heterologous squalene synthase genes in the <i>E. coli</i> strain engineered with the dual MEP/MVA pathway.....	64
2.4.3.1 Generation of the Dual-hSQS3 and Dual-SQS3 strains .....	64
2.4.3.2 Characterization of squalene production from the engineered <i>E. coli</i> dual MEP/MVA pathway with SQS genes .....	64
2.4.3.3 Comparison of the squalene production in the three approaches .....	66
2.5 Discussion.....	68
2.6 Conclusions .....	77
2.7 Recommendations and prospects .....	77
2.8 References .....	78
<b>Chapter III. Genome editing in rice <i>APL2</i> gene using CRISPR/Cas9 .....</b>	<b>94</b>
3.0 Abstract.....	94
3.1 Introduction .....	94
3.2 Materials and methods.....	100
3.2.1 Target sites and sgRNA design .....	100
3.2.2 Vector construction.....	100
3.2.3 Rice transformation and recovery of transgenic plants.....	100
3.2.4 Confirmation of the presence of Cas9 and gRNA .....	100
3.2.5 Analysis of induced mutations .....	101
3.2.6 Protein structural modelling .....	102
3.2.7 Enzymatic activity and carbohydrate levels.....	103
3.2.8 RNA extraction and real-time qRT-PCR analysis .....	103
3.3 Results.....	103
3.3.1 Design of a CRISPR/Cas9 mutation strategy .....	103



3.3.2 Recovery and analysis of mutant lines.....	104
3.3.3 Structural comparisons .....	105
3.3.4 Analysis of AGPase and sucrose synthase activity .....	106
3.3.5 Analysis of AGPase family gene expression .....	107
3.3.6 Analysis of starch and sugar levels.....	108
3.4 Discussion.....	109
3.5 Conclusion.....	114
3.6 References .....	115
<b>Chapter IV. Genome editing in rice <i>Waxy</i> gene using CRISPR/CAS9 .....</b>	<b>124</b>
4.0 Abstract.....	124
4.1 Introduction .....	124
4.2 Materials and methods.....	127
4.2.1 Target sites and sgRNA design .....	127
4.2.2 Vector construction.....	127
4.2.3 Rice transformation and recovery of transgenic plants.....	128
4.2.4 Confirmation of the presence of cas9 and gRNA in regenerated rice plants .....	128
4.2.5 Analysis of induced mutations .....	128
4.2.5 Protein structural modeling and phylogenetic analysis.....	129
4.2.6 Enzymatic activity assays .....	129
4.2.7 Starch and soluble sugars.....	129
4.2.8 RNA extraction and real-time quantitative RT-PCR analysis.....	130
4.2.9 Seed phenotype and microscopy.....	130
4.2.10 Statistical analysis .....	131
4.2.11 Accession numbers.....	131
4.3 Results.....	132
4.3.1 Recovery and characterization of GBSSI mutants.....	132
4.3.2 Structural comparisons and phylogenetic analysis of protein sequence ....	133
4.3.3 Changes in enzymatic activity of GBSS by the loss of GBSSI activity .....	137

4.3.4 Deregulation of starch-related family gene expression induced by GBSSI mutations .....	139
4.3.5 Changes in starch, amylose and soluble sugar levels in GBSSI mutants.....	142
4.3.6 Phenotype and microscopy changes produced by GBSSI mutations .....	143
4.4 Discussion.....	145
4.5 Conclusions .....	149
4.6 References .....	150
<b>General discussion.....</b>	<b>155</b>
General discussion .....	157
References .....	158
<b>General conclusions .....</b>	<b>162</b>

## Index of figures

### Chapter I.

**Figure 1.** Analysis of searching peer-reviewed literature results by Scopus database showed as log 10 of the number of publications as a function of year of publication.

**Figure 2.** Number of publications on the SDGs and SynbioTech, reveals the gap between these two both topics that aims to remain constant over time. Source: Scopus database.

**Figure 3.** Research activities carried out during this thesis and their relationship with the objectives of sustainable development

### Chapter II.

**Figure 1.** Chemical structure of squalene; linear structure: double bond geometry and squalene in a coiled form.

**Figure 2.** US Squalene Market Volume (Tons) by Sectors, 2013 – 2024.

**Figure 3.** US squalene market, 2014-2022 (millions of dollars).

**Figure 4.** Natural sources of squalene in different organisms. Modified figures from Gohil et al., 2019.

**Figure 5.** Squalene biosynthesis pathways via MVA in plants, fungi, algae and yeast and via MEP in some bacteria (except *E. coli*).

**Figure 6.** Framework to test the three different approaches for squalene production in *E. coli*.

**Figure 7.** Parameters of the *E. coli* culture conditions.

**Figure 8.** A general overview of Gibson Assembly cloning method adapted from Gibson Assembly cloning guide (2017).

**Figure 9.** TP length prediction for A, *DXS 1*; B, *DXS 2*; C, *DXS 3*; D, *IDI 1*; E, *IDI 2*; F, *FPS*.

**Figure 10.** *E. coli* colony PCR for selected genes in the plasmid pET-32a(+).

**Figure 11.** Sanger sequencing alignment of *OsDXS1* with the corresponding GeneBank wild type sequence (*ApE: A plasmid Editor, 2022*).

**Figure 12.** Sanger sequencing alignment of *OsIDI1* with the corresponding GeneBank wild type sequence (*ApE: A plasmid Editor, 2022*).

**Figure 13.** Sanger sequencing alignment of *GIFPS* with the corresponding GeneBank wild type sequence (*ApE: A plasmid Editor, 2022*).

**Figure 14.** Sanger sequencing alignment of *HSQS* with the corresponding GeneBank wild type sequence (*ApE: A plasmid Editor, 2022*).

**Figure 15.** Sanger sequencing alignment of *TSQS* with the corresponding GeneBank wild type sequence (*ApE: A plasmid Editor, 2022*).

**Figure 16.** Confirmation by PCR of the genes inserted in the BL21(DE3) *E. coli* strain.

**Figure 17.** Squalene production in the approach 1 with the MEP-hSQS and MEP-tSQS *E. coli* strains at different growth conditions.

**Figure 18:** Colony PCR of the key genes inserted in the engineered *E. coli* BL21 (DE3) strain overexpressing MEP genes.

**Figure 19.** Squalene production in approach 2 from the engineered hSQS and tSQS strains (MEP-hSQS2 and MEP-tSQS2 stains) at different growth conditions.

**Figure 20.** PCR analysis of the dual pathway genes carried out in individual *E. coli* colonies.

**Figure 21.** Evaluation of the optimal IPTG concentration tested in three different *E. coli* strains containing the dual MEP/MVA pathway and a SQS gene.

**Figure 22.** Squalene production in dual MEP/MVA *HSQS* or *TSQS* at different growth conditions.

**Figure 23.** Squalene production of the *E. coli* strains tested with two different squalene synthases (human SQS and TSQS) designed through three different synthetic biology approaches.

**Figure 24.** Instant squalene productivity in mg/L/h of the different engineered *E. coli* strains obtained in this work compared to those reported in the literature.

### Chapter III.

**Figure 1.** The starch biosynthetic pathway and their coordination of different genes in rice.

**Figure 2.** Comparisons between wild type and mutant APL2 protein 3-D model structure.

**Figure 3.** Wild type and mutant comparisons of 3-D heterotetrameric structure.

**Figure 4.** gRNA sites of the CRISPR/Cas9 system and sequencing results in the rice mutant lines.

**Figure 5.** Expected APL2 protein sequences using ExpASy of the wild type and the mutant lines.

**Figure 6.** Enzymatic activity of AGPase and sucrose synthase activity in rice flag leaves.

**Figure 7.** Relative expression levels of rice AGPase family genes in the flag leaves of wild-type and mutants.

**Figure 8.** Starch and soluble sugar content in flag leaves of wild-type and mutants.

## Chapter IV

**Figure 1.** The coordination of different starch biosynthesis genes in rice (modified from Thitisaksakul et al. 2012).

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

**Figure 3.** The gRNA target sites and sequencing results showing the nature of the six *Wx* mutant obtained in this work.

**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.

**Figure 6.** Structure of GBSSI in wild-type rice and mutant line 1.

**Figure 7.** Sequence analysis of the GBSSI protein. Sequence alignment of GBSSI proteins from various monocot and dicots plants.

**Figure 8.** Enzyme activity in wild-type and mutant rice plants.

**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-changes in the expression of starch biosynthesis and degradation pathway genes in T2 seeds in wild-type and mutant rice plants.

**Figure 11.** Mean normalized expression correlating different plant tissues with the different starch pathway gene isoforms depending on gene family (Gene) and the enzymatic function (Gene\_Type).

**Figure 12.** Seed carbohydrate content in wild-type and mutant rice plants. (a) Total starch content of T2 seeds from wild-type (WT) plants, the *Wx* mutant lines and the two *Wx* irradiation mutants KUR and Musa.

**Figure 13.** Seed phenotypes of wild-type (WT) plants, the *Wx* mutant lines and the two *Wx* irradiation mutants KUR and Musa.

**Figure 14.** Optical microscopy 40x 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.

## Index of tables

### Chapter II

**Table 1.** Physical and chemical properties of squalene (Adapted from Spanova and Daum et al., 2011 and Naiziri et al., 2011 and Popa et al., 2015).

**Table 2.** Squalene content in different shark liver oils. (Adapted from Rosales-Garcia et al., 2017).

**Table 3.** Squalene content in some plants (Modified from Gohil et al., 2019).

**Table 4.** Squalene content in deodorizer distillates from some plants.

**Table 5.** Squalene production from wild type microorganisms. (Modified from Ghimire et al., 2016 and Naziri et al., 2011).

**Table 6.** Engineered microorganism strains with the highest squalene production reported in the literature (adapted and updated from Gohil et al., 2019).

**Table 7.** Best ranked microorganisms as a source of squalene (adapted and updated from Gohil et al., 2019).

**Table 8.** Bacterial strains and plasmids used in approach 1.

**Table 9.** Oligonucleotides used in approach 1.

**Table 10:** Primers used for sequencing analysis.

**Table 11.** Oligonucleotides used for Gibson Assembly.

**Table 12.** Genes encoding key enzymes involved in squalene biosynthesis and their function in the pathway.

**Table 13.** Vector constructs for MEP pathway overexpression

### Chapter III

**Table 1.** Primer sequences for RT-qPCR analysis.

### Chapter IV

**Table 1.** Characteristics of the six mutated lines generated in this study and the irradiation mutants KUR and Musa.

**Table 2.** Analysis of variance for normalized expression of different gene types, genes, isoforms and genotypes in different tissues on log-transformed data.



## Abbreviations

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
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
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
DXS	1-deoxy-D-xylulose-5-phosphate synthase
<i>E. coli</i>	<i>Escherichia coli</i>
F-6-P	Fructose-6-phosphate
FPP	Farnesyl diphosphate
FPS	Farnesyl-diphosphate synthase
<i>Gl</i>	<i>Gentiana lutea</i>
gRNA	guide RNA
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
hSQS	Human squalene synthase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	High-performance liquid chromatography
IDI	Isoprenyl diphosphate isomerase
ISA1	Isoamylase 1
ISA2	Isoamylase 2
ISA3	Isoamylase 3
IU	International unit
KUR	GBSS mutant strain

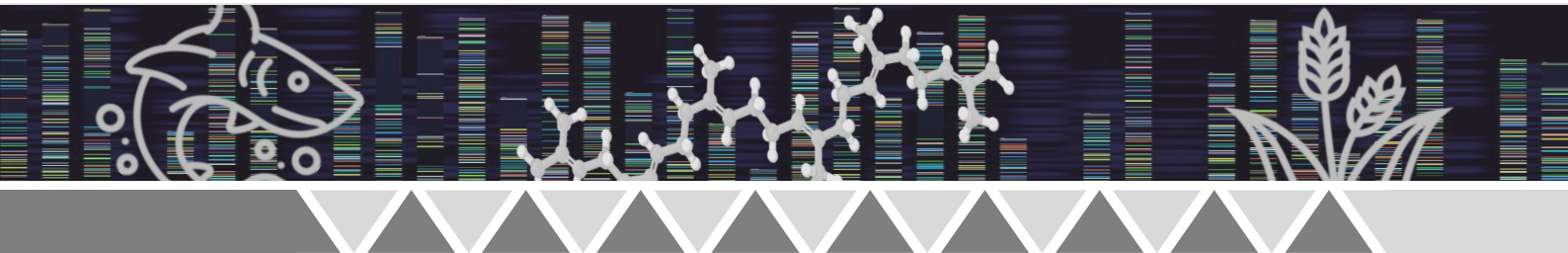
Lys	Lysine
MEP	2C-methylerythritol 4-phosphate
mev	Mevalonate
MgCl <sub>2</sub>	Magnesium Chloride
mRNA	Messenger RNA
MS	Murashige and Skoog medium
Musa	Musashimochi GBSS mutant strain
MVA	Mevalonic acid
MVK	Mevalonate kinase
NHEJ	Non-homologous end joining
N-terminal	Amino-terminal
Os	Oryza sativa
PAM	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
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>
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
TEM	Transmission Electron Microscopy
Thr	Threonine
tSQS	<i>Thermocynecoccus enlongatus</i> squalene syntase
Tyr	Tyrosine
UBQ5	Ubiquitin 5
UDP	Uridine diphosphate
UDP-Glc	UDP-glucose
Waxy	GBSSI
WR1	WRINKLED 1
WT	Wild-type
Wx	Waxy





# Outputs



## Outputs

### List of publications and manuscripts related to this thesis

**Soto, E.**, Pérez, L., Farré, G., Juanos, J., Villorbina, G., Bassie, L., Medina, V., Serrato, A. J., Sahrawy, M., Rojas, J. A., Romagosa, I., Muñoz, P., Zhu, C., & Christou, P. (2019). CRISPR/Cas9 mutations in the rice Waxy/GBSSI gene induce allele-specific and zygosity-dependent feedback effects on endosperm starch biosynthesis. *Plant cell reports*, 38(3), 417–433. <https://doi.org/10.1007/s00299-019-02388-z>

**Soto, E.**, Pérez, L., Villorbina, G., Bassie, L., Medina, V., Muñoz, P., Capell, T., Zhu, C., Christou, P., & Farré, G. (2018). CRISPR/Cas9-induced monoallelic mutations in the cytosolic AGPase large subunit gene *APL2* induce the ectopic expression of *APL2* and the corresponding small subunit gene *APS2b* in rice leaves. *Transgenic research*, 27(5), 423–439. <https://doi.org/10.1007/s11248-018-0089-7>

### Publications that I have contributed to during the time of my Ph.D. studies but are not included in the thesis.

Pérez, L., Alves, R., Perez-Fons, L., Albacete, A., Farré, G., **Soto, E.**, Vilaprinnyó, E., Martínez-Andújar, C., Basallo, O., Fraser, P.D. and Medina, V., (2022). Multilevel interactions between native and ectopic isoprenoid pathways affect global metabolism in rice. *Transgenic Research*, 31(2), pp.249-268. <https://pubmed.ncbi.nlm.nih.gov/35201538/>

Bortesi, L., Zhu, C., Zischewski, J., Perez, L., Bassié, L., Nadi, R., Forni, G., Lade, S. B., **Soto, E.**, Jin, X., Medina, V., Villorbina, G., Muñoz, P., Farré, G., Fischer, R., Twyman, R. M., Capell, T., Christou, P., & Schillberg, S. (2016). Patterns of CRISPR/Cas9 activity in plants, animals and microbes. *Plant biotechnology journal*, 14(12), 2203–2216. <https://doi.org/10.1111/pbi.12634>

### Dissemination at conferences

**Erika Soto**, Ludovic Bassie, Teresa Capell, Paul Christou, Gemma Villorbina. (2020). Towards squalene shark-free products in microbes: lessons learned from metabolic engineering strategies using synthetic biology tools. European Federation of Biotechnology. Biocatalysis Open Day, 26th of November, 2020.

**Erika Soto**, Ludovic Bassie, Teresa Capell, Paul Christou, Gemma Villorbina (2018). Oral presentation: 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. European Congress of Biotechnology, 1 to 4 July 2018 CICG, Geneva (Switzerland).

**Erika Soto**, Ludovic Bassie, Teresa Capell, Paul Christou, Gemma Villorbina (2017). Poster: CRISPR/Cas9 mutations of the rice APL2 gene. University of Lleida. II Research Conference at the UdL: The Doctorate as an Engine of Innovation, Lleida (Spain).

**Erika Soto**, Ludovic Bassie, Teresa Capell, Paul Christou, Gemma Villorbina (2016). Poster: Targeted mutagenesis of the rice waxy gene using CRISPR/Cas9. University of Lleida. I Research Conference at the UdL: The Doctorate as an Engine of Innovation, Lleida (Spain).

### **Awards and recognitions**

**Erika Soto**, Ludovic Bassie, Teresa Capell, Paul Christou, Gemma Villorbina (2019). Oral presentation: How Genome editing can save sharks. Winner in the contest “Thesis in three minutes”. Campus Iberus, 28 November 2019, Tudela (Spain).

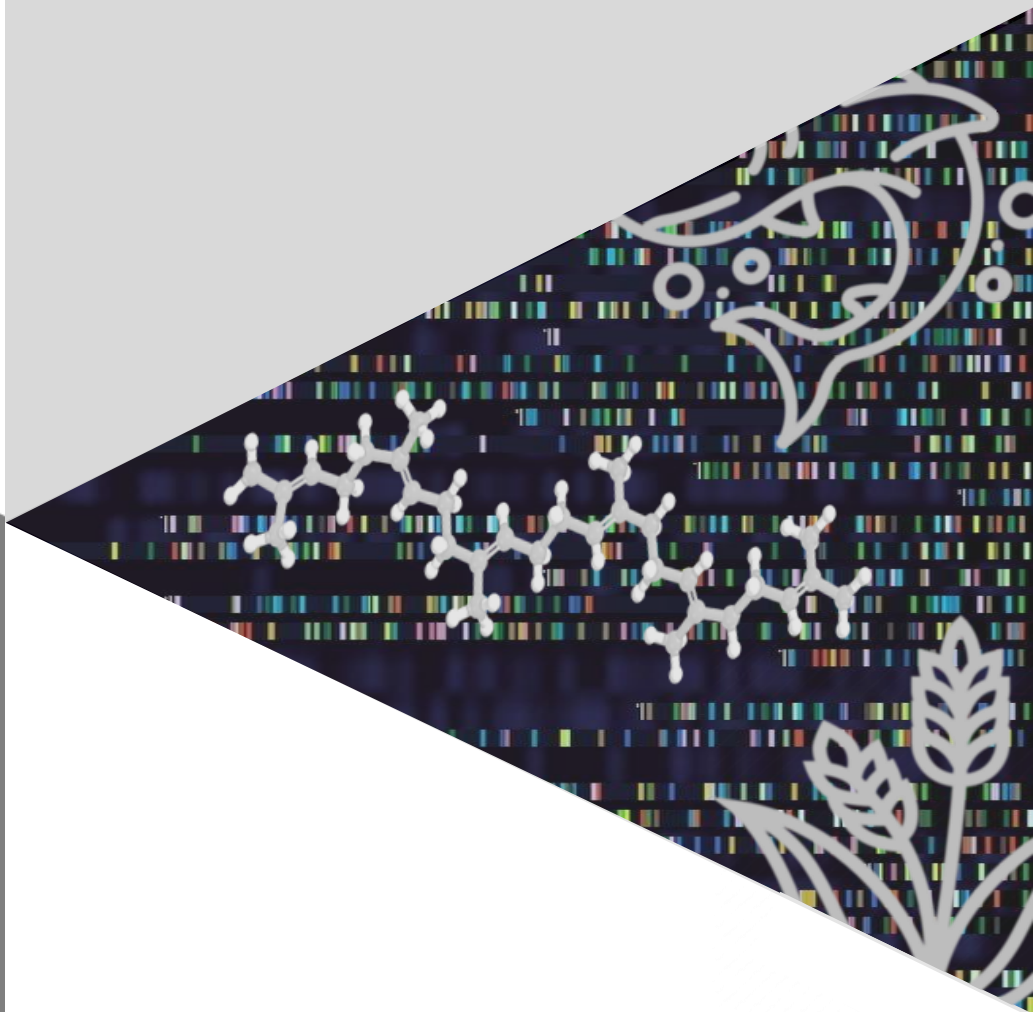
It was very important for me during the development of this thesis to participate in science communication events to point out important environmental problems, like shark’s slaughter. [https://youtu.be/gcn68eE\\_H4o?t=4705](https://youtu.be/gcn68eE_H4o?t=4705)







# General Introduction





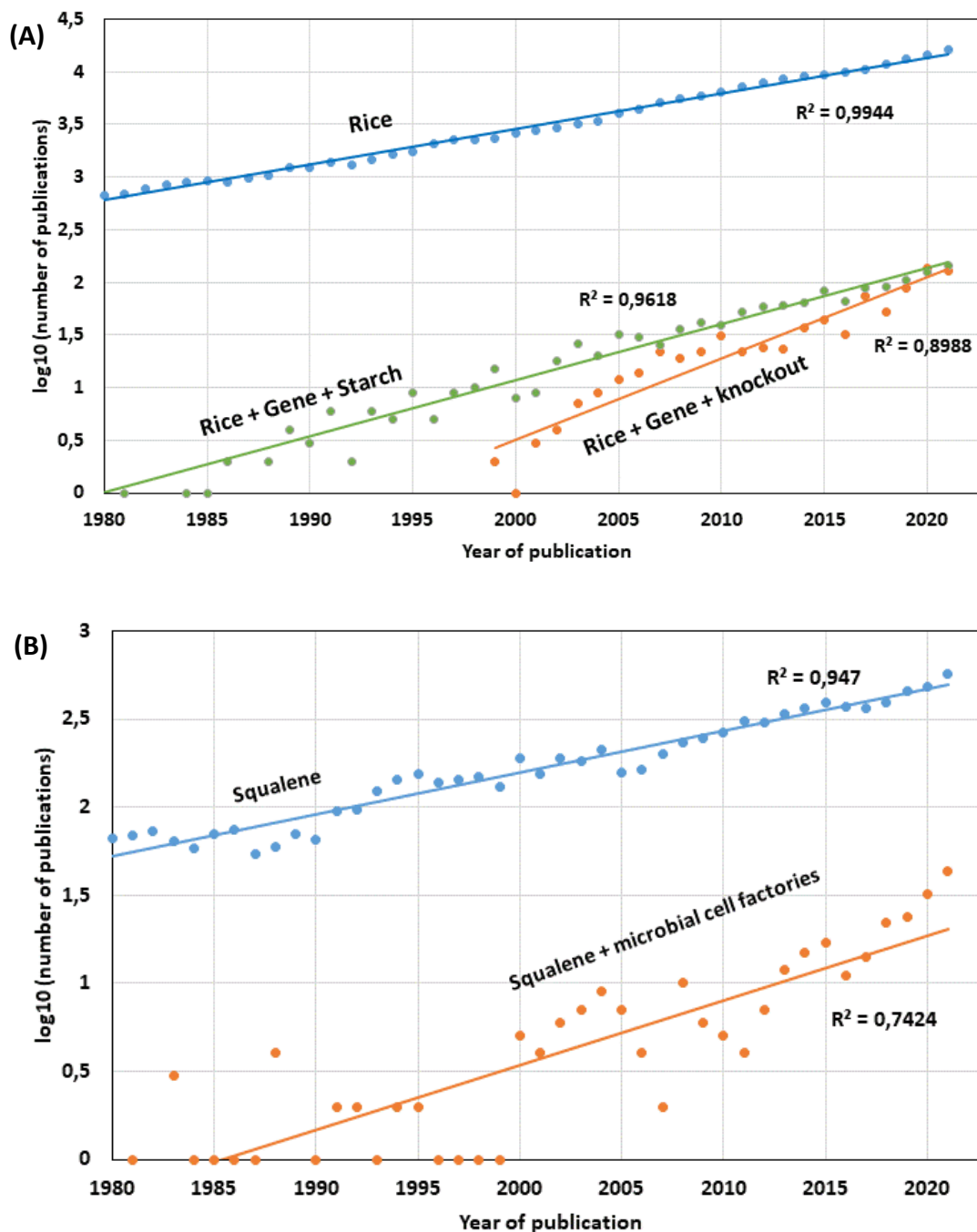
## 1.1 General introduction

### What is this thesis all about?

This thesis addresses the study of the new valuable materials production using the approach provided by advanced biological technologies such as genome editing and synthetic biology. In particular, two of these new valuable materials are discussed: on the one hand, the natural biosynthesis of a complex polymeric carbohydrate from the rice plant, starch, which is essential for human nutrition, is investigated; on the other hand, the artificial biosynthesis of a secondary terpenoid in *E. coli* bacteria, squalene, a compound of great interest for its pharmacological and cosmetic applications, is investigated. In the first case, mutant rice plants were created, genetically edited to knockout genes associated with starch synthesis, and the effect on the biochemistry of primary metabolism was studied, with the aim of understanding the response of the organism to this type of controlled genomic manipulation. In the second case, new forms of artificial life called engineered microorganisms were created, using *E. coli* cells as chassis, incorporating genes from other species, and different metabolic pathways for squalene biosynthesis were studied, with the aim of understanding the possibilities and limitations of this type of synthetic biology techniques for this case and hopefully, one day to produce squalene at a commercial level without having to slaughter sharks to achieve it.

You may be wondering, how are issues as disparate as the production of materials of commercial interest related to the use of cutting-edge technologies and environmental problems such as feeding the growing world population and the indiscriminate hunting of wild animals? Here is some technical information that will serve as a compass to guide you through this journey. Next, it is proposed an interpretation of the broad panorama in which this research is inserted.

## Rice and Squalene in the scientific literature



**Figure 1.** Analysis of searching peer-reviewed literature results by Scopus database showed as log 10 of the number of publications as a function of year of publication. (A) rice and (B) squalene scientific literature.

A searching in the peer-reviewed literature (**Figure 1**), shows that there are hundreds of thousands of publications in the last 40 years related to the study of rice (**Figure 1A**) and squalene (**Figure 1B**), the two materials that this thesis addresses. As can be seen, the linear trend of the data evidences the "exponential" growth of the interest caused by

these substances in the world scientific community. Of all this literature, a smaller part corresponds to research that addresses the issue of genes related to starch biosynthesis and its knockout.

The same can be seen in the case of squalene, where there is a more modest growth in research related to the production of this compound using microorganisms (microbial cell factories). However, it is still talking about hundreds of scientific articles, to which more and more reports on new advances and discoveries are added every day. It is not an easy task to navigate through all the currently available scientific literature and this work was no exception. The reader will find in this thesis a good amount of analyzed and cited literature, so you can delve into the aspects in which you want to deepen, understanding that today such a review of the literature can hardly claim to be exhaustive.

The case of the literature related to the squalene production is interesting. One of the first publications that deals with the subject is the one in which its purification and chemical identification were reported by extracting it from the liver of the shark (shark) (Tsujiyama, 1916). Much time has passed and, however, the extraction of squalene from shark liver oil is still a method used to produce this substance (Rosales-Garcia et al., 2017), reporting the scandalous number of 3000 sharks slaughtered to obtain a ton of the product (Ciriminna et al., 2014).

Alternative squalene production methods using microorganisms were first reviewed by (Popa et al., 2015). One of the most recent studies where the state of the art of this technology is reviewed by Paramasivan & Mutturi (2022). The use of microorganisms such as yeasts, fungi, bacteria and algae to produce chemical substances is as old as civilization itself. However, with current technology the genes of these we can be manipulated beings at will, temporarily or permanently altering their genome, overexpressing or silencing genes and/or inserting genes from other species (Liu C, et al., 2022; Montaño López et al., 2022). In this way, the concept of "microbial cell factories" emerged, the designed microorganisms that are used as biofactories to produce highly specialized chemical substances, which are currently being researched assiduously at a basic level (Gohil et al., 2021) and whose greatest commercial achievements already constitute a multimillion-dollar business (Kordi et al., 2022; Philippidis, 2021). To highlight, research is currently being carried out on the use of these synthetic biotechnologies for the production of cannabinoids (Luo et al., 2019), and on the creation of autotrophic engineered microorganisms (Gleizer et al., 2019) and last but not least, the creation of non-alcoholic beer aroma profiles (Melton, 2022; Dusséaux et al., 2020).

The case of the rice study is even more difficult to narrate. Not only is the number of publications much higher than for squalene, but also the studies on the improvement of the cereal are very varied (Hassan et al., 2022; Sukegawa et al., 2022). Of note are the studies that led to the creation and marketing of "Golden Rice", a rice obtained through genetic modifications to have greater nutritional value by incorporating carotenoids, which gave it a golden color, hence its name (Paine et al., 2005).

The complete rice genome was first published in 2006 for the International Rice Genome Sequencing project (Jackson, 2016). Among the first studies on genetic engineering to manipulate the biosynthesis of starch in rice is the study of Mizuno et al., (1993). One of the first investigations that proposed to improve the biosynthesis of carotenes in rice by controlling the role of squalene was the work of Furubayashi et al., (2014). The most recent studies on this subject, especially with the knockout of genes associated with starch biosynthesis, are those published by Xu et al., (2021). To highlight, research is currently being carried out on the use of these genome editing biotechnologies for the improvement of potato crops (Toinga-Villafuerte et al., 2022; Tussipkan & Manabayeva, 2021), the creation of crops with less recalcitrant and easy-to-value residual biomass (Velvizhi et al., 2021) and the creation of cellulosomes designed to its application in biorefineries (Šuchová et al., 2022).

In 2010, Craig Venter made the official announcement of the successful synthesis, assembly and transplantation of the artificial synthesized genome (*M. mycoides* JCVI-syn1.0) in a natural bacterial chassis (*M. capricolum*) creating a new self-replicate bacteria controlled by a synthetic genome (*M. mycoides*) (Gibson et al., 2010). This is one of the landmarks of Synthetic Biology, an emerging field that is broadly described as the design and construction of novel artificial or natural biological pathways, organisms or devices, in order to understand the basis of biological mechanisms as well as to create new and useful biological functions (Freemont et al., 2012). To differentiate the synthetic genome and the natural one, Venter's team included "watermarks" in the genome, among them a quotation from Richard Feynman "what I cannot build, I cannot understand". This thesis presents one of the first attempts in our research group in Applied Plant Biotechnology and the Centre for Biotechnological and Agrofood Developments (Universitat de Lleida) to understand life through re-build it.

### **Interpretation of the broader context of this research**

Human societies depend on materials that are valuable for their survival. The way in which these materials are obtained and mastered determines a large part of the relationship with the environment that surrounds them, with nature. Materials to feed, to dress, to shelter, to cure their illnesses, to adorn their body and their home, as well as materials to build endless tools. The ways in which these materials are produced is determined by the level of technological development, in other words, by the degree of integration/symbiosis between science, technology, the environment and society (Mauser et al., 2013).

It was only until the middle of the 20th century that it began to understand that anthropogenic action on the environment to obtain and produce materials has had a series of negative effects that significantly reduce the availability of natural resources, deteriorate the biosphere and therefore affect human health and put the survival of humanity and the future generations at risk, who would not have the material resources they will need (Barnosky et al., 2012) neither in quantity nor in quality. This environmental problem is independent of scale. Small populations can devastate their surrounding natural resources and pollute their environment to the point of stopping

their development and disappearing; in turn, the growth of the world population as well as the technologies that allow us to live in the global village show us that the entire planet already suffers from the consequences of the over-exploitation of natural resources, due to environmental pollution processes that go imaginary borderlines between countries and that attack nature in its fantastic and complex network of biophysical interactions. Phenomena such as global warming and climate change, whose most debated consequences have already been verified (Malla et al., 2022), are examples of what our current ways of being and living, as well as our current level of socio-technological development, have made damage to the planet.

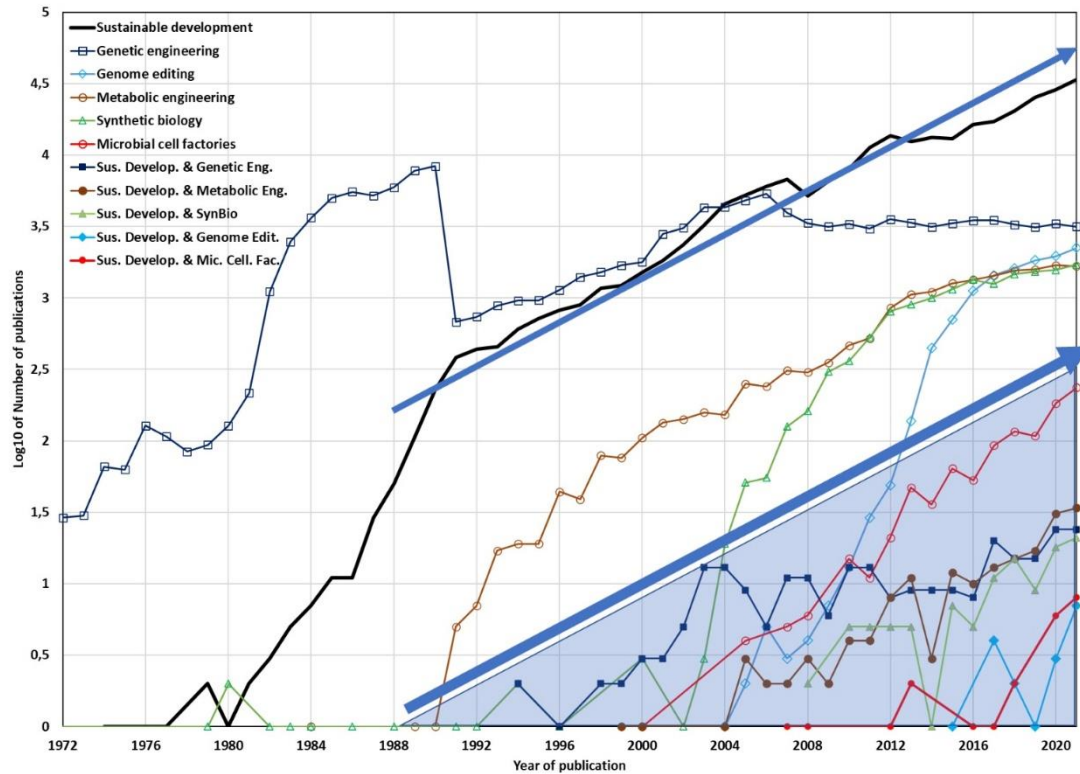
The complex relationship between science, technology and society is not a new topic. Its importance is, however, increasing (Mumford, 2010). Concepts such as sustainable development derive from re-conceptualizations about what is technology (the artificial), how it evolves and how it should be related to the natural environment in which it operates. Historically we are at a time when we can marvel at the problems we have solved and the technologies we have invented to do so; it is also a key moment to question the approach with which technologies can be developed and applied technologies, better anticipating the adverse effects they may have on ecosystems, on human in particular and on humanity in general. Whether the future development of humanity is to be called sustainable or otherwise is not relevant; the important thing is that societies find ways to use science and technology to protect life as a whole planet, to restore the environmental balance of the biosphere and allow future generations to live fully and continue advancing along the path of human evolution on planet earth and on other new worlds.

An extraordinary *corpus* of scientific knowledge is now available in all branches of knowledge, as well as unprecedented access to that knowledge through information and communication technologies. With this new and powerful scientific knowledge, a group of new technologies are being created that have the potential to offer alternative paths of technological development for humanity (Zhongming & Wei, 2019). These technologies can revolutionize the way we live in the same way that other technologies that we take for granted today did in their time. The democratization of scientific and technological knowledge (universal and free access to knowledge) as well as the possibility of making use of these new technologies (access to power), will be central elements for humanity to achieve the sustainable development goals that it has set for itself.

The 17 Sustainable Development Goals (SDG) were formulated in 2015 by the United Nations to create a reference framework that would save the planet from the most adverse effects of environmental deterioration, with the goal of being met by 2030. Some researchers have discussed the role of various key technologies that would contribute significantly to achieving these goals (World Economic Forum, 2021). It is considered that one of such technologies is the set of knowledge that today could be grouped under the name of synthetic biotechnology or SynBioTech (which includes genetic engineering, metabolic engineering, genome editing, synthetic biology and



microbial cell factory technology) (Gallup et al., 2021). In **Figure 2**, it is shown the number of publications on the SDGs and SynbioTech, the graph reveals the gap between these two both topics that aims to remain constant over time



**Figure 2.** Number of publications on the SDGs and SynbioTech. Source: Scopus database.

Of the 17 SDGs, the ones that it is believed can have the greatest impact due to the application of SBTs are: (2) Zero Hunger, (3) Good Health and Well-being, (6) Clean Water and Sanitation, (7) Affordable and Clean Energy, (12) Responsible Consumption and Production, (13) Climate action, (14) Life Below Water, (15) Life on Land. SBTs have the potential to transform food production with GMOs with examples such as Golden Rice (Paine et al., 2005), the creation of new and better drugs for personalized medicine (Courdavault et al., 2020), the treatment of wastewater and solid waste such as plastics using designed microorganisms (Lu et al., 2022), biofuels (Liu Z, et al., 2022), protect endangered animal species by replacing indiscriminate hunting with engineered microorganisms fermentation, such as the case of sharks to obtain squalene (Mendes et al., 2022) or the protection of natural genomes with controlled genetic modifications in salmon fish farming (Fedoroff et al., 2022), as well as the substitution of illicit crops of plant species for fermentations with engineered microorganisms (Galanie et al., 2015). I believe that we can use these technologies to conserve nature and contribute to the solution of many environmental conflicts around the world. During the completion of this doctoral thesis, various scientific topics that It is observed related to the SDGs were discussed, as illustrated in **Figure 3**.



**Figure 3.** Research activities carried out during this thesis and their relationship with the objectives of sustainable development. The work with yeasts to produce alkaloids using enzyme engineering techniques was developed during the international internship at the Technical University of Denmark DTU, at the Novo Nordisk Foundation Center for Biosustainability, participating in the MIAMi project: Refactoring monoterpene indole alkaloid production in microbial cell factories, details of which are not presented here due to confidentiality agreements.

However, although the number of publications on the SDGs continues its impressive exponential growth, as can be seen in Figure 2, the same does not happen with the research on SynbioTech that explicitly points to compliance with the SDGs, which separately and as areas of scientific knowledge grow exponentially but have a lower growth rate and a much lower volume of publications than studies by the SDGs, which reveals a gap between both topics that aims to remain constant over time. This can be verified in various recent studies that analyze the role of key enabling technologies (KET) in compliance with the SDGs (Laibach et al., 2022; Mabkhot et al., 2021).

I strongly believe that bridging this gap with more research that explicitly aims to use SynBioTech to meet the SDGs I have mentioned will allow us to begin to see the changes towards sustainability that we as humanity need to ensure our survival without compromising our chances of survival of our future generations. My conviction that this is necessary and possible has served as a motivation for me to carry out this work. I consider that the results that the reader will find in this thesis are like small pieces of a

great puzzle that we must solve: a sustainable development of humanity with technologies based on the power of life.

## 1.2 References

- Barnosky, A. D., Hadly, E. A., Bascompte, J., Berlow, E. L., Brown, J. H., Fortelius, M., & Smith, A. B. (2012). Approaching a state shift in Earth's biosphere. *Nature*, *486*(7401), 52-58.
- Ciriminna, R., Pandarus, V., Béland, F., & Pagliaro, M. (2014). Catalytic hydrogenation of squalene to squalane. *Organic Process Research & Development*, *18*, 1110-1115.
- Courdavault, V., O'Connor, S. E., Oudin, A., Besseau, S., & Papon, N. (2020). Towards the microbial production of plant-derived anticancer drugs. *Trends in Cancer*, *6*(6), 444-448.
- Dusséaux, S., Wajn, W. T., Liu, Y., Ignea, C., & Kampranis, S. C. (2020). Transforming yeast peroxisomes into microfactories for the efficient production of high-value isoprenoids. *Proceedings of the National Academy of Sciences*, *117*(50), 31789-31799.
- Fedoroff, N., Benfey, T., Giddings, L. V., Jackson, J., Lichatowich, J., Lovejoy, T., & Williams, R. N. (2022). Biotechnology can help us save the genetic heritage of salmon and other aquatic species. *Proceedings of the National Academy of Sciences*, *119*(19), e2202184119.
- Freemont, P. S., & Kitney, R. I. (Eds.). (2012). *Synthetic Biology-A Primer*. World Scientific Publishing Company.
- Furubayashi, M., Li, L., Katabami, A., Saito, K., & Umeno, D. (2014). Construction of carotenoid biosynthetic pathways using squalene synthase. *FEBS letters*, *588*(3), 436-442.
- Galanie, S., Thodey, K., Trenchard, I. J., Filsinger Interrante, M., & Smolke, C. D. (2015). Complete biosynthesis of opioids in yeast. *Science*, *349*(6252), 1095-1100.
- Gallup, O., Ming, H., & Ellis, T. (2021). Ten future challenges for synthetic biology. *Engineering Biology*, *5*(3), 51-59.
- Gibson, D. G., Glass, J. I., Lartigue, C., Noskov, V. N., Chuang, R. Y., Algire, M. A., & Venter, J. C. (2010). Creation of a bacterial cell controlled by a chemically synthesized genome. *Science*, *329*, 52-56.
- Gleizer, S., Ben-Nissan, R., Bar-On, Y. M., Antonovsky, N., Noor, E., Zohar, Y., & Milo, R. (2019). Conversion of *Escherichia coli* to generate all biomass carbon from CO<sub>2</sub>. *Cell*, *179*(6), 1255-1263.
- Gohil, N., Bhattacharjee, G., & Singh, V. (2021). An introduction to microbial cell factories for production of biomolecules. In *Microbial cell factories engineering for production of biomolecules* (pp. 1-19). Academic Press.

- Hassan, A., Shahzad, A. N., & Qureshi, M. K. (2022). Rice Production and Crop Improvement Through Breeding and Biotechnology. In *Modern Techniques of Rice Crop Production* (pp. 605-627). Springer, Singapore.
- Jackson, S. A. (2016). Rice: the first crop genome. *Rice*, 9(1), 1-3.
- Kordi, M., Salami, R., Bolouri, P., Delangiz, N., Asgari Lajayer, B., & van Hullebusch, E. D. (2022). White biotechnology and the production of bio-products. *Systems Microbiology and Biomanufacturing*, 1-17.
- Laibach, N., & Bröring, S. (2022). The Emergence of Genome Editing—Innovation Network Dynamics of Academic Publications, Patents, and Business Activities. *Frontiers in bioengineering and biotechnology*, 556.
- Liu, C. L., Xue, K., Yang, Y., Liu, X., Li, Y., Lee, & Tan, T. (2022). Metabolic engineering strategies for sesquiterpene production in microorganism. *Critical Reviews in Biotechnology*, 42(1), 73-92.
- Liu, Z., Wang, J., & Nielsen, J. (2022). Yeast synthetic biology advances biofuel production. *Current opinion in microbiology*, 65, 33-39.
- Lu, H., Diaz, D. J., Czarnecki, N. J., Zhu, C., Kim, W., Shroff, R., & Alper, H. S. (2022). Machine learning-aided engineering of hydrolases for PET depolymerization. *Nature*, 604(7907), 662-667.
- Luo, X., Reiter, M. A., d'Espaux, L., Wong, J., Denby, C. M., Lechner, A., & Keasling, J. D. (2019). Complete biosynthesis of cannabinoids and their unnatural analogues in yeast. *Nature*, 567(7746), 123-126.
- Mabkhot, M. M., Ferreira, P., Maffei, A., Podrżaj, P., Mądział, M., Antonelli, D., & Lohse, N. (2021). Mapping industry 4.0 enabling technologies into united nations sustainability development goals. *Sustainability*, 13(5), 2560.
- Malla, F. A., Mushtaq, A., Bandh, S. A., Qayoom, I., & Hoang, A. T. (2022). Understanding climate change: scientific opinion and public perspective. In *Climate Change* (pp. 1-20). Springer, Cham.
- Mausser, W., Klepper, G., Rice, M., Schmalzbauer, B. S., Hackmann, H., Leemans, R., & Moore, H. (2013). Transdisciplinary global change research: the co-creation of knowledge for sustainability. *Current opinion in environmental sustainability*, 5(3-4), 420-431.
- Melton, L. (2022). Synbio salvages alcohol-free beer. *Nature biotechnology*, 40(1), 8.
- Mendes, A., Azevedo-Silva, J., & Fernandes, J. C. (2022). From Sharks to Yeasts: Squalene in the Development of Vaccine Adjuvants. *Pharmaceuticals*, 15(3), 265.

- Mizuno, K., Kawasaki, T., Shimada, H., Satoh, H., Kobayashi, E., Okumura, S., & Baba, T. (1993). Alteration of the structural properties of starch components by the lack of an isoform of starch branching enzyme in rice seeds. *Journal of Biological Chemistry*, 268(25), 19084-19091.
- Montaño López, J., Duran, L., & Avalos, J. L. (2022). Physiological limitations and opportunities in microbial metabolic engineering. *Nature Reviews Microbiology*, 20(1), 35-48.
- Mumford, L. (2010). *Technics and civilization*. University of Chicago Press.
- Paine, J. A., Shipton, C. A., Chaggar, S., Howells, R. M., Kennedy, M. J., Vernon, G., & Drake, R. (2005). Improving the nutritional value of Golden Rice through increased provitamin A content. *Nature biotechnology*, 23(4), 482-487.
- Paramasivan, K., & Mutturi, S. (2022). Recent advances in the microbial production of squalene. *World Journal of Microbiology and Biotechnology*, 38(5), 1-21.
- Philippidis, A. (2021, July 2). Top 10 Synthetic Biology Companies. Genetic Engineering & Biotechnology News (GEN). Retrieved June 6, 2022, from <https://www.genengnews.com/topics/genome-editing/top-10-synthetic-biology-companies/>
- Popa, O., Băbeanu, N. E., Popa, I., Niță, S., & Dinu-Pârvu, C. E. (2015). Methods for obtaining and determination of squalene from natural sources. *BioMed research international*, 2015.
- Rosales-García, T., Jimenez-Martinez, C., & Dávila-Ortiz, G. (2017). Squalene extraction: biological sources and extraction methods. *International Journal of Environment, Agriculture and Biotechnology*, 2, 238838.
- Šuchová, K., Fehér, C., Ravn, J. L., Bedř, S., Biely, P., & Geijer, C. (2022). Cellulose-and xylan-degrading yeasts: Enzymes, applications and biotechnological potential. *Biotechnology Advances*, 107981.
- Sukegawa, S., Toki, S., & Saika, H. (2022). Genome Editing Technology and Its Application to Metabolic Engineering in Rice. *Rice*, 15(1), 1-10.
- Toinga-Villafuerte, S., Vales, M. I., Awika, J. M., & Rathore, K. S. (2022). CRISPR/Cas9-mediated mutagenesis of the granule-bound starch synthase gene in the potato variety Yukon Gold to obtain amylose-free starch in tubers. *International Journal of Molecular Sciences*, 23(9), 4640.
- Tsujimoto, M. (1916). A highly unsaturated hydrocarbon in shark liver oil. *Industrial & Engineering Chemistry*, 8, 889-896.
- Tussipkan, D., & Manabayeva, S. A. (2021). Employing CRISPR/Cas Technology for the Improvement of Potato and Other Tuber Crops. *Frontiers in Plant Science*, 12.

Velvizhi, G., Goswami, C., Shetti, N. P., Ahmad, E., Pant, K. K., & Aminabhavi, T. M. (2022). Valorisation of lignocellulosic biomass to value-added products: Paving the pathway towards low-carbon footprint. *Fuel*, 313, 122678.

Xu, Y., Lin, Q., Li, X., Wang, F., Chen, Z., Wang, J., & Gao, C. (2021). Fine-tuning the amylose content of rice by precise base editing of the *Wx* gene. *Plant biotechnology journal*, 19(1), 11.

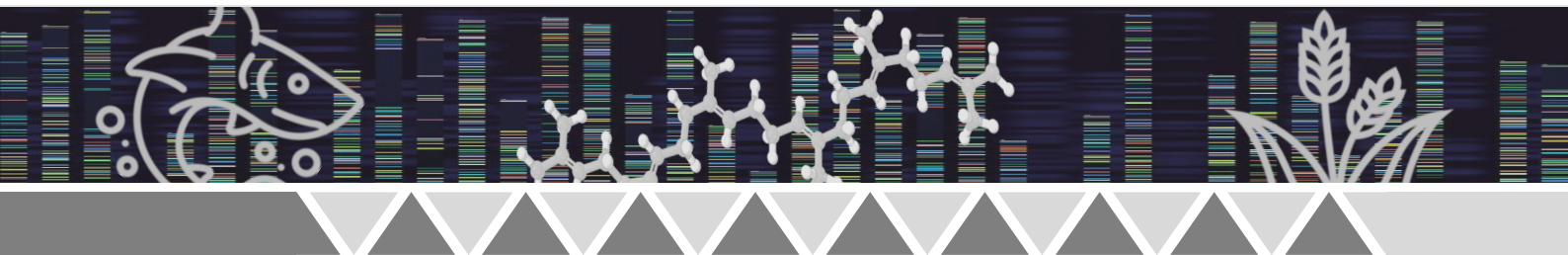
World Economic Forum. Why tech will be key in our quest to hit the SDGs. (2021, December 7). Retrieved May 5, 2022, from <https://www.weforum.org/agenda/2019/09/technology-global-goals-sustainable-development-sdgs/>

Zhongming, Z., & Wei, L. (2019). The future is now: Science for achieving sustainable development. Global Sustainable Development report.





# Aims and Objectives







## Thesis aim and objectives

### Aim

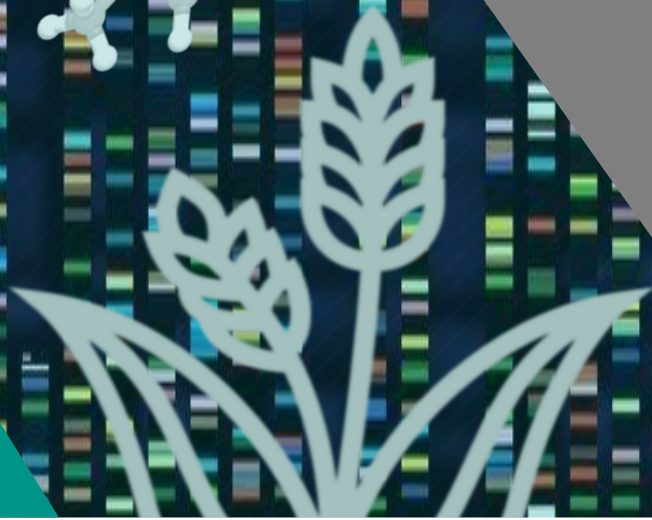
The overall aim of this thesis was to explore different approaches for the production of novel valuable materials using metabolic engineering and synthetic biology tools.

### Objectives

1. To engineer *E. coli* strains using metabolic engineering strategies
2. To compare different approaches of engineering *E. coli* strains for squalene production
3. To test the effects of the CRISPR/Cas9 induced mutations of *APL2* gene, encoding the first enzyme of the starch metabolism, in the rice plants.
4. To introduce CRISPR/Cas9 mutations and determine how they affect the *Waxy* (*Wx*) locus encoding granule-bound starch synthase I (GBSSI) in rice endosperm.



**Metabolic engineering in  
*Escherichia coli* for squalene  
production**





## Chapter II: Metabolic engineering in *E. coli* for squalene production

### 2.0 Abstract

Squalene is a chemical substance discovered in 1916 in the shark liver oil (Tsujiimoto M. 1916) that has found important applications in the pharmaceutical, cosmetic and food markets. It has been used since 1985 as an adjuvant in vaccines (Allison and Byars 1986), and recently have gained increased attention for its use in numerous vaccines against COVID-19 (Zhang et al., 2020; Peng et al., 2020).

Increasing demand of squalene in the last decades has become a problem for wildlife conservation because squalene is primarily obtained from shark liver oil (Tsoi et al., 2016; Ciriminna et al., 2014). If no other sources of production are considered, soon it could lead to the extinction of different shark species. With this concern, the synthesis and production of this compound in other organisms, particularly microbes, is being studied. The present contribution is aimed to explore three methodological approaches for the microbial metabolic engineering leading to the production of squalene by the application of different synthetic biology tools.

Three genetic engineering strategies were designed and tested to be applied to *E. coli*. 1) Insertion of heterologous MEP pathway genes and squalene synthase genes in *E. coli*; 2) Overexpression of *E. coli* MEP pathway genes and heterologous squalene synthase genes in *E. Coli*; 3) Expression of heterologous squalene synthase genes in the *E. coli* strain engineered with the dual MEP/MVA pathway

The highest squalene production was achieved with strategy 1, which resulted in 90 and 43 mg/L with *HSQS* and *TSQS*, respectively. The other strategies showed promising results, but they need to be optimized; otherwise, they can result in low squalene yield or to cell death. Metabolic engineering using synthetic biology toolkits can help to create new sustainable products and at the same time, to protect marine life, as established by the commitment to Sustainable Development Goals claims.

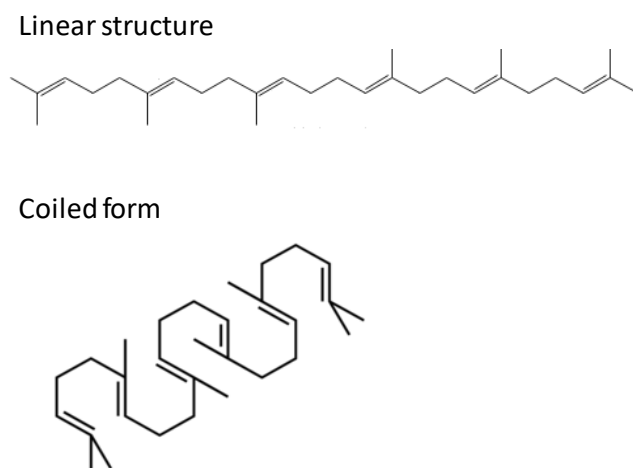
## 2.1 Introduction

Squalene is a linear triterpene, derived from lipid metabolism, widely distributed in the bacteria, fungi, algae, plants, and animals (Ghimire et al., 2016). They are formed through MEP (methylerythritol phosphate) pathway which mainly occurs in bacteria and the plastids of photosynthetic organisms (Ghimire et al., 2016; Kuzuyama T. 2002) and MVA (mevalonic acid) biosynthetic pathway found in plants, animals, and fungi (Spanova and Daum 2011; Popa et al., 2015; Rani et al., 2018). This compound has an essential role in lipid metabolism and consequently is a key intermediate in the formation of complex secondary metabolites like sterols, hopanoids and terpenes. Also, squalene is a molecule with very interesting physicochemical properties for the pharmaceutical, food and cosmetics industry (Naziri et al., 2011). In the last decade, the use of squalene from shark liver oil has increased exponentially due to its great demand in the market, which has triggered an environmental problem (Tsoi et al., 2016; Ciriminna et al., 2014). Consequently, if no other sources of production are considered, it could lead to the extinction of different shark species. To search for solutions, several research groups have studied the synthesis and production of this molecule in other organisms (Katabami et al., 2015; Wei et al. 2018; Huang et al., 2018; Choi et al., 2016; Han et al., 2018; Liu et al., 2020).

In the last decade, novel microbial platforms have been developed through synthetic biology and genome engineering tools to increase the production of high-added value compounds such as squalene, in the so-called microbial cell factories (Kleasling J.D 2010; Shepelin et al 2018; Wang et al 2018). Despite the impressive results obtained at the laboratory scale, these alternatives have not yet been able to reach commercial manufacturing due to the technical issues to scale-up the process at industrial conditions (Gohil et al., 2019). Nevertheless, instead of using shark liver oil and vegetable oil as the conventional source of squalene, microbial cell factories are the most promising current technology for sustainable production of this and many other speciality chemicals. This chapter describes several approaches experimentally tested for squalene production in microbial cell factories using genome editing tools and discuss some of the several challenges that the technology is facing to go commercial application.

### Physical properties

Squalene (2,6,10,15,19,23-hexamethyl-2,6,10,14,18,22-tetracosahexane) is a polyunsaturated lipid with a molecular formula  $C_{30}H_{50}$  and molecular weight of 410.73 g/mol. Its chemical structure consists of 6 double bonds that allow it to adopt different conformations, such as linear or coiled for, as shown in **Figure 1** (Xu et al., 2016).



**Figure 1.** Chemical structure of squalene; linear structure: double bond geometry and squalene in a coiled form (Xu et al., 2016).

Squalene is a very lipophilic molecule, with a non-polar nature, and together with the high surface tension it presents (Spanova and Daum et al., 2011), it makes it an excellent emulsifier for hydrophobic compounds. Some physical and chemical properties of squalene are shown in **Table 1**.

**Table 1.** Physical and chemical properties of squalene (Adapted from Spanova and Daum et al., 2011 and Naiziri et al., 2011 and Popa et al., 2015).

Properties	Value
Molecular weight	410.7 g/mol
Octanol/water partitioning coefficient (log P)	10.67
Solubility in water	0.124 mg/L
Viscosity	~11 cP
Surface tension	~32 mN/m
Density	0.858 g/mL
Melting point	-75°C
Boiling point	285°C
Refractive index	1.499
Iodine number	381 g/100g
Heat of combustion	10.773 cal/g

### Industrial interest

New applications of this lipid are being discovered in the medical and pharmaceutical sector, which prove to be beneficial for health. Its emulsifying power is used in a large number of vaccines and medicines because it decreases the release rate of the active components, improving their absorption (Kim et al., 2003; Gopakumar 2012; Popa et al., 2015).

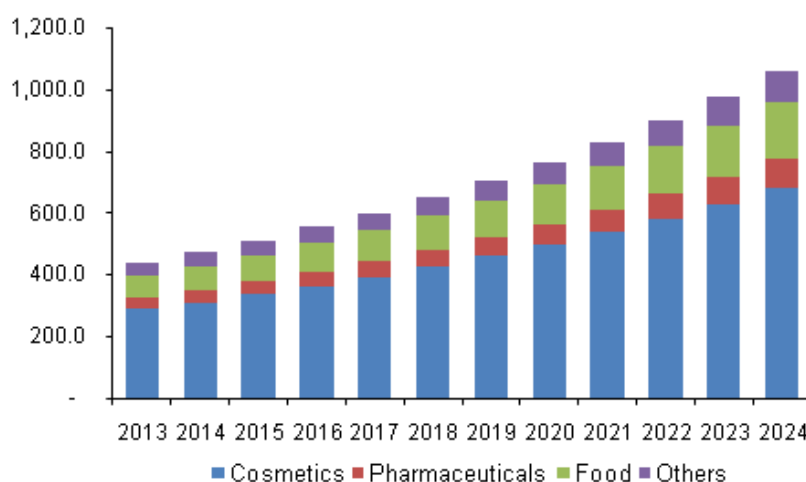


Several studies have also demonstrated the chemopreventive activity of squalene (Budiyanto et al., 2000; Ronco and De Stéfani, 2013; Xu et al., 2016). There are two specific mechanisms: directly such as the inhibition of oncoproteins or the elimination of free radicals; and indirectly by increasing the efficacy of anticancer drugs (Das et al., 2003; Reddy and Couvreur 2009). This molecule also offers protection against coronary heart diseases, due to the ability to reduce cholesterol levels in the blood (Lou-Bonafonte et al., 2018; Bhilwade et al., 2010; Strandberg et al., 1990).

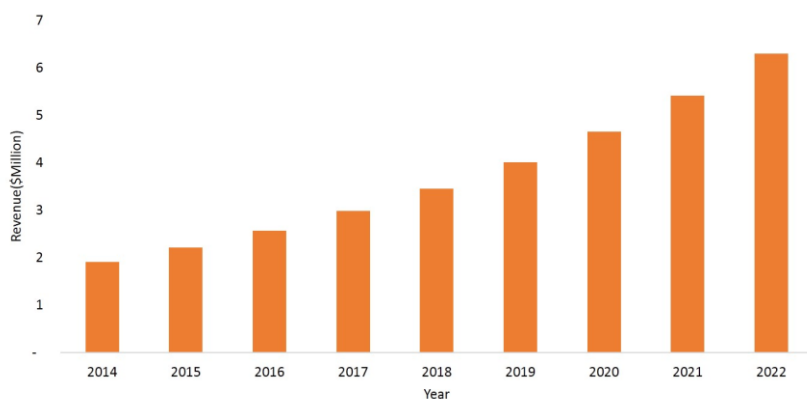
Due to squalene is a molecule with antioxidant properties, showing high scavenging activity towards ROS (Reactive oxygen species), is also used in the food sector as a supplement and functional food to increase the nutritional value of different products (Lou-Bonafonte et al., 2018).

In the cosmetic sector, its emollient and moisturizing properties, along with its compatibility with lipids on the skin's surface, stimulate its use in personal care products, sunscreen and anti-ageing creams (Naziri et al., 2011, Güneş 2013, Huang 2009).

**Figure 2** shows the evolution and projection of the volume of the US squalene market by sectors from 2013 to 2024. The squalene world market is experiencing significant growth and, according to estimates made, for 2023 in the United States there will be a demand for 700 tons of this product and, as indicated in **Figure 3**, will involve a revenue of more than 6 million dollars (Prasad and Roy 2016).



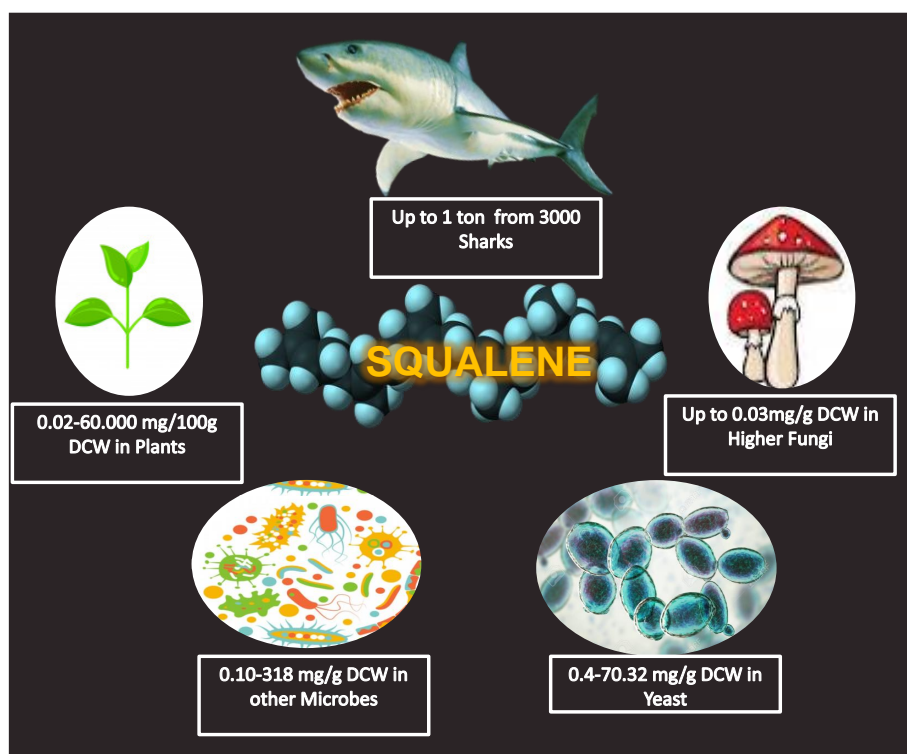
**Figure 2.** US Squalene Market Volume (Tons) by Sectors, 2013 - 2024 (Prasad and Roy 2016)



**Figure 3.** US squalene market, 2014-2022 (millions of dollars) (Prasad and Roy 2016).

### Natural sources of squalene

The large commercial and industrial applications of squalene due to its properties, mentioned above, make this compound arouse the interest of companies and research groups to find natural sources for its extraction. The sources of squalene in different organisms, such as fungi, plants, yeast and deep-sea sharks are shown in **Figure 4**.
















**Figure 4.** Natural sources of squalene in different organisms. Modified figures from Gohil et al., 2019.

## Squalene from sharks

The main source of squalene is from the liver oil of several species of sharks, especially those that live at depths below 400 m (**Table 2**) (Rosales-Garcia et al., 2017). In fact, the name "squalene" came from the family taxonomy rank of sharks, Squalidae. This compound was discovered by the Japanese researcher Mitsumaru Tsujimoto in 1906, when he achieved the separation of the unsaponifiable fraction from the shark liver oil (Tsujimoto, 1916). Although this animal is a great producer, to get a ton of squalene, it is necessary to kill 3000 sharks (333 g/shark) (Ciriminna et al., 2014). An uncontrolled slaughter should be carried out to obtain 700 tons of squalene, which would cause a decrease in biodiversity, severely damaging the environment and even the very viability of obtaining it in the future (Naziri et al., 2011; Popa et al., 2015; Rosales-Garcia et al., 2017).

Moreover, taking into account that these animals possess a long reproductive cycle and slow growth, many of these species are close to extinction (Rosales-Garcia et al., 2017, Popa et al., 2015, Gohil et al., 2019). Also, it is interesting to note that the purification process of squalene is inconvenient, because this molecule shares similarities with other lipid compounds that exist in shark liver oil and for obtaining a purity of >98% it is required to perform a single distillation phase under vacuum at temperatures of 200-230°C (Gohil et al., 2019).

**Table 2.** Squalene content in different shark liver oils. (Adapted from Rosales-Garcia et al., 2017)

Shark specie	Squalene content in the liver (%)	Phenotype
<i>Centroscymnus crepidater</i>	35.7-59.4	
<i>Centroscymnus owstoni</i>	37.1-53.1	
<i>Centroscymnus coelolepis</i>	31.1-47.1	
<i>Deania calcea</i>	43.4-66.1	
<i>Etmopterus baxteri</i>	14.3-51.5	
<i>Etmopterus sp. nov.</i>	20.8	
<i>Dalatias licha</i>	43.4	
<i>Centrophorus squamosus</i>	79.6	
<i>Centroscymnus plunketi</i>	0.9	
<i>Etmopterus granulosus</i>	50.3-60.5	
<i>Deania calcea</i>	69.6	
<i>Centroscymnus crepidater</i>	73	
<i>Centrophorus squamosus</i>	65.5	

## Squalene from plants

For several years, alternative sources of production that do not involve the slaughter of animals have been studied. In some reports, it has been seen that it can also be obtained from the saponifiable fraction of different vegetable oils (**Table 3**), among which amaranth oil stands out. However, although amaranth has the highest concentration of squalene from all reported plants (600 g/kg), (Wejnerowska et al., 2013) it has not been used in industry. The explanation is that the lipid content of its seeds is low (4.8-8.1%), compared to the lipid content from olive (6.67-26.67%) (United States Department of Agriculture, 2018), which is the only source that is currently used for commercial purposes. Despite this, the squalene production using plants is still insufficient to cover the global demand.

**Table 3.** Squalene content in some plants (Modified from Gohil et al., 2019)

Plant source	Concentration (mg/100g DCW)
Amaranth	60.000
Olive	99-1.245
Ginseng seed	514-569
Pumping seed	523
Rice bran	320
Brazil nut	146
Peanuts	133
White sesame seed	61
Black sesame seed	57
Palm	20-50

In addition, deodorizer distillates which come from the refinement of plant oils, are rich in compounds with biological activity such as tocopherols, fatty acids, sterols and squalene has become a more efficient alternative compared to the performance of raw plant oils (Popa et al., 2015; Sherazi and Mahesar, 2016). To give a concrete example, the squalene content in olive oil is about 0.99-12.45 g/kg, (Giacometti and Milin, 2001), in contrast, the deodorizing distillates of olive oil contain 100-300 g/kg of squalene (Naziri et al., 2011). In a similar context, the deodorizing distillates from soybean, sunflower, canola, and palm contain 18-55, 43-45, 30-35, and 2-13 g/kg of squalene, respectively (Dumont and Narine, 2007; Naziri et al., 2011; Naz et al., 2014). The deodorizer distillates from some plants are shown in **Table 4**.

**Table 4.** Squalene content in deodorizer distillates from some plants.

Distillates	Concentration (mg/100g DCW)
Olive oil	10.000-30.000
Soybean oil	5.500
Sunflower oil	4.300-4.500
Canola oil	3.000-3.500
Palm fatty acid	200-1.300
Wine less	6.000

Even though plants have good squalene content, they are not considered an appropriate source of this compound. This because they depend on environmental, agronomic and climatic factors that can vary the squalene content. Also, many of these plants require a laborious and time-consuming cultivation process. For these reasons, the plants cannot meet the current demand for squalene in the markets.

### Squalene from microorganisms

In the literature it has been reported, that some microorganisms are capable of producing a high amount of squalene (Ghimire et al., 2016), such as *Pseudozyma sp.* JCC207 and *Aurantiochytrium sp.* 18W-13a. (**Table 5**). However, applying genetic engineering to them is complicated due to lack of information of their metabolic pathways. That is why it was proposed to choose model microorganisms such as *Saccharomyces cerevisiae* or *Escherichia coli*, which are characterized by having a rapid growth and by synthesizing low amount of squalene (in the case of *S. cerevisiae*), to genetically engineer them for enhancing its production.

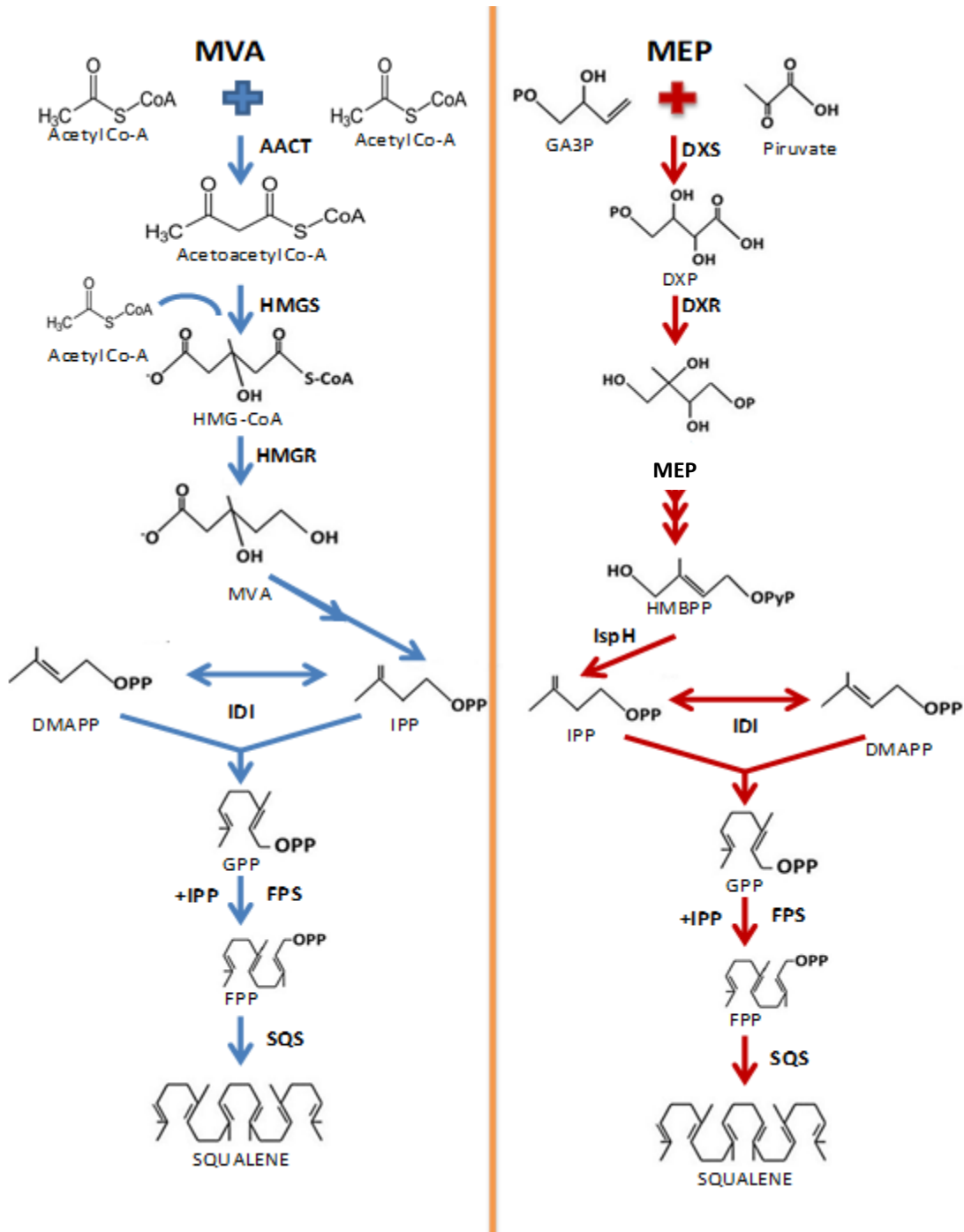
**Table 5.** Squalene production from wild type microorganisms. (Modified from Ghimire et al., 2016 and Naziri et al., 2011)

Group	Microorganism	Squalene Yield
Unicellular eukaryotic protists (Marine heterotrophic organism)	<i>Aurantiochytrium sp.</i> Yonez 5-1	318 mg/g
	<i>Aurantiochytrium sp.</i> 18W-13a	198 mg/g
	<i>Aurantiochytrium sp.</i>	5.9 mg/L
Marine bacteria	<i>Aurantiochytrium sp.</i> BR-MP4-A1	0.57 mg/g
	<i>Aurantiochytrium mangrovei</i> FB3	0.37 mg/g
	<i>Schizochytrium mangrovei</i>	1.31 mg/L
Green algae	<i>Rubritalea squalenifaciens sp. nov.</i>	15 mg/g
	<i>Chlamydomonas sp.</i>	1.1 mg/g
Fungi (yeast)	<i>Torulaspora delbrueckii</i>	0.24 mg/g
	<i>Pseudozyma sp.</i> JCC207	70.32 mg/g
	<i>Kluyveromyces lactis</i>	30 mg/L
	<i>Saccharomyces cerevisiae</i> (baker's yeast)	0.04 mg/g
	<i>Saccharomyces cerevisiae</i> (BY4741)	2.97 mg/L
	<i>Saccharomyces cerevisiae</i> (EGY48)	3.13 mg/L

### Biosynthetic pathways of squalene

In order to increase the production of squalene in microorganisms, it is essential to study the metabolic pathways where it is formed. These are the mevalonic acid pathway (MVA) and the methylerythritol phosphate pathway (MEP) that give rise to the compounds isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate or dimethylallyl diphosphate (DMAPP), both precursors of squalene and others isoprenoids (**Figure 5**).

The union of these two molecules through the action of a transferase forms geranylpyrophosphate (GPP) and, subsequently, with the incorporation of another IPP unit, farnesyl pyrophosphate (FPP) is obtained, through the action of the enzyme farnesyl pyrophosphate synthase (FPS). This molecule has a 12-carbon chain with three methyl groups, three double bonds, and a pyrophosphate radical. Thanks to the action of the enzyme known as squalene synthase (SQS), two FPPs fuse at the ends, losing the pyrophosphate groups, and giving rise to squalene, both FPS and SQS enzymes are not present in *E. coli*. From here, numerous studies prove that, by introducing specific genes into *E. coli* or by combining the two routes mentioned above, the production of squalene can be achieved.



**Figure 5.** Squalene biosynthesis pathways via MVA in plants, fungi, algae and yeast and via MEP in some bacteria (except *E. coli*).

### Engineering microorganisms for squalene production

Microorganisms such as *S. cerevisiae* and *E. coli* have advantages over other organisms such as plants, animals or higher fungi. Among these advantages, it can be highlighted that they have rapid growth; for example, they duplicate their growth in a short time, they are easy to grow and can develop under different growth conditions. Also, their genomes are well characterized, which facilitates their genetic manipulation. As a



matter of fact, for the advantages mentioned above, these microorganisms have been used successfully in the industry for the production of valuable products (biofuels, chemicals, secondary metabolites from plants) as microbial cell factories. (Rabinovitch-Deere et al., 2013; Singh, 2014, 2016; Chubukov et al., 2016; Gohil et al., 2017).

Additionally, with the development of areas such as molecular biology, synthetic biology, and metabolic engineering, it has been possible to understand, manipulate, and modulate the metabolic pathways and their flux to increase the production of desired metabolites, such as squalene (Gohil et al., 2019). In the literature, different strategies have been reported, alone or in combination, for trying to increase the squalene content in different strains of microorganisms, among which stand out *S. cerevisiae* and *E. coli* (**Table 6**). These strategies are mentioned below:

- 1) Genetic modification: Introduction of the heterologous gene(s) involved in squalene production, overexpression of key genes of the pathway and knock out and/or down-regulation of downstream genes from the pathway that allows squalene accumulation (Bunch and Harris, 1986; Singh et al., 2017, 2018; Valachovič and Hapala, 2017).
- 2) Fed-batch fermentation with optimized conditions (Chen et al., 2010; Fan et al., 2010; Nakazawa et al., 2012).
- 3) Use of an inhibitor (e.g., terbinafine) that blocks the competitive pathway thus allowing squalene accumulation (Han et al., 2018).
- 4) Dual modulation of cytoplasmic and peroxisomal pathway engineering and two-stage fed-batch fermentation by hybrid *S. cerevisiae* strains (Liu et al., 2020).
- 5) Extension of cell membrane to overcome the storage limitation of *E. coli* for successful squalene production (Meng et al., 2020).
- 6) A combinatorial strategy of cytoplasmic and mitochondrial engineering to alleviate the metabolic burden caused by the compartmentalized MVA pathway in mitochondria and improve cell growth in *S. cerevisiae* (Zhu et al., 2021).

**Table 6.** Engineered microorganism strains with the highest squalene production reported in the literature (adapted and updated from Gohil et al., 2019).

Strain	Strategy	Yield mg/L	References
<i>Saccharomyces cerevisiae</i> BY4741	Overexpression of <i>ERG9</i> and <i>tHMG1R</i> , insertion mutation in <i>ERG1</i>	270	Zhuang and Chappell, 2015
<i>S. cerevisiae</i> INVSc1	Overexpression of <i>tHMG1</i> , <i>IDI1</i> , <i>ERG20</i> , <i>ERG9</i> , <i>ERG10</i> (encoding acetyl-CoA C-acetyltransferase), <i>ERG13</i> (HMG-CoA synthase), <i>ERG12</i> (mevalonate kinase), <i>ERG8</i> (phosphomevalonate kinase), and <i>MVD1</i> (diphosphomevalonate decarboxylase)	304.49	Rasool et al., 2016
<i>S. cerevisiae</i> D452-2	Overexpression of <i>tHMG1</i> and <i>DGA1</i> , fed-batch fermentation in nitrogen restricted minimal media	445.6	Wei et al., 2018
<i>S. cerevisiae</i> SR7	Co-expression of <i>tHMG1</i> and <i>ERG10</i> gene in xylose-rich medium	532	Kwak et al., 2017
<i>S. cerevisiae</i> Y2805	Overexpression of <i>tHMG1</i> , expression of <i>ispA</i> , fed-batch fermentation	1026	Han et al., 2018
<i>S. cerevisiae</i> Y2805	Overexpression of <i>tHMG1</i> , expression of <i>ispA</i> , fed-batch fermentation with supplementation of terbinafine	2011	Han et al., 2018
<i>S. cerevisiae</i> SquCP2	Dual modulation of cytoplasmic and peroxisomal engineering and two-stage fed-batch fermentation	11000	Liu et al., 2020
<i>S. cerevisiae</i>	Compartmentalized MVA pathway in mitochondria, combinatorial cytoplasmic and mitochondria engineering and two-stage fed-batch fermentation	21100	Zhu et al., 2021
<i>E. coli</i> XL1-Blue	Expression of <i>hSQS</i>	2.7	Furubayashi et al., 2014

Strain	Strategy	Yield mg/L	References
<i>E. coli</i> BL21(DE3)	Expression of <i>hopA</i> and <i>hopB</i> (squalene/phytoene synthases) together with <i>hopD</i> (farnesyl diphosphate synthase) from <i>Streptomyces peucetius</i>	4.1	Ghimire et al., 2009
<i>E. coli</i> BL21(DE3)	Overexpression of <i>dxs</i> and <i>idi</i> (rate-limiting enzymes), expression of <i>hopA</i> and <i>hopB</i> together with <i>hopD</i> from <i>Streptomyces peucetius</i>	11.8	Ghimire et al., 2009
<i>E. coli</i> XL1-Blue	Co-expression of <i>hSQS</i> , chimeric mevalonate pathway containing <i>tHMGR</i> , <i>ERG13</i> (hydroxymethylglutaryl-CoA synthase), <i>ERG12</i> (mevalonate kinase), <i>ERG8</i> (phosphomevalonate kinase) and <i>MVD1</i> (mevalonate diphosphate decarboxylase) from <i>S. cerevisiae</i> , overexpression of <i>atoB</i> (acetyl-CoA acetyltransferase), <i>idi</i> (isoprenyl diphosphate isomerise) and <i>ispA</i> (farnesyl diphosphate synthase)	230	Katabami et al., 2015
<i>E. coli</i> XL1-Blue	Overexpression of the entire MVA pathway and IPP enzymes	32	Choi et al., 2019
<i>E. coli</i> DH5 $\alpha$	Extension of cell membrane to overcome the storage limitation by overexpression of some membrane proteins	612	Meng et al., 2020

Regarding *S. cerevisiae*, several studies had demonstrated with promising data, an improvement of squalene production in engineered strains. Due that yeast is considered a best-studied eukaryotic experimental model organism, is inexpensive to grow it and manipulate it genetically; *S. cerevisiae* is also called “a swiss army knife” for the versatility of its use in biotechnology (Pretorius 2017).

On the other hand, *E. coli* has the advantage of not producing triterpenes endogenously which means that squalene is converted into only three key products such as, cholesterol, hopanoids and ergosterol and this allows to control the downstream metabolic routes, blocking the competitive metabolites (Katabami et al., 2015).

In the most specific cases of squalene production in *E. coli*, recent studies showed the development of engineered strains via different strategies generated by plasmid

transformation such as by inserting heterologous genes or by overexpressing endogenous MEP genes or by co-expressing genes in combination from both MEP and MVA pathways. For example, in 2009 Ghimire et al., inserted and expressed three different hopanoid genes from *Streptomyces peucetius* in *Escherichia coli*; *hopA* and *hopB* encoding for squalene synthases and *hopD* encoding a farnesyl diphosphate synthase. The engineered strain produced 4.1 mg/mL of squalene.

In the same study, the yield of squalene was elevated up to 11.8 mg/L by overexpressing *hopA*, *hopB*, *hopD* and also two additional genes, *DXS* and *IDI* (that encode rate limiting enzymes). Later, in 2015 Pan et al., carried out a similar study by introducing hopanoid genes, *hpnC*, *hpnD*, and *hpnE* from *Zymomonas mobilis* and *Rhodospseudomonas palustris* in *Escherichia coli*. This study reveals a new pathway for squalene biosynthesis in bacteria where the three enzymes HpnC, HpnD, HpnE together catalyse the reactions from farnesyl diphosphate to squalene. The resulting modified *E. coli* strain could be promising in squalene production, but its yield has not yet been tested. In 2015 Katabami et al. co-expressed a chimeric MEP/MVA pathway in a combination of the human squalene synthase gene from *Homo sapiens* or *Thermosynechococcus elongatus* in *E. coli*, obtaining up to 230 mg/L of squalene in the flask culture. More recently, Meng et al. 2020, engineered a *E. coli* membrane to overcome the limitation of the squalene storage, obtaining a yield of 612 mg/L in flask culture.

### **Industrial scale-up of squalene production using engineering microbes**

As far as we know, there are no reports about squalene production at industrial scale in the literature. As mentioned by Gohil et al. (2019) in the review about squalene production using engineering strategies in microorganisms, the “scale up problem” still remains as an open issue in fermentation biotechnology. These authors claim that “as far as both plant and microbial sources are concerned, productivity still remains a major issue for scale-up, thereby, limiting their use as sources in industrial and biomedical applications”. So, productivity (mass of squalene obtained per quantity of biomass used) is one of several key factors that must be taken into account. On the other hand, following the argument of Gohil et al. (2019), “in order to elevate the industrial scale production of valuable products, fermentation technology uses the adaptability of the natural pathways to produce a desired molecule from the organisms by utilizing cheaper sources of substrates”, which points out towards fermentation conditions (titer, cell concentration, substrate type and concentration, etc.) as a key second factor for successful scale up.

But, why scale-up is important? According to Crater & Lievens (2018), “the financial investment to scale up a microbial process to manufacturing scale is usually greater than the cost to develop the production microbe and lab-scale process. This can be on the order of US \$100 million to \$1 billion, including intermediate process validation (pilot

and demo scales) and construction and start-up of the manufacturing plant. The annual operating cost of the manufacturing plant is on the same order. The time required to transition from lab-scale to manufacturing is typically 3–10 years. Under these circumstances, the financial risk is high, so deterioration in process performance during scale-up will be costly and disruptive, potentially even leading to project failure. Short of failure, even incremental (5–10%) under-performance and/or delays (3–12 months) during scale-up will substantially reduce financial returns to investors and undermine stakeholder and customer confidence”.

### **Estimation of the bioreactor volume required for a squalene manufacture facility to produce 0.1% of the global market demand**

GD: Global demand (ton/year) = 267000 tons in 2014 (Rosales-Garcia et al., 2017)

GD1: 0.1% of global demand (kg/h) = 30.5

BP: Bioreactor productivity (kg/L\*h)  $\approx 6.5 \cdot 10^{-5}$  (Liu et al., 2020, see **Table 7**)

BVind: Bioreactor volume (m<sup>3</sup>) = GD1/BP  $\approx 471$

SF: Scale Factor = BVind / BVlab = 471000 / 3 = 157000

Note that Amyris uses six 200000 L bioreactors for commercial production of semi-synthetic artemisinin, and of  $\beta$ -farnesene, an isoprenoid produced by fermentation of sugar (Hill et al., 2020). In order to find the real bioreactor volume and the best operation mode, squalene biochemical kinetics and mass balances are required.

In the present investigation, squalene synthesis experiments were performed using genetically designed microorganisms in reaction volumes of 250 and 500 mL. Scaling up the process, as discussed above, is beyond the scope of this thesis.

**Table 7.** Best ranked microorganisms as a source of squalene (adapted and updated from Gohil et al., 2019)

Microorganism	Yield (mg/g DCW)	Titer (mg/L)	Fermentation volume/mode	Reference
<i>Aurantiochytrium sp.</i> Yonez 5–1 (WT)	317.74	1073.66	ND	Nakazawa et al., 2014
<i>Aurantiochytrium sp.</i> 18 W-13a (WT)	198	1290	ND	Kaya et al., 2011
<i>Schizochytrium mangrovei</i> PQ6 (FO)	33.00	991,65	15 L	Hoang et al., 2014
<i>Schizochytrium mangrovei</i> PQ6 (FO)	33.04	1019,28	100 L	Hoang et al., 2014
<i>Aurantiochytrium sp.</i> strain 18W-13a (FO)	171	900	200 mL	Nakazawa et al., 2012
<i>S. cerevisiae</i> BY4741, Overexpression of tHMG1 and POS5 with mitochondrial presequence (EM)	58.6	28.4	50 mL*	Paramasivan and Mutturi, 2017
<i>S. cerevisiae</i> Y2805, Overexpression of tHMG1, expression of <i>ispA</i> , fed-batch fermentation with supplementation of terbinafine (EM)	ND	2011	5 L	Han et al., 2018
<i>Yarrowia lipolytica</i> yeast HLYaliS02 overexpressing native <i>HMG-CoA</i> (3-hydroxy-3-methylglutaryl-CoA) reductase (EM)	32.60	502.75	ND	Liu et al., 2020
<i>S. cerevisiae</i> SquCP1, dual cytoplasmic-peroxisomal engineering (EM and FO)	350	11000	3 L	Liu et al., 2020
<i>E. coli</i> , MVA pathway and <i>S. cerevisiae</i> squalene synthase harnessed, additional Tsr expression to extending membrane volume for squalene storage (EM)	ND*	612	200 mL*	Meng et al., 2020

- ND: no data available. \*Values estimated from the same publication; WT: wild type; FO: fermentation optimization; EM: Engineered microorganism.

## 2.2 Scope of this work

Several hypotheses were conceived for the development of this research: based on what has been reported in the specialized literature, it is speculated that the production of squalene must be strongly influenced by the metabolic stress (burden) experienced by the microorganism according to the genetic load that is incorporated. For this reason, the possibility of testing at least three genetic engineering approaches was conceived.

Firstly, since that the insertion of some plant MEP pathway genes, such as DXS, *IDI* and *FPS*, for carotenoid biosynthesis have already been proven in our lab (Jin et al., 2020; 2021), it was considered as hypothesis that, the heterologous expression of this genes from rice (*Oriza sativa*) and *Gentiana lutea* (*OsDXS*, *OsIDI* and *GIFPS*) could represent an effective way for squalene synthesis in *E. coli*.

Secondly, it was hypothesized that the overexpression of the native genes of the MEP pathway in *E. coli* up to the production of IPP (isopentenyl diphosphate), combined with the heterologous expression of the genes encoding the enzymes FPS (farnesyl pyrophosphate synthase) and SQS (squalene synthase), could reach significant levels of squalene in *E. coli*.

Thirdly, the use of a dual MEP/MVA metabolic pathway has been previously proven in *E. coli* to be efficient for isoprene production in the work of Yang et al., (2016) as well as by Katabami et al., (2015) by using a chimeric MEP/MVA pathway in the production of squalene. In this way, the hypothesis was stated that a double MEP/MVA pathway could also favor the synthesis of squalene, since in this case some genes are inserted into the bacterial genome, which could lead to greater stability of squalene gene expression over time.

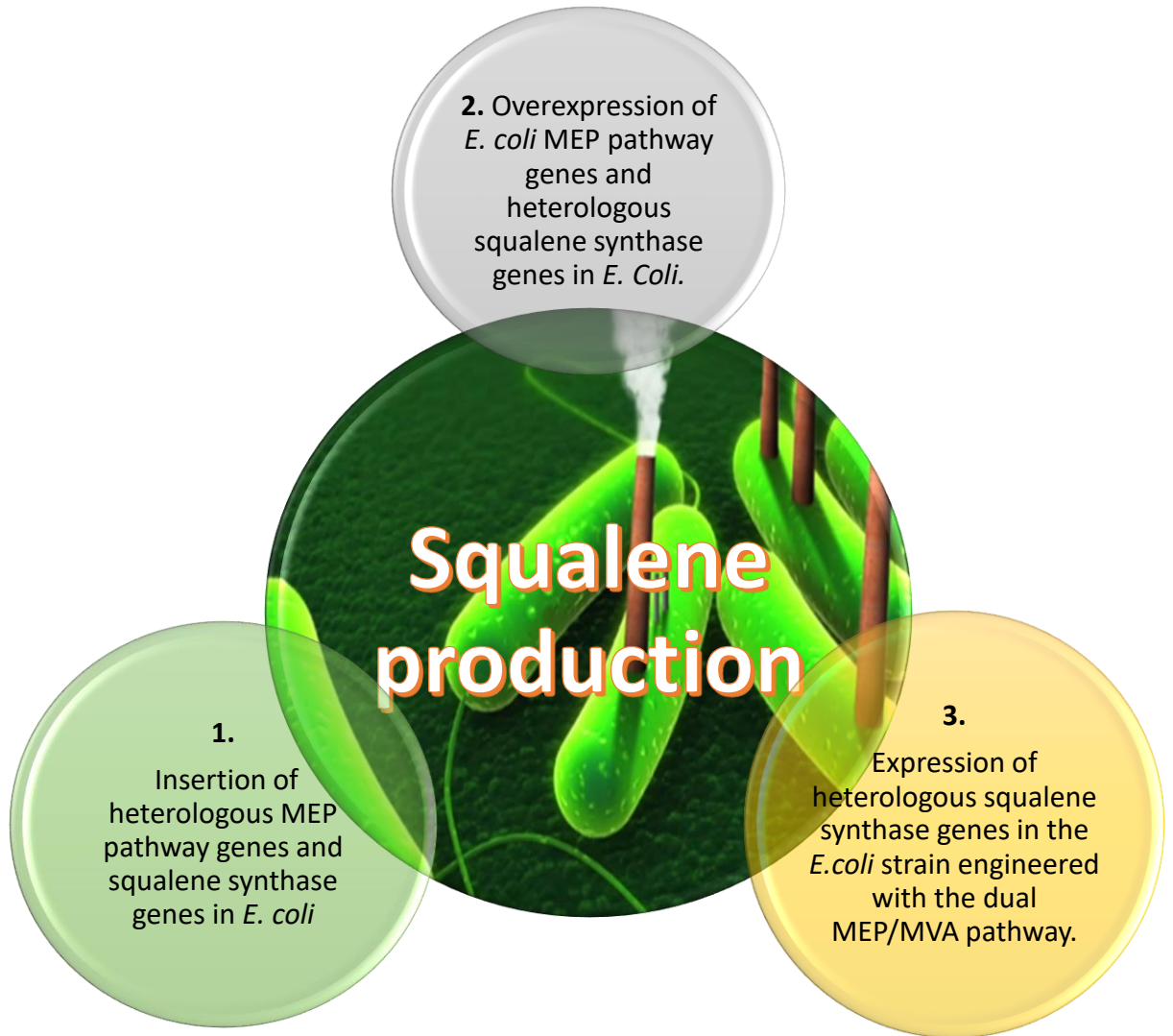
From the above, three metabolic engineering strategies were proposed as follows:

(1) Insertion of heterologous MEP pathway genes and squalene synthase genes in *E. coli*.  
(2) Overexpression of *E. coli* MEP pathway genes and heterologous squalene synthase genes in *E. Coli* and (3) Expression of heterologous squalene synthase genes in the *E. coli* strain engineered with the dual MEP/MVA pathway. Two genes of squalene synthase (from bacteria and human obtained from Katabami et al., 2015) were tested separately in the already engineered dual pathway (a combination of two metabolic pathways, MEP and MVA) in *E. coli* strains that were obtained from Yang et al., 2016 (China).

The major aim of the work described in this chapter was to explore three methodological approaches for the microbial metabolic engineering leading to the production of squalene, a valuable terpenoid compound, by the application of different synthetic biology tools.

## 2.3 Materials and methods

Three different approaches for squalene production in *E. coli* were evaluated, as shown in **Figure 6**.



**Figure 6.** Framework to test the three different approaches for squalene production in *E. coli*.



### **2.3.1 Approach 1: Insertion of heterologous MEP pathway genes and squalene synthase genes in *E. coli***

#### **2.3.1.1 Bacterial strains and reagents**

The bacterial strains and plasmids used in this study are listed in **Table 8**, and the oligonucleotides used in this approach are summarized in **Table 9**. All chemicals, solvents, and media components were purchased from New England Biolabs (UK), Sigma-Aldrich (St. Louis, MO), Fisher Scientific (Pittsburgh, PA), or VWR (West Chester, PA) unless otherwise noted.

#### **2.3.1.2 Remotion of transit peptide signal**

The *OsDXS1*, *OsDXS2*, *OsDXS3*, *OsIDI1* and *OsIDI2* gene nucleotide sequences were translated into amino acid sequences using expasy program (<http://web.expasy.org/translate/>). Then, TargetP 1.1 server software (<http://www.cbs.dtu.dk/services/TargetP-1.1/index.php>) developed at the Technical University of Denmark DTU was used to predict the subcellular location of eukaryotic proteins without any loss of enzyme function. The location assignment is based on the predicted presence of any of the N-terminal presequences: chloroplast transit peptide (cTP), mitochondrial targeting peptide (mTP) or secretory pathway signal peptide (SP) (Emanuelsson et al., 2000). The sequence of the farnesyl pyrophosphate synthase gene from *Gentiana lutea* (Gl FPS), does not contain targeting peptide. Once the target peptide (TP) sequences were predicted, the rest of the coding sequences were amplified by PCR using the corresponding primers.

**Table 8.** Bacterial strains and plasmids used in approach 1

Strain or Plasmid	Description	Source
<b>Strain</b>	T7 Expression Strain	
	B strain carrying F- <i>ompThsdSB</i> (rB <sup>-</sup> mB <sup>-</sup> ) <i>gal dcm</i> (DE3)	
	<b>Genotype</b>	New England Biolabs
<i>Escherichia coli</i> BL21(DE3)	<i>fhuA2 [lon] ompT gal (λ DE3) [dcm]</i> <i>ΔhsdS</i> <i>λ DE3 = λ sBamHI ΔEcoRI-B</i> <i>int::(lacI::PlacUV5::T7 gene1) i21</i> <i>Δnin5</i>	
	<b>Genotype</b>	
DH5α	F- φ80lacZΔ M15 Δ (lacZYA-argF) U169 recA1 endA1 hsdR17 (rK <sup>-</sup> mK <sup>+</sup> ) phoA supE44 λ- thi-1 gyrA96 relA1	New England Biolabs
<b>Plasmids</b>		
	<i>E. coli</i> expression vector; Amp <sup>r</sup>	PROMEGA
pET-32a(+)		
pET-32a- <i>dxs1</i>	pUC8- <i>DXS1</i> carrying PCR products of <i>dxs1</i> gene from <i>Oryza sativa</i>	This study
pET-32a- <i>dxs2</i>	pUC8- <i>DXS2</i> carrying PCR products of <i>dxs2</i> gene from <i>Oryza sativa</i>	This study
pET-32a- <i>dxs3</i>	pUC8- <i>DXS3</i> carrying PCR products of <i>dxs3</i> gene from <i>Oryza sativa</i>	This study
pET-32a- <i>idi1</i>	pUC8- <i>OsIPPI1</i> carrying PCR products of <i>idi1</i> gene from <i>Oryza sativa</i>	This study
pET-32a- <i>idi2</i>	pUC8- <i>OsIPPI2</i> carrying PCR products of <i>idi2</i> gene from <i>Oryza sativa</i>	This study
pET-32a- <i>fps</i>	pUC8- <i>FPS</i> carrying PCR products of <i>fps</i> gene from <i>Gentiana lutea</i>	This study
pET-32a- <i>hsqs</i>	pUC18m- <i>hsqs</i> pUC- <i>hsqs</i> carrying PCR products of <i>hsqs</i> from human	Katabami et al; 2015
pET-32a- <i>tsqs</i>	pJE404- <i>tsqs</i> pUC- <i>tsqs tsqs</i> carrying PCR products of <i>Thermosynechococcus</i> <i>elongatus</i>	Katabami et al; 2015

**Table 9.** Oligonucleotides used in approach 1

Gene	Size (bp)	Accession number	Forward primer (5'- 3')	Reverse primer (5'- 3')
<i>OsDXS1</i>	2010	NM_001062059.1	GCGTCGCTGTCCACGGAGAGGGAGG	CTACGCGTTGGGCACCGTCATGATG
<i>OsDXS2</i>	1989	XM_015787004.1	GTGGCTGCGCTGCCAGATGTCGATG	TTACTTCATCAACAAAAGTGCCTCT
<i>OsDXS3</i>	2052	NM_001065621.1	GCGAGCGCGGCGGCGGCGGCGGCGGA	TCAGCTGAGCTGAAGTGCCTCCAAT
<i>OsIPPI1</i>	714	NM_001066454.1	ATGGCCGGCGCCGCCGCCCGTGG	TTACTTCAGCTTGTGGATGGTCTCC
<i>OsIPPI2</i>	723	NM_001062082.1	GCGGTGATGGGGAAGGCCGGCACCG	CTACAATTATGAATTGTTTTTCATG
<i>GIFPS</i>	1047	AB017371.1	ATGGCAAATCTGAACGGAACACTACAT	TTATTTTCAGCCTTGTATATCTTA
<i>TeSQS</i>	1077	--	ATQCGTGTTGGTGITAACCCACCTA	CTACAGACCACCCAGGATAAACGGC
<i>HSQS</i>	1020	--	ATGGACCGGAACCTCGCTCAGCAACA	CTAGTTCTGCGTCCGGATGGTGGAG

### 2.3.1.3 Plasmid preparation for bacterial transformation

Pairs of primers (Table 9) were designed to amplify the coding sequence, without the transit peptide signal, of the following genes: *DXS1*, *DXS2*, *DXS3*, *IDI1* and *IDI2*. Primers for *GIFPS* (*Gentiana lutea FPS*), *hsQS* (*Homo sapiens SQS*) and *tsQS* (*Thermosynechococcus elongatus SQS*) were including the restriction sites for EcoRI (GAATTC) at 5'-end from forward primers and for HindIII (AAGCTT) at 5'-end from reverse primers. A spacer of six nucleotides was also included: TAAGCA in forward primers and TGCTTA in reverse primers (Table 9). The genes that were already cloned in our lab in the vector pUC8 or pUC18m-*hsqs* and pJE404-*tsqs* from Katabami et al., 2015, were transferred into pET32a vector.

The linearized vector pET32 a+ (PROMEGA, see circular map in Figure 10), suitable for bacterial protein expression, and PCR products of the genes (generated using the primers listed above), were purified using the GeneClean II Kit (MP Biomedicals) and were digested and ligated by the Anza Restriction Enzyme Cloning System (INVITROGEN) to build the plasmids listed above (Table 8). Then, ligation products were used for bacterial transformation by heat shock into *E. coli* BL21(DE3).

### 2.3.1.4 PCR screening and clone sequencing

The bacterial clones were grown and selected onto solid agar plates LB (Luria-Bertani) medium with ampicillin 100 mg/L and incubated at 37°C for 24 hours. For each set of

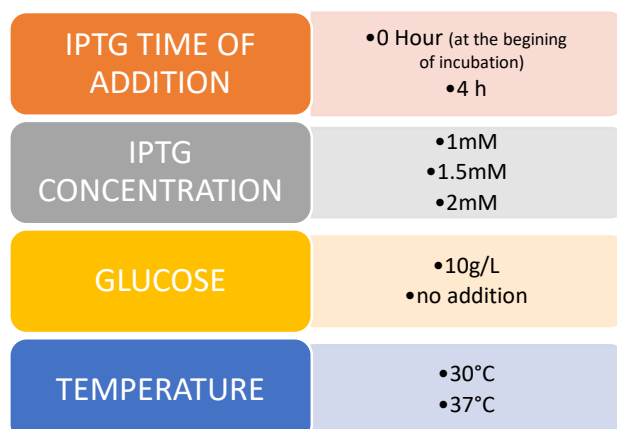
transformation, ten bacterial colonies were screened by PCR to determine the presence of the genes of interest. Following electrophoresis, PCR products of expected sizes were purified and sequenced by using an ABI 3730xl DNA analyzer (Stabvida). At least five clones for each amplified coding sequence were sequenced by using various sets of primers (F1, F2, F3, R1, R2, R3 from **Table 10**).

**Table 10:** Primers used for sequencing analysis

GENE	FWD	REV
<i>DXS1</i>	F1: GCAGGCGTACGAGGCGATGAATAA	R1: AGGCGTGTGAACACGTAGTGGAG
	F2: GAAGACGCTGTCGTACGAACTACT	R2: TACTGCACCGCCGAACCGTACCCAA
	F3: AGCGTCGCTGGTGGAGCGGCACGG	R3: CATCATGACGGTGCCCAACGCGTAG
<i>DXS2</i>	F1: CTCTCTAGTGCTCTGAGCAAGG	R1: ACCAGCCATAGTTGTCCAGTACTT
	F2: TGCAATCACAAGTGCAGGTCTGGT	R2: GATTGAACCTTGCTCAGAGCACTAGAGAG
	F3: GATACCAGTTCGCTTTGCAATCACAAGTG	R3: GTGGCTGCGCTGCCAGATGTCGATG
<i>DXS3</i>	F1: GCGAGCGCGGCGGCGGCGGCGCGGA	R1: CTTCTCGCCGAGTAGTCGATCTT
	F2: AGTCCAAGTGTTCGACGTTGCTCT	R2: CTTCTCGCCGAGTAGTCGATCTT
	F3: GAAGAAGAACCACGTCATCTCG	R3: TCAGCTGAGCTGAAGTGCCTCCAAT
<i>IDI1</i>	F1: CCGCCGTGGAGGACCCGGGATG	R1: TTACTIONCAGCTTGTGGATGGTCTCC
<i>IDI2</i>	F1: GCGGTGATGGGGAAGGCCGGCACCCG	R2: CTACAACTTATGAATTGTTTTCATG
<i>FPS</i>	F1: GCAAATCTGAACGGAACACTACATCCG	R1: TTATTTGAGCCTCTTGTATATCTTA
	F2: CCGGACAGATGATAGATTTGATTAC	R2: ATGCAGTCTGGAACCTAACCTCGTT
<i>TSQS</i>	F1: CGTGTGGTGTAAACCCACCTATGA	R1: CGCGACGTAATAACAATATTCTTTG
	F2: CACGGTGGGTTACTTGTGACGGAT	R2: CTACAGACCACCCAGGATAAACGGC
<i>HSQS</i>	F1: GACCGGAACTCGCTCAGCAACAGCC	R1: ACTAAGGGGTCTCAAACCTCTGAGG
	F2: AAGATACAGAACGTGCCAACTCTAT	R2: CTAGTTCTGCGTCCGGATGGTGGAG

### 2.3.1.5 Determination of *E. coli* culture parameters

In order to find the best conditions of culture for *E. coli*, the following parameters were adjusted: presence or absence of glucose; IPTG (isopropyl- $\beta$ -D-1-thiogalactopyranoside) concentration, which is a small molecule analogue of allolactose that allows gene transcription in the lac operon; time point of IPTG induction; temperature (see **Figure 7**). After 48 hours of growth, the optical density at 600 nm ( $OD_{600}$ ) was measured.



**Figure 7:** Parameters of the *E. coli* culture conditions.

### 2.3.1.5 Shake flask culture

The conditions for *E. coli* growth in a shake flask were: 37°C, IPTG (Isopropyl  $\beta$ -D-1-thiogalactopyranoside) 2mM added at the beginning of incubation without glucose. An *E. coli* inoculum from solid media was precultured (seed culture) at 37°C for about 6 hours in 10 mL of liquid LB medium consisted of 10 g/L tryptone; 5 g/L yeast extract; and 10 g/L NaCl, and the appropriate antibiotic (ampicillin 100  $\mu$ g/mL). Approximately 1 mL of the seed culture was inoculated to a 250 mL liquid LB media into a 500 mL flask. 2 mM of IPTG was added to the media with an  $OD_{600}$  of 0.2 and the cultures were carried out at 30°C or 37°C in a rotary shaker (220 rpm) approximately 24 hours.

### 2.3.1.6 Quantitative Determination of Squalene

#### Squalene extraction

The culture broth was harvested by centrifugation at 6000 $\times$ g for 15 min. The supernatant was discarded and the pellet was washed twice with deionized sterile water and frozen overnight at -20°C. The lipids were extracted using the method described by Bligh and Dyer (1959), with slight modification. To freeze-dried cells, chloroform, methanol, and water were added in the ratio of 1:2:0.8 by volume. Samples were

vortexed for 1 min immediately following the addition of each solvent, and allowed to stand for about 1 hour, in the tube rotator mixer. Phase separation of the biomass-solvent mixtures was achieved by adding chloroform and water to obtain final chloroform, methanol, and water ratio of 1:1:0.9 by volume. The lipid extract was recovered from the lower chloroform phase; this phase was dried under nitrogen. Dried extracts were resuspended in chloroform and then stored at -20°C until chromatographic analysis.

### **Chromatographic analysis**

All samples and standards were analyzed by as GC-FID previously described by Han et al., (2018). The instrument used was a Hewlett-Packard Series II gas chromatograph, model 5890 (Hewlett-Packard, Avondale, PA), equipped with FID (flame-ionization detector) and an on-column injector. A BPX 5 fused-silica capillary column (15 m ×0.32 mm), coated with 5%phenyl/95% polysilphenylsiloxane, with a film thickness of 0.25 µm (SGE, Austin, TX), was used. The GC oven temperature was initially held at 40°C for 3 min, followed by an increase to 240°C at a rate of 10°C/min where it was maintained for 5 min until increasing to 300°C at a rate of 10°C/min and maintained for 1 min. The injector temperature was 300°C and the detector temperature was 320°C. Squalene quantification was performed using a 5-point calibration assay of pure squalene and squalane as external standard, with  $R^2$  values of >0.9900. All analyses were performed in duplicate.

### **2.3.2 Approach 2: Overexpression of *E. coli* MEP pathway genes and heterologous squalene synthase genes in *E. Coli*.**

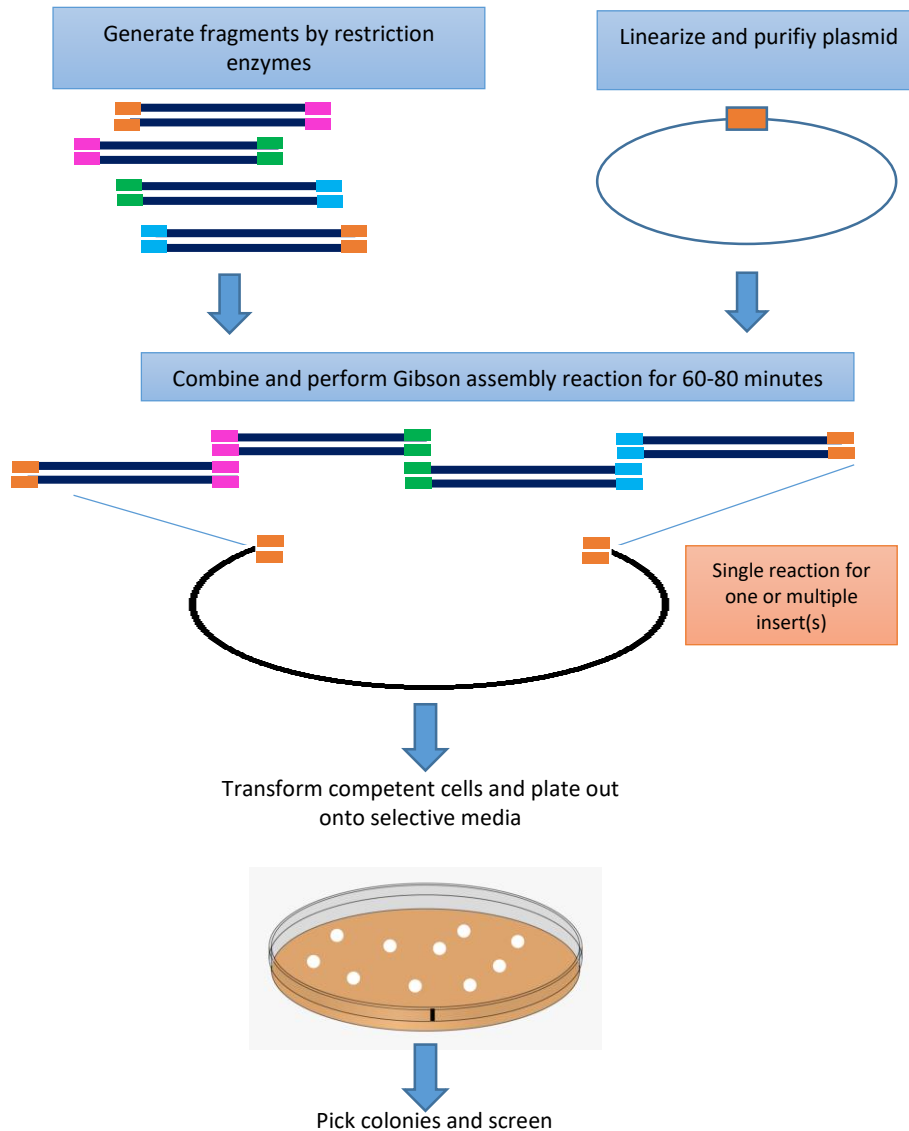
#### **2.3.2.1 Bacterial strains and reagents**

The bacterial strains, chemicals, solvents and media components used in this set of experiments were mentioned and described previously in approach 1. The *E. coli* BL21(DE3) strain was used for DNA subcloning and for squalene production. Ampicillin (AMP) were used at 100 mg/mL. Squalene synthase genes from *Thermosynechococcus elongatus* and *Homo sapiens* were obtained from Katabami et al. (2015).

#### **2.3.2.2 Plasmid construction**

MEP pathway genes *DXS*, *DXR*, *ispD*, *ispE*, *ispF*, *ispG*, *ispH*, *fldA* and *IDI* from *E. coli*, *ERG20* (also called *FPS*) from *Saccharomyces cerevisiae* and *TSQS* or *HSQS* (used in approach 1) were subcloned into the KpnI/EcoRI or XbaI/SacI sites from pET32a (+) vector by using the Invitrogen Anza Cloning System. Gibson Assembly® Cloning Kit (Gibson et al., 2010) was used to build vectors containing several expression cassettes according to the manufacturer's instructions: an overview of the procedure is shown in Figure 9. This method has been successfully used for assembly of multigene constructs in metabolic

engineering (Weissmann et al., 2018; Wang et al., 2015; Jiang et al., 2015). Different plasmids were constructed as intermediate vectors (pET32a (+) **Figure 8**) and expression cassettes were assembled through the Gibson system in two sets of constructs as is indicated in **Table 11**.



**Figure 8.** A general overview of Gibson Assembly cloning method adapted from Gibson Assembly cloning guide (2017). With this method, it was possible to insert several DNA fragments simultaneously in a single reaction in a short time (1 hour) by the action of three different enzymes: **1) T5 Exonuclease** creates single-strand DNA 3' overhangs by chewing back DNA from the 5' end. Complementary DNA fragments can subsequently anneal to each other. **2) Phusion DNA Polymerase** incorporates nucleotides to "fill in" the gaps in the annealed DNA fragments. **3) Taq DNA Ligase** - covalently joins the annealed complementary DNA fragments, removing any nicks and creating a contiguous DNA fragment. The Gibson Assembly method requires a linearized vector and 20–80 bp sequence overlaps at the ends of the DNA elements to be assembled.

### 2.3.2.3 PCR Primer design

The primers used in Gibson Assembly were designed with an overlap sequence that was required for creating homologous overlap regions. To assemble several inserts with a linearized vector, each DNA partner carried at its ends a 20-40 base pair region homologous to the adjacent fragments. A restriction site was added between the overlap region and the sequence-specific segment, enabling the subsequent release of the insert from the vector for diagnostic analysis. NEBuilder Assembly Tool from New England Biolab was used to design the primers, available in <https://nebuilder.neb.com/>. Details of the primers are shown in **Table 11**.

**Table 11.** Oligonucleotides used for Gibson Assembly

Transformation construct	Genes	Gibson assembly Primers fwd	Gibson assembly Primers rev
vector 1	<i>EcDXS</i>	F1( <i>DXS</i> ): GGAATTGTGAGCGGATAACAATTCGGT ACCATGAGTTTTGA	R1( <i>DXS</i> ): CCTGGCTGGCATAAGAATTC AAGGAGATAT ACATATGAGC
	<i>EcDXR</i>	F2 ( <i>DXR</i> ): AAGGAGATATACATATGAGCGGTACCAT GAAGCAACTCAC	R2( <i>DXR</i> ): GTCTCGCAAGCTGAGAGCTCAAGGAGTTT GACACGGATGT ACTCAA
vector 2	<i>EcispD</i>	F1( <i>ispD</i> ): GGAATTGTGAGCGGATAACAATTCGGT ACCATGGCAACCAC	R1( <i>ispD</i> ): AGGAGAATACATAAGAATTC AAGGAGATAT ACATATGAGC
	<i>EcispE</i>	F2( <i>ispE</i> ): AAGGAGATATACATATGAGCGGTACCAT GCGGACACAGTG	R2( <i>ispE</i> ): GAGCCATGCTTTAAGAGCTCAAGGAGTTG ACACGGATGTA CTCAA
	<i>EcispF</i>	F3( <i>ispF</i> ): AAGGAGTTTGACACGGATGTGGTACCA TGCGAATTGGACA	R3( <i>ispF</i> ): AGGCAACAAAATGAAGAATTC AAGGAGTTT CTGGGCAGAGTGGTGCG



		F1 ( <i>ispG</i> ):	R1 ( <i>ispG</i> ):
	<i>EcispG</i>	<u>GGAATTGTGAGCGGATAACAATTCGGT</u> <u>ACCATGCATAACCA</u>	<u>AGGTTGAAAAATAAGAATTCAAGGAGATAT</u> ACATATGAGC
		F2 ( <i>ispH</i> ):	R2 ( <i>ispH</i> ):
vector 3	<i>EcispH</i>	<b>AAGGAG</b> ATATACATATGAGCGGTACCAT <b>GCAGATCCTGTT</b>	<b>GTGAAGTCGATTAA</b> <u>GAGCTCAAGGAGTTT</u> GACACGGATGTACTION
		F3 ( <i>FLDA</i> ):	R3 ( <i>FLDA</i> ):
	<i>EcFLDA</i>	<b>AAGGAG</b> TTTGACACGGATGTGGTACCA <b>TGGCTACTCTGG</b>	<b>TCTCAATGCCTGAGAATTCAAGGAGTCTGG</b> GCAGAGTGGTGCCG
	<i>ERG20</i> ( <i>FPS</i> )	F1( <i>ERG20</i> ): <u>GGAATTGTGAGCGGATAACAATTCGA</u> <u>ATTCTATTTGCTTCT</u>	R1( <i>ERG20</i> ): <u>TTTTCTGAAGCCATCTCGAGAAGGAGATA</u> TACATATGAGC
	<i>EcDI</i>	F2( <i>DI1</i> ): <b>AAGGAG</b> ATATACATATGAGCGAATTC <b>ATGGCCGGCGCCGC</b>	R2( <i>DI1</i> ): <b>ACAAGCTGAAGTAA</b> <u>AAGCTTAAGGAGTTT</u> GACACGGATGTACTION
vector 4	<i>HSQS</i>	F3( <i>HSQS</i> ): <b>AAGGAG</b> ATATACATATGAGCGAATTC <b>ATGGACCGAACTC</b>	R3( <i>HSQS</i> ): <b>GGACGCAGAACTAG</b> <u>AAGCTTAAGGAGTTT</u> ACACGGATGTACTION
	Or	F3( <i>TSQS</i> ): <b>AAGGAG</b> ATATACATATGAGCGAATTC <b>ATGCGTGTGGTGT</b>	R3( <i>TSQS</i> ): <b>TGGGTGGTCTGTAG</b> <u>AAGCTTAAGGAGTTT</u> GACACGGATGT ACTCAA
	<i>TSQS</i>		

\*Lac operator in blue, restriction enzyme underlined, gene complement in red, RBS in bold, the reverse primers containing a complementary sequence to the pET32a(+) vector, which contain T7 promoter and terminator.

### 2.3.2.4 Confirmation of Gene presence

To evaluate the success of the assembly reaction, a PCR reaction was performed to confirm the integration of DNA fragments in the vectors and then were analyzed on a gel electrophoresis with 50% of the assembly reaction on a 0.8-2% agarose gel.

### **2.3.2.5 Bacterial transformation and clone screening**

Two microliters of the Gibson assembly product were mixed gently with NEB 5-alpha *E. coli* competent cells (provided with the kit). The mixture was placed on ice for 30 min, followed by a heat shock at 42°C for 30 seconds, and transferred back to ice for 2 minutes. Following the addition of 950 µL of SOC media, cells were incubated at 37°C for 1 hour, spread onto solid LB plate with ampicillin (100mg/L) and grown at 37°C overnight. Following colony subculturing and plasmid isolation, genes of interest were sequenced.

### **2.3.2.6 Shake flask culture**

The 250 mL shake flask cultures were performed as described in approach 1, using colonies with the correct integration cassettes.

### **2.3.2.7 Quantitative Determination of Squalene**

The metabolite quantification analysis was carried out as previously described in approach 1.

## **2.3.3 Approach 3: Expression of heterologous squalene synthase genes in the *E. coli* strain engineered with the dual MEP/MVA pathway.**

### **2.3.3.1 Recovery of bacterial strain from a transferred biological sample material**

Dr. Sheng Yang<sup>1</sup> provided the bacterial strain named 1440 (Yang et al., 2016), containing the MEP/MVA dual pathway, which was constructed from other strains that had resistance to three different antibiotics: ampicillin, spectinomycin and chloramphenicol. After signing a MTA (Material Transfer Agreement), we received the bacterial strain in filter paper format, immediately, a piece of this paper was immersed in liquid LB in a 2 mL sterile Eppendorf tube and incubated at 37°C/180rpm, overnight. Then, 10 µL of culture was plated out onto an appropriate antibiotic (ampicillin 100 µg/mL, chloramphenicol 30 µg/mL and spectinomycin 50 µg/mL) agar plate and grew overnight. Single colonies were isolated and inoculated for an overnight liquid culture. A glycerol stock was created by mixing gently 500 µL of the overnight culture with 500 µL of 50% glycerol in a 2 mL cryovial and stored at -80°C until further usage.

---

<sup>1</sup> Laboratory of Synthetic Biology, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, China.

### **2.3.3.2 Confirmation by PCR of the presence of the key genes**

With a sterile loop, a small amount of glycerol stock was used to do streaks of single colonies onto a solid agar plate with the appropriate antibiotics. Plate was incubated at 37 °C until colonies grew enough. Then, they were collected with a pipette tip and transferred to an Eppendorf tube with 30 µL TE (Tris-EDTA) buffer and mixed gently. One microliter of this mixture was used to performed PCR analysis of the key genes from the dual pathway with the primers and conditions reported previously in Yang et al. (2016).

### **2.3.3.3 Transformation of the dual pathway strain with two squalene synthase genes**

*Gentiana lutea* FPS, *Thermosynechococcus elongatus* or *Homo sapiens* squalene synthase genes (pUC18m-*hsqs* and pJE404-*tsqs*) were transformed into the *E. coli* BL21 (DE3) strain containing the MEP/MVA dual pathway, and transformants were selected onto LB-agar plates with ampicillin, chloramphenicol and spectinomycin according to the strain antibiotic resistance and incubated at 37 °C overnight.

### **2.3.3.4 Shake flask culture**

The 250mL shake flask cultures were performed as described in approach 1 after screening successfully transformed colonies.

### **2.3.3.5 Quantitative Determination of Squalene**

Squalene quantification analysis was carried out as previously described in approach 1.

## 2.4 RESULTS

### 2.4.1 Approach 1: Heterologous expression of MEP pathway genes from plants and two squalene synthase genes in *E. coli*.

#### 2.4.1.1 Selection of genes

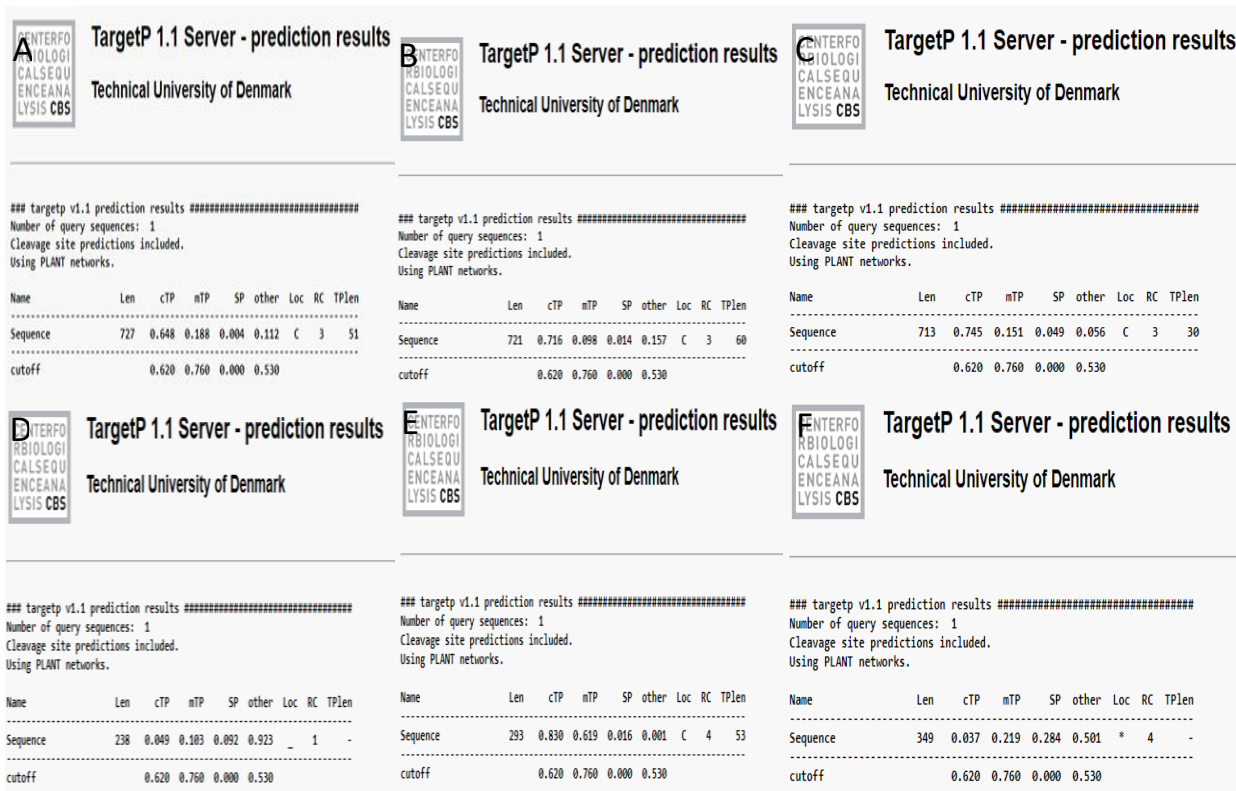
The genes encoding key enzymes involved in squalene biosynthesis and their function in the metabolic pathway are shown in **Table 12**. We selected the genes *DXS* and *IDI* from *Oryza sativa* and *FPS* from *Gentiana lutea* because our lab has extensive experience working with those genes (Jin et al., 2020, 2021; Bai et al., 2016, 2014; Pérez et al., 2022; Cervantes-Cervantes et al., 2006), and *SQS* from human and the bacteria *Thermosynechococcus elongatus* because they have been expressed successfully for squalene production in microbes (Katabami et al., 2015; Lee and Poulter 2008; Thompson et al., 1998).

**Table 12.** Genes encoding key enzymes involved in squalene biosynthesis and their function in the pathway

Gene	Function
<i>OsDXS</i>	Condensation of glyceraldehyde 3-phosphate and pyruvate in deoxyxylulose-5-phosphate (DXP)
<i>OsIDI</i>	Catalyses the 1,3-allylic rearrangement of the homoallylic substrate isopentenyl (IPP) to the isomer DMAPP
<i>GIFPS</i>	Catalyses the formation of farnesyl diphosphate (FPP), a sterol precursor
Human or <i>T. elongatusSQS</i>	Catalyses the condensation of two farnesyl pyrophosphate moieties to form squalene. It is the first committed enzyme of the sterol biosynthesis pathway

#### 2.4.1.1 Truncation of *OsDXS* 1, 2, 3 and *IDI2*

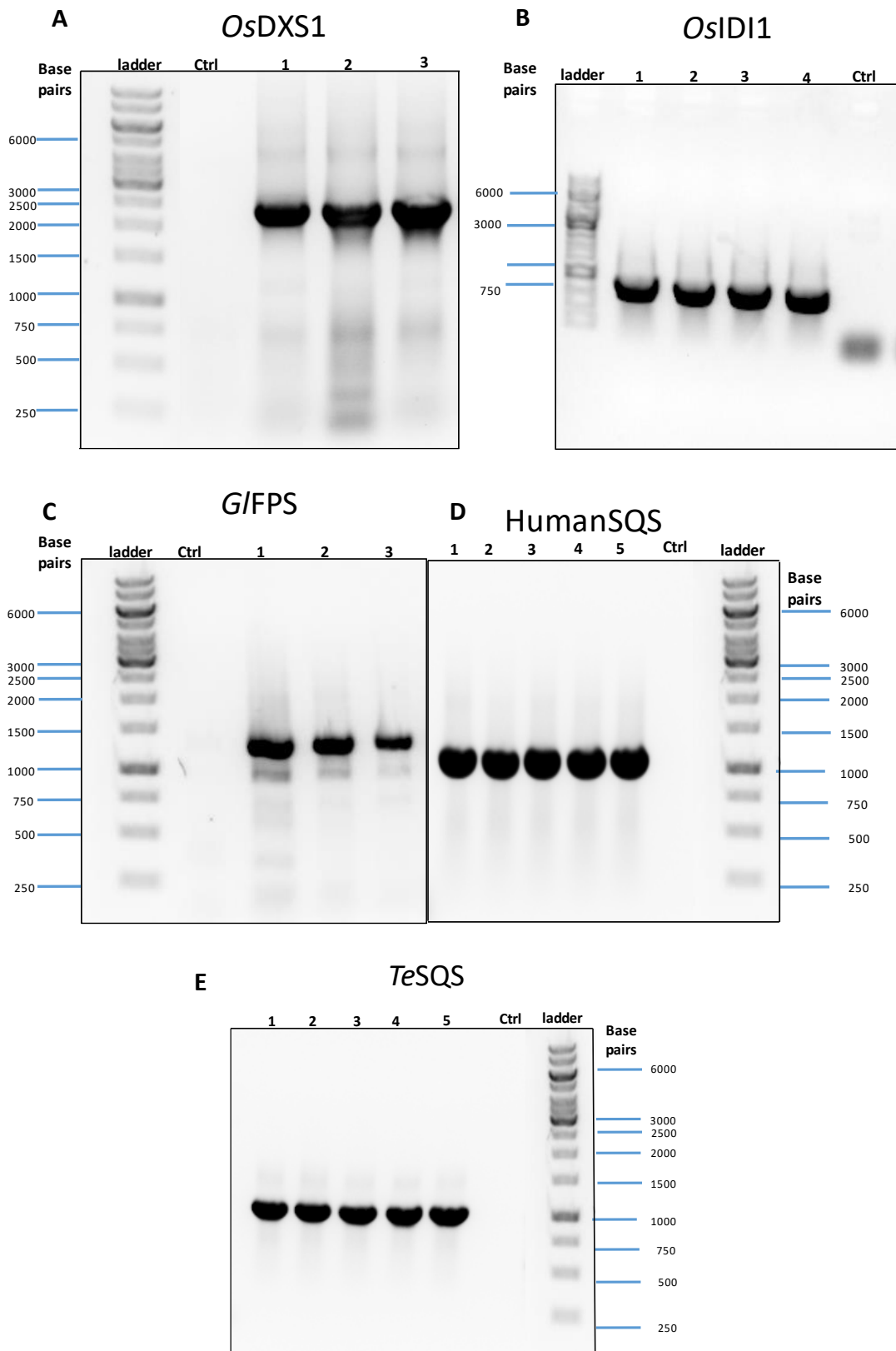
In order to express *OsDXS* 1,2,3 and *OsIDI2* in a prokaryotic system, transit peptide (TP), which allows protein sub-cellular localization in eukaryotes, has been removed from the gene coding sequences. The target TP 1.1 server software indicated the presence of a putative 51-residue transit peptide for *DXS1*, for *DXS2*, 60; *DXS3*, 30; and for *IDI2*; 53. A transit peptide for *IDI1* and *FPS* were not predicted for the software (**Figure 9**). Therefore, the corresponding TP sequences have been avoided during PCR amplification of the plant coding sequences. The truncated version of the human *SQS* gene performed by Katabami et al. (2015) gene was used in this study.



**Figure 9.** TP length prediction for A, *DXS 1*; B, *DXS 2*; C, *DXS 3*; D, *IDI 1*; E, *IDI 2*; F, *FPS*.

### 2.4.1.2 Strain building

Based on the results obtained from the functional analysis in our laboratory with the isoforms *DXS1/2/3* and *IDI1/2* (Jin et al., 2020 and 2021), the *E. coli* strain was engineered to express the isoforms *DXS1* and *IDI1* coding sequences. The genes *DXS1*, *IDI1*, *FPS* and *SQS*, were inserted in the plasmid pET32a+. The presence of the genes in the plasmids was confirmed by PCR (**Figure 10**). A colony containing the right plasmid was subcultured and sequenced. The sequencing data for each gene was aligned with the corresponding GeneBank sequence and no mutations were observed (**Figure 11-15**). The plasmids containing the selected genes were transformed in *E. coli* BL21(DE3) competent cells, and PCR screening was carried to confirm the presence of those genes (**Figure 16**), resulting in the strains MEP-hSQS1 or MEP-tSQS1.



**Figure 10.** *E. coli* colony PCR for selected genes in the plasmid pET-32a(+). (A) *OsDXS1*, with three individual colonies; (B) *OsIDI1*, four colonies; (C) *G/FPS*, and three colonies (D, E) *SQS* from Human or *T. elongatus* with five individual colonies. For each gene amplification, water instead of culture lysis was used as a template for the control.

OsDXS1

DXS1 sanger sequencing.ape from 1 to 2189  
Alignment to  
DXS1 gene bank.ape from 1 to 2189

Matches():2189  
Mismatches(#):0  
Gaps():0  
Unattempted(.):0

```

      *      *      *      *      *      *      *      *      *      *
1 ATGGCGCTCAGACGTTCTCCATTTTCGAGAGGAGGCTTCGTCGGCGCGGTGCCGCAAGGAGGGGCAITTCGCTCCGGCGCGGGCGGAGCTCAGTCTCCACA 100
|
|
|
1 ATGGCGCTCAGACGTTCTCCATTTTCGAGAGGAGGCTTCGTCGGCGCGGTGCCGCAAGGAGGGGCAITTCGCTCCGGCGCGGGCGGAGCTCAGTCTCCACA 100
      *      *      *      *      *      *      *      *      *      *
101 AGTCCAGAGCAGGCCACACAAGGCTAGGCGGAGGTCGTCGTCGAGCATCTCGGCGTCGCTGTCCACGGAGAGGGAGGGCGGGAGTACCCTCGCAGCG 200
|
|
|
101 AGTCCAGAGCAGGCCACACAAGGCTAGGCGGAGGTCGTCGTCGAGCATCTCGGCGTCGCTGTCCACGGAGAGGGAGGGCGGGAGTACCCTCGCAGCG 200
      *      *      *      *      *      *      *      *      *      *
201 GCCACCAGCCCGCTGCTGGACACGGTCAACTACCCCATCCACATGAAGAACCTGTCCCTCAAGGAGCTCCAGCAGCTCGCCGACGAGCTCCGCTCCGAC 300
|
|
|
201 GCCACCAGCCCGCTGCTGGACACGGTCAACTACCCCATCCACATGAAGAACCTGTCCCTCAAGGAGCTCCAGCAGCTCGCCGACGAGCTCCGCTCCGAC 300
      *      *      *      *      *      *      *      *      *      *
301 GTCATCTTCCACGTCTCCAAGACCGGGGACATCTCGGGTCCAGCCTCGGGCTCGTTCGAGCTCACCGTCCGCTCCACTACGTGTCAACACGCTCAGG 400
|
|
|
301 GTCATCTTCCACGTCTCCAAGACCGGGGACATCTCGGGTCCAGCCTCGGGCTCGTTCGAGCTCACCGTCCGCTCCACTACGTGTCAACACGCTCAGG 400
      *      *      *      *      *      *      *      *      *      *
401 ACAAGATCCTCTGGGACGTCGGCCACCCAGTCTACCCCTCACAAGATTCTGACCGGGCGGGCGACAAGATGCCGAGATGCGTCAGACCAACGGCTTGTG 500
|
|
|
401 ACAAGATCCTCTGGGACGTCGGCCACCCAGTCTACCCCTCACAAGATTCTGACCGGGCGGGCGACAAGATGCCGAGATGCGTCAGACCAACGGCTTGTG 500
      *      *      *      *      *      *      *      *      *      *
501 GGGATTACCAAGCGGTTCGGAGAGCGAGTACGACTCCTTCGGCACCGGCCACAGCTCCACCACCATCTCCGCGCCCTCGGGATGGCGGTGGGGAGGGAT 600
|
|
|
501 GGGATTACCAAGCGGTTCGGAGAGCGAGTACGACTCCTTCGGCACCGGCCACAGCTCCACCACCATCTCCGCGCCCTCGGGATGGCGGTGGGGAGGGAT 600
      *      *      *      *      *      *      *      *      *      *
601 CTCGAAGGAGGGAAGAACAACGTGGTGGCGGTGATCGGCGACGGCCCATGACGGCCGGCAGGCGTACGAGGCGATGAATAACCGGGGTATCTCGACT 700
|
|
|
601 CTCGAAGGAGGGAAGAACAACGTGGTGGCGGTGATCGGCGACGGCCCATGACGGCCGGCAGGCGTACGAGGCGATGAATAACCGGGGTATCTCGACT 700
      *      *      *      *      *      *      *      *      *      *
701 CCGATATGATCGTGATTCTCAACGACAACAAGCAGGTGTCGCTGCGGACGGCGACGCTCGACGGGCGGGCGCCGCGGTGGGCGCGCTCAGCAGCGCCCT 800
|
|
|
701 CCGATATGATCGTGATTCTCAACGACAACAAGCAGGTGTCGCTGCGGACGGCGACGCTCGACGGGCGGGCGCCGCGGTGGGCGCGCTCAGCAGCGCCCT 800
      *      *      *      *      *      *      *      *      *      *
801 CAGCAAGCTGCACTCCAGCCGCCACTCAGGGAGCTCAGGGAGGTGGCAAAGGGCGTGACGAAGCAAATCGGAGGGTCCGTTGCACGAGCTGGCGCGCAAG 900
|
|
|
801 CAGCAAGCTGCACTCCAGCCGCCACTCAGGGAGCTCAGGGAGGTGGCAAAGGGCGTGACGAAGCAAATCGGAGGGTCCGTTGCACGAGCTGGCGCGCAAG 900
      *      *      *      *      *      *      *      *      *      *
901 GTGGACGAGTACGCCCGGGCATGATCAGCGGCTCGGCTCGAGCTCTTCGAGGAGCTCGGCCTCTACTACATCGGCCCGGTCGACGGCCACAACATCG 1000
|
|
|
901 GTGGACGAGTACGCCCGGGCATGATCAGCGGCTCGGCTCGAGCTCTTCGAGGAGCTCGGCCTCTACTACATCGGCCCGGTCGACGGCCACAACATCG 1000
      *      *      *      *      *      *      *      *      *      *
1001 ACGACCTCATACCATCCTCCGCGAGGTCAAGAGCACCAAGACCACAGGCCCGGTGCTCATCCAGTCTGTCACCGAGAAAAGCCGCGGCTACCCCTACGC 1100

```

```

|||||
1001 ACGACCTCATCACCATCTCCGCGAGGTCAAGAGCAACAGACCACAGGCCCGGTGCTCATCCACGTGTCACCGAGAAAAGCCCGGGTACCCCTACGC 1100
      * * * * *
      * * * * *
1101 CGAGCGCGCCCGCAAGTACCACGGCGTGGCGAAGTTCGATCCGGCGACGGGGAAGCAGTTCAAGTCGCCCGGCAAGACGGTGTCTACACGAATAC 1200
      * * * * *
1101 CGAGCGCGCCCGCAAGTACCACGGCGTGGCGAAGTTCGATCCGGCGACGGGGAAGCAGTTCAAGTCGCCCGGCAAGACGGTGTCTACACGAATAC 1200
      * * * * *
      * * * * *
1201 TTCGCGGAGGCGCTCATCGCCGAGGCGGAGCAGGACAAAGGTCGTGGCCATCCACGCGGCCATGGGGGAGGCACGGGGCTCAACTACTTCTCCGGC 1300
      * * * * *
1201 TTCGCGGAGGCGCTCATCGCCGAGGCGGAGCAGGACAAAGGTCGTGGCCATCCACGCGGCCATGGGGGAGGCACGGGGCTCAACTACTTCTCCGGC 1300
      * * * * *
      * * * * *
1301 GCTTCCCGAACCGTGTCTCGACGTCGGGATCGCCGAGCAGCACGCCGTACGTTCCGCCCGCGCCCTCGCCTGCGAGGGCCCTCAAGCCGTTCTCGCCAT 1400
      * * * * *
1301 GCTTCCCGAACCGTGTCTCGACGTCGGGATCGCCGAGCAGCACGCCGTACGTTCCGCCCGCGCCCTCGCCTGCGAGGGCCCTCAAGCCGTTCTCGCCAT 1400
      * * * * *
      * * * * *
1401 CTACTCTCTTCTTCCGAGAGGCTACGACAGGTGGTGACGACGCTGGACCTCCAGAAGTCCCGGTGAGGTTGCCATGGACAGGGCCGGGCTCGTG 1500
      * * * * *
1401 CTACTCTCTTCTTCCGAGAGGCTACGACAGGTGGTGACGACGCTGGACCTCCAGAAGTCCCGGTGAGGTTGCCATGGACAGGGCCGGGCTCGTG 1500
      * * * * *
      * * * * *
1501 GCGCGGAGGGCGGACGCACTGCGGCGGTTTCAAGTACCTACATGGCGTCCGCAACATGGTGTCTATGGCCCCGTCCGACGAGGCGGAGCTCT 1600
      * * * * *
1501 GCGCGGAGGGCGGACGCACTGCGGCGGTTTCAAGTACCTACATGGCGTCCGCAACATGGTGTCTATGGCCCCGTCCGACGAGGCGGAGCTCT 1600
      * * * * *
      * * * * *
1601 GCCACATGGTCCGACCCGCGGCCATCGACAGCCGCCCTCTGCTTCCGCTACCCAGAGGCAACGGCATCGCGTCCCGCTACCCACCAACTACAA 1700
      * * * * *
1601 GCCACATGGTCCGACCCGCGGCCATCGACAGCCGCCCTCTGCTTCCGCTACCCAGAGGCAACGGCATCGCGTCCCGCTACCCACCAACTACAA 1700
      * * * * *
      * * * * *
1701 AGCGGTTCCCTCGAGGTAGGCAAAAGGGAGGTTACTGTGAGGGGCGAGAGGTTGGCGTCTTGGGTACGGTTCGGCGGTGCACTACTGCTCGCCGCA 1800
      * * * * *
1701 AGCGGTTCCCTCGAGGTAGGCAAAAGGGAGGTTACTGTGAGGGGCGAGAGGTTGGCGTCTTGGGTACGGTTCGGCGGTGCACTACTGCTCGCCGCA 1800
      * * * * *
      * * * * *
1801 GCGTCGCTGGTGAGCGGCGACGGCCCAAGGTGACCGTCGCCGACGCGAGGTTCTGCAAGCCGCTGGGAAGACGCTGTCGTACACGAACTACTACCAAC 1900
      * * * * *
1801 GCGTCGCTGGTGAGCGGCGACGGCCCAAGGTGACCGTCGCCGACGCGAGGTTCTGCAAGCCGCTGGGAAGACGCTGTCGTACACGAACTACTACCAAC 1900
      * * * * *
      * * * * *
1901 GCTCATCAGGAGGCTGGCAGCTCCACGAGGTGCTCCTACCGTCGAGGAAGGCTCCATCGGGGTTCCGGTCCACGTCGCGCAGTTCATGGCCCTC 2000
      * * * * *
1901 GCTCATCAGGAGGCTGGCAGCTCCACGAGGTGCTCCTACCGTCGAGGAAGGCTCCATCGGGGTTCCGGTCCACGTCGCGCAGTTCATGGCCCTC 2000
      * * * * *
      * * * * *
2001 GACGGCTCTCTGACGGCAAACTCAAGTGGCGGCGTGGTGTACCCGACAGGTACATCGACCAGGGTACCCGCGGATCAGTGGCGGAGGCAGGGC 2100
      * * * * *
2001 GACGGCTCTCTGACGGCAAACTCAAGTGGCGGCGTGGTGTACCCGACAGGTACATCGACCAGGGTACCCGCGGATCAGTGGCGGAGGCAGGGC 2100
      * * * * *
      * * * * *
2101 TGACGCCGTCGACATCGCGGCGACGGTGTCAACGTGCTGGGCCAGGCGAGGAGGCGCTGCCATCATGACGGTGCCCAACCGGTAG 2189
      * * * * *
2101 TGACGCCGTCGACATCGCGGCGACGGTGTCAACGTGCTGGGCCAGGCGAGGAGGCGCTGCCATCATGACGGTGCCCAACCGGTAG 2189
      * * * * *

```

**Figure 11.** Sanger sequencing alignment of *OsDXS1* with the corresponding GeneBank GeneBank wild type sequence (*ApE: A plasmid Editor, 2022*)



*OsIDI1*

```
-
ID11 Sanger sequencing.apc from 1 to 717
Alignment to
OsIDI1.apc from 1 to 717

Matches (|):717
Mismatches (#):0
Gaps ( ):0
Unattempted (.):0

      *      *      *      *      *      *      *      *      *      *
1 ATGGCCGGCCGCCCGCCCGCGCTGGAGGACGCCGGGATGGACGAGGTCAGAAAGCGGCTCATGTTTCGACGACGAATGCATTTTGGTGGATGAACAAGACA 100
  |
  |
  |
1 ATGGCCGGCCGCCCGCCCGCGCTGGAGGACGCCGGGATGGACGAGGTCAGAAAGCGGCTCATGTTTCGACGACGAATGCATTTTGGTGGATGAACAAGACA 100
      *      *      *      *      *      *      *      *      *      *

      *      *      *      *      *      *      *      *      *      *
101 ATGTTGTGGCCATGAATCAAAATATAACIGCCATCTGATGGAAAAATCGAATCTGAAAAATCTACTTCATAGGGCTTCAGTGTATTCTCTGTTCAACTC 200
  |
  |
  |
101 ATGTTGTGGCCATGAATCAAAATATAACIGCCATCTGATGGAAAAATCGAATCTGAAAAATCTACTTCATAGGGCTTCAGTGTATTCTCTGTTCAACTC 200
      *      *      *      *      *      *      *      *      *      *

      *      *      *      *      *      *      *      *      *      *
201 AAAATATGAACCTCTACTCCAGCAACGATCTGCAACAAGGTTACATTTCCCTCTAGTTTGGACCAACACTTGGTCGAGCCATCCCTGTACCCGTGAGTCT 300
  |
  |
  |
201 AAAATATGAACCTCTACTCCAGCAACGATCTGCAACAAGGTTACATTTCCCTCTAGTTTGGACCAACACTTGGTCGAGCCATCCCTGTACCCGTGAGTCT 300
      *      *      *      *      *      *      *      *      *      *

      *      *      *      *      *      *      *      *      *      *
301 GAGCTTATACAGGAAAACTACCTTGGTGTAGAAAATGCTGCTCAGAGGAAGCTCTTGGATGAGCTGGGCATCCCAGCTGAAGATGTGCCAGTTGACCAAT 400
  |
  |
  |
301 GAGCTTATACAGGAAAACTACCTTGGTGTAGAAAATGCTGCTCAGAGGAAGCTCTTGGATGAGCTGGGCATCCCAGCTGAAGATGTGCCAGTTGACCAAT 400
      *      *      *      *      *      *      *      *      *      *

      *      *      *      *      *      *      *      *      *      *
401 TCACCCCTCTTGGTCGGATGCTTTACAAGGCCCATCTGATGGAAAAATGGGGTGAACACGAGCTTGACTACCTGCTTCATCGTCCGGCAGCTGAAGGT 500
  |
  |
  |
401 TCACCCCTCTTGGTCGGATGCTTTACAAGGCCCATCTGATGGAAAAATGGGGTGAACACGAGCTTGACTACCTGCTTCATCGTCCGGCAGCTGAAGGT 500
      *      *      *      *      *      *      *      *      *      *

      *      *      *      *      *      *      *      *      *      *
501 AGTCCCGAACCCGGACGAAGTGGCCGATGTGAAATACGTGAGCCGTGAGCAGCTGAAGGAGCTCATCCGAAAGCGGACGCCGGAGAGGAGGCCCTGAAG 600
  |
  |
  |
501 AGTCCCGAACCCGGACGAAGTGGCCGATGTGAAATACGTGAGCCGTGAGCAGCTGAAGGAGCTCATCCGAAAGCGGACGCCGGAGAGGAGGCCCTGAAG 600
      *      *      *      *      *      *      *      *      *      *

      *      *      *      *      *      *      *      *      *      *
601 CTGTCTCCCTGGTTCGGCTGGTGTGACAACTTCCTCATGGGCTGGTGGGATCAGTCGAGAAAGGCACCCTCAACGAGGCCGTGGACATGGAGACCA 700
  |
  |
  |
601 CTGTCTCCCTGGTTCGGCTGGTGTGACAACTTCCTCATGGGCTGGTGGGATCAGTCGAGAAAGGCACCCTCAACGAGGCCGTGGACATGGAGACCA 700
      *      *      *      *      *      *      *      *      *      *

      *
701 TCCACAAGCTGAAGTAA 717
  |
  |
  |
701 TCCACAAGCTGAAGTAA 717
      *
```

**Figure 12.** Sanger sequencing alignment of *OsIDI1* with the corresponding GeneBank GeneBank wild type sequence (*ApE: A plasmid Editor*, 2022)

**GIFPS**

FPS Sanger sequencing.ape from 1 to 1050  
 Alignment to  
 G1FPS.ape from 1 to 1050

Matches (|):1050  
 Mismatches (#):0  
 Gaps ( ):0  
 Unattempted (.):0

```

      *      *      *      *      *      *      *      *      *      *
1  ATGGCAAATCTGAAACGGAACTACATCCGATCTGAGAACTACATTCCTGCAAGTCTATCCGTA
  |
1  ATGGCAAATCTGAAACGGAACTACATCCGATCTGAGAACTACATTCCTGCAAGTCTATCCGTA
  |
      *      *      *      *      *      *      *      *      *      *
101 GGTCTCCTGATTCTCGTCAATGGGTCGAACGGATGCTGGATTACAATGTTCTGGAGGAAAGCTCAACAGAGGCCTTTCTGTCATTGATAGCTATGAACA 200
  |
101 GGTCTCCTGATTCTCGTCAATGGGTCGAACGGATGCTGGATTACAATGTTCTGGAGGAAAGCTCAACAGAGGCCTTTCTGTCATTGATAGCTATGAACA 200
  |
      *      *      *      *      *      *      *      *      *      *
201 GCTAAAAGAAGGAAAGGAGTTAACAAAGGATGAGGCTTTCTTGCAGTCTAGGCTGGTATTGAATGGCTCCAGGCATACTTCTTGTCTCTGAT 300
  |
201 GCTAAAAGAAGGAAAGGAGTTAACAAAGGATGAGGCTTTCTTGCAGTCTAGGCTGGTATTGAATGGCTCCAGGCATACTTCTTGTCTCTGAT 300
  |
      *      *      *      *      *      *      *      *      *      *
301 GACATTATGGATGGCTCTCACACTAGGCGTAGCCAACTTGTGGTTCAGATTGCCAAAGGTTGGAATGATAGCTGTAATGACGGTGTCTCTCTCGCA 400
  |
301 GACATTATGGATGGCTCTCACACTAGGCGTAGCCAACTTGTGGTTCAGATTGCCAAAGGTTGGAATGATAGCTGTAATGACGGTGTCTCTCTCGCA 400
  |
      *      *      *      *      *      *      *      *      *      *
401 ACCACATACCAGAAATTCCTCAAGAGCACTTTAGAGAAAAGCCTTACTATGTAGACCTATTGGACCTTTTCAACGAGGTTGAGTCCAGACTGCATCCGG 500
  |
401 ACCACATACCAGAAATTCCTCAAGAGCACTTTAGAGAAAAGCCTTACTATGTAGACCTATTGGACCTTTTCAACGAGGTTGAGTCCAGACTGCATCCGG 500
  |
      *      *      *      *      *      *      *      *      *      *
501 ACAGATGATAGATTGATTACTACTCTTGAAGGAGAAAAGGATCTATCAAATATTCTTGTCTCTACACCGGAGGATTGTTTCAGTACAAAAGTGCCTGAC 600
  |
501 ACAGATGATAGATTGATTACTACTCTTGAAGGAGAAAAGGATCTATCAAATATTCTTGTCTCTACACCGGAGGATTGTTTCAGTACAAAAGTGCCTGAC 600
  |
      *      *      *      *      *      *      *      *      *      *
601 TACTCGTTTTACCTTCCGTTGCATGTGATTGCTTATGGCTGGTAAAAATTGGACAATTATGTAGATGTTAAGAAGCTTCTTATTGACATGGGAATAT 700
  |
601 TACTCGTTTTACCTTCCGTTGCATGTGATTGCTTATGGCTGGTAAAAATTGGACAATTATGTAGATGTTAAGAAGCTTCTTATTGACATGGGAATAT 700
  |
      *      *      *      *      *      *      *      *      *      *
701 ACTTCCAAGTTCAGGATGACTATTTGGACTGCTTTGGTGAACAGAGAAAGATTGGAAAGATTGGAACCGATATTGAAGATTTCAAGTGCCTTTGGCTTGT 800
  |
701 ACTTCCAAGTTCAGGATGACTATTTGGACTGCTTTGGTGAACAGAGAAAGATTGGAAAGATTGGAACCGATATTGAAGATTTCAAGTGCCTTTGGCTTGT 800
  |
      *      *      *      *      *      *      *      *      *      *
801 TGTCAAAGCATTAGAGAAGTGTAAAGAGCAAAAACAACATTTATTAGAACACTATGGAAAGGCAATCCTGAAGATGTCGAGAAAGTCAAACGCCTT 900
  |
801 TGTCAAAGCATTAGAGAAGTGTAAAGAGCAAAAACAACATTTATTAGAACACTATGGAAAGGCAATCCTGAAGATGTCGAGAAAGTCAAACGCCTT 900
  |
      *      *      *      *      *      *      *      *      *      *
901 TACAATGATCTGGATCTTCAGGGCGTATTTAAAGAATTGAAAGCAAGAGCTACGAGAAACTGAATACCTCCATTGAAGCTCATCCAACAAGCTGTGC 1000
  |
901 TACAATGATCTGGATCTTCAGGGCGTATTTAAAGAATTGAAAGCAAGAGCTACGAGAAACTGAATACCTCCATTGAAGCTCATCCAACAAGCTGTGC 1000
  |
      *      *      *      *      *
1001 AAGCTGTACTCAAGTCATTTTGGGTAAGATATACAAGAGGCTGAATAA 1050
  |
1001 AAGCTGTACTCAAGTCATTTTGGGTAAGATATACAAGAGGCTGAATAA 1050
  |
      *      *      *      *      *

```

**Figure 13.** Sanger sequencing alignment of *GIFPS* with the corresponding GeneBank GeneBank wild type sequence (*ApE: A plasmid Editor*, 2022)

HSQS

HSQS sanger sequencing.ape from 1 to 1023  
Alignment to  
HSQS.ape from 1 to 1023

Matches():1023  
Mismatches(#):0  
Gaps():0  
Unattempted():0

```

      *      *      *      *      *      *      *      *      *      *
1  ATGGACCGGAACTCGGCTCAGCAACAGCCTGAAAACCTGCTACAAGTATCTCAATCAGACCAGTCGCAGTTTCGCAGCTGTTATCCAGGCCCTGGATGGGG 100
   |
1  ATGGACCGGAACTCGGCTCAGCAACAGCCTGAAAACCTGCTACAAGTATCTCAATCAGACCAGTCGCAGTTTCGCAGCTGTTATCCAGGCCCTGGATGGGG 100
      *      *      *      *      *      *      *      *      *      *

101 AAATGCGCAACGCAAGTGTGCATATTTTATCTGGTTCCTCCGAGCTCTGGACACACTGGAAGATGACATGACCATCAGTGTGGAAAAGAAGTCCCCTGTT 200
   |
101 AAATGCGCAACGCAAGTGTGCATATTTTATCTGGTTCCTCCGAGCTCTGGACACACTGGAAGATGACATGACCATCAGTGTGGAAAAGAAGTCCCCTGTT 200
      *      *      *      *      *      *      *      *      *      *

201 ACACAACCTTCACTCTTCCCTTTACCAACAGACTGGCGGTTTCATGGAGAGCAAGGAGAAGGATCGCCAGGTGCTGGAGGACTTCCCAACGATCTCCCTT 300
   |
201 ACACAACCTTCACTCTTCCCTTTACCAACAGACTGGCGGTTTCATGGAGAGCAAGGAGAAGGATCGCCAGGTGCTGGAGGACTTCCCAACGATCTCCCTT 300
      *      *      *      *      *      *      *      *      *      *

301 GAGTTTAGAAATCTGGCTGAGAAATACCAACAGTGAATGCCGACATTTGCCGAGAGATGGGCATGGGATGGCAGAGTTTTGGATAAGCATGTGACCT 400
   |
301 GAGTTTAGAAATCTGGCTGAGAAATACCAACAGTGAATGCCGACATTTGCCGAGAGATGGGCATGGGATGGCAGAGTTTTGGATAAGCATGTGACCT 400
      *      *      *      *      *      *      *      *      *      *

401 CTGAACAGGAGTGGGACAAGTACTGCCACTATGTTGCTGGGCTGGCGAATTGGCCTTTCCCGCTTTTCTCAGCCTCAGAGTTTGAAGACCCCTTAGT 500
   |
401 CTGAACAGGAGTGGGACAAGTACTGCCACTATGTTGCTGGGCTGGCGAATTGGCCTTTCCCGCTTTTCTCAGCCTCAGAGTTTGAAGACCCCTTAGT 500
      *      *      *      *      *      *      *      *      *      *

501 TGGTGAAGATACAGAACGTGCCAATCTATGGGCCGTGTTTCTGCAGAAAACAACATCATCCGTGACTATCTGGAAGACCAGCAAGGAGGAGAGAGTTC 600
   |
501 TGGTGAAGATACAGAACGTGCCAATCTATGGGCCGTGTTTCTGCAGAAAACAACATCATCCGTGACTATCTGGAAGACCAGCAAGGAGGAGAGAGTTC 600
      *      *      *      *      *      *      *      *      *      *

601 TGGCCTCAAGAGGTTTGGAGCAGGTATGTTAAGAAGTTAGGGGATTTGCTAAGCCGGAGAATATTGACTTGGCCGTGCAGTGCCTGAATGAACITATAA 700
   |
601 TGGCCTCAAGAGGTTTGGAGCAGGTATGTTAAGAAGTTAGGGGATTTGCTAAGCCGGAGAATATTGACTTGGCCGTGCAGTGCCTGAATGAACITATAA 700
      *      *      *      *      *      *      *      *      *      *

701 CCAATGCACCTGCACCACATCCAGATGTCATCACCTACCTTTCGAGACTCAGAAACAGAGTGTGTTAACTTCTGTGCTATTCCACAGGTGATGGCCAT 800
   |
701 CCAATGCACCTGCACCACATCCAGATGTCATCACCTACCTTTCGAGACTCAGAAACAGAGTGTGTTAACTTCTGTGCTATTCCACAGGTGATGGCCAT 800
      *      *      *      *      *      *      *      *      *      *

801 TGCCACTTTGGCTGCCTGTTATAATAACCAGCAGGTGTTCAAAGGGGCAAGTGAAGATTGGAAGGGCAAGCAGTACCCCTGATGATGGATGCCACCAAT 900
   |
801 TGCCACTTTGGCTGCCTGTTATAATAACCAGCAGGTGTTCAAAGGGGCAAGTGAAGATTGGAAGGGCAAGCAGTACCCCTGATGATGGATGCCACCAAT 900
      *      *      *      *      *      *      *      *      *      *

901 ATGCCAGCTGTCAAAGCCATCATATATCAGTATATGGAAGAGATTTATCATAGAATCCCCGACTCAGACCCATCTTCTAGCAAAAACAAGGCAGATCATCT 1000
   |
901 ATGCCAGCTGTCAAAGCCATCATATATCAGTATATGGAAGAGATTTATCATAGAATCCCCGACTCAGACCCATCTTCTAGCAAAAACAAGGCAGATCATCT 1000
      *      *      *      *      *      *      *      *      *      *

1001 CCACCATCCGGACGCAGAACTAG 1023
   |
1001 CCACCATCCGGACGCAGAACTAG 1023
      *      *

```

Figure 14. Sanger sequencing alignment of HSQS with the corresponding GeneBank GeneBank wild type sequence (ApE: A plasmid Editor, 2022)

TSQS

TSQS Sanger sequencing.ape from 1 to 1080  
Alignment to  
TSQS.ape from 1 to 1080

Matches():1080  
Mismatches(#):0  
Gaps():0  
Unattempted(.):0

```

      *      *      *      *      *      *      *      *      *      *
1  ATGCGTGTGGTGTAAACCCACCTATGATTCAGATGGTTTTAGAAAATCCGGTTAGCGTTAAAGACGCAGAGTGCTTCTGTCAAGAAATCTGCCGCAAG 100
   |
1  ATGCGTGTGGTGTAAACCCACCTATGATTCAGATGGTTTTAGAAAATCCGGTTAGCGTTAAAGACGCAGAGTGCTTCTGTCAAGAAATCTGCCGCAAG 100
      *      *      *      *      *      *      *      *      *      *

101 TTTCTCGCACCTTCGCACTGTCCATCCGTTTCTGCGGGTAACTGGCCGTGCAGTTCTGGTGGCTATCTGATCTGCCGATTGCGGATACCGTGA 200
   |
101 TTTCTCGCACCTTCGCACTGTCCATCCGTTTCTGCGGGTAACTGGCCGTGCAGTTCTGGTGGCTATCTGATCTGCCGATTGCGGATACCGTGA 200
      *      *      *      *      *      *      *      *      *      *

201 AGATGACCCGGTGGCGAGCATCGCGGCCAAAACCGCTTGTCTGGACATTTGCTTCAGTGTCTTGATAGCCAGCGCTGGCAAACAGCTACGGTGA 300
   |
201 AGATGACCCGGTGGCGAGCATCGCGGCCAAAACCGCTTGTCTGGACATTTGCTTCAGTGTCTTGATAGCCAGCGCTGGCAAACAGCTACGGTGA 300
      *      *      *      *      *      *      *      *      *      *

301 GCGCGTGGCGTGCAGGGCGAGCCGGCTCATGTCACGTTGCAAGCACACCGGCATCGTTTTTACCCTGTACCGCAGCGCTGCCGCTACCTCCAGCAAC 400
   |
301 GCGCGTGGCGTGCAGGGCGAGCCGGCTCATGTCACGTTGCAAGCACACCGGCATCGTTTTTACCCTGTACCGCAGCGCTGCCGCTACCTCCAGCAAC 400
      *      *      *      *      *      *      *      *      *      *

401 ACGTCCAAACGTTGGGTGAGCGAGATGGTGCACGGTATGAAGAAATTCATTAAATCTGTATCCGAATGGTATTCGCATCCAGACCCGCGGAATACAAAGA 500
   |
401 ACGTCCAAACGTTGGGTGAGCGAGATGGTGCACGGTATGAAGAAATTCATTAAATCTGTATCCGAATGGTATTCGCATCCAGACCCGCGGAATACAAAGA 500
      *      *      *      *      *      *      *      *      *      *

501 ATATTGTTATTACGTCGCGGGCACGGTGGTTACTTGTGACGGATCTGTGGCATGAGCACAGCCGAGCATGGTGGGACGAGTATCAAGTCCGCTG 600
   |
501 ATATTGTTATTACGTCGCGGGCACGGTGGTTACTTGTGACGGATCTGTGGCATGAGCACAGCCGAGCATGGTGGGACGAGTATCAAGTCCGCTG 600
      *      *      *      *      *      *      *      *      *      *

601 AAGCGTTGTGCAGCGTTGCGCGAGGCCCTGCAGACGGTGAATATCTTGAAGACATTGCTTGGGATGCGGAACACGAGAACAGCATCTACATCCCGAATG 700
   |
601 AAGCGTTGTGCAGCGTTGCGCGAGGCCCTGCAGACGGTGAATATCTTGAAGACATTGCTTGGGATGCGGAACACGAGAACAGCATCTACATCCCGAATG 700
      *      *      *      *      *      *      *      *      *      *

701 AGAGCCTGATTCTGCAAGGTAGCAGCCATCAGACGCTGCTGTCTGCGGAGCATCTGCAACAGAACCACGAGCGATCAAAGAATGATTGCTCTGGCCG 800
   |
701 AGAGCCTGATTCTGCAAGGTAGCAGCCATCAGACGCTGCTGTCTGCGGAGCATCTGCAACAGAACCACGAGCGATCAAAGAATGATTGCTCTGGCCG 800
      *      *      *      *      *      *      *      *      *      *

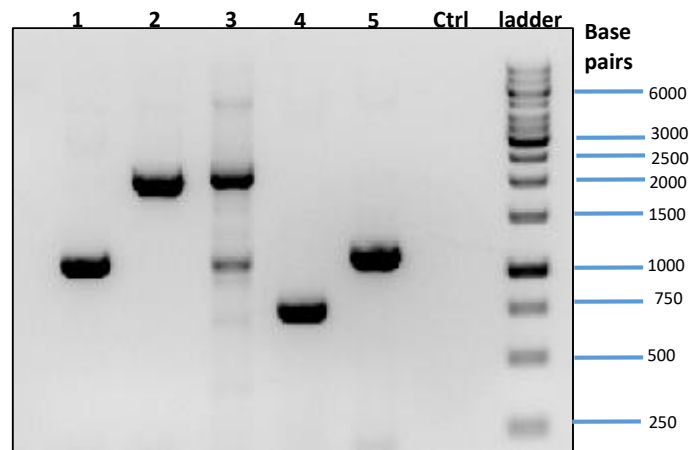
801 GCACGATCTGGACGAAGCCCAAGCGTATCTGCTGTCCGTGCCAAGGCCGCGATTCCGATTGCGCTGTTTGGCTGCTGCCGCTGCTGTTGCATACGCG 900
   |
801 GCACGATCTGGACGAAGCCCAAGCGTATCTGCTGTCCGTGCCAAGGCCGCGATTCCGATTGCGCTGTTTGGCTGCTGCCGCTGCTGTTGCATACGCG 900
      *      *      *      *      *      *      *      *      *      *

901 ACTCTGCGGAGCTGACGCACTCGACCGCATGCTGCAACCTGGCGGTGGTGTAAAGATCAGCCGTGCAGAGGTCAAAAGCCTCATGGTATGGGTCGCG 1000
   |
901 ACTCTGCGGAGCTGACGCACTCGACCGCATGCTGCAACCTGGCGGTGGTGTAAAGATCAGCCGTGCAGAGGTCAAAAGCCTCATGGTATGGGTCGCG 1000
      *      *      *      *      *      *      *      *      *      *

1001 TGAGCATCTGAGCAACCACGGCCTGCGTTGGTTGATCGGCCAGGTTCCGCCAGAAGCCGTTTATCCTGGGTGGTCTGTAG 1080
   |
1001 TGAGCATCTGAGCAACCACGGCCTGCGTTGGTTGATCGGCCAGGTTCCGCCAGAAGCCGTTTATCCTGGGTGGTCTGTAG 1080
      *      *      *      *      *      *      *      *      *      *

```

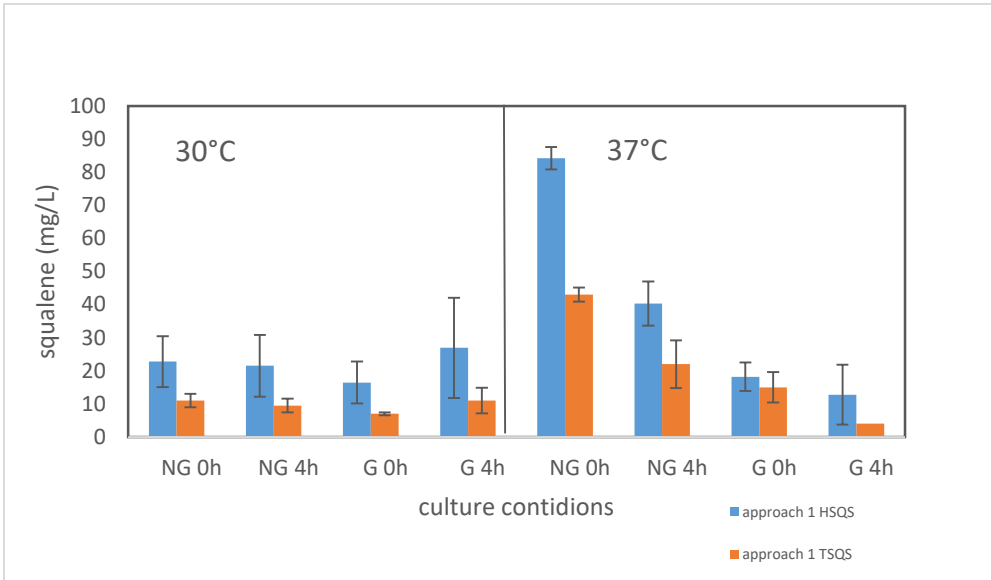
Figure 15. Sanger sequencing alignment of TSQS with the corresponding GeneBank wild type sequence (ApE: A plasmid Editor, 2022)



**Figure 16.** Confirmation by PCR of the genes inserted in the BL21(DE3) *E. coli* strain. Lane 1, human *SQS*; lane 2, *DXS1*; lane 3, *DXS1* (using a different pair of primers), lane 4, *IDI1*; lane 5, *FPS*; Ctrl, control with water instead of DNA; Ladder 1 Kb.

#### 2.4.1.3 Characterization of squalene production in the engineered strains

We examined squalene production from the *E. coli* strains expressing the heterologous genes from plants and the two squalene synthases genes (*hSQS* and *tSQS*). The strains were cultured in 250mL LB media into a 500mL shake flask with different culture conditions: induction with IPTG at time zero and four hours later, containing or not glucose and grown at 30°C or 37°C. As shown in **Figure 17**, the highest squalene titer after 24 hours of culture was reached at 37°C, without glucose supplementation and with IPTG induction (2mM) at beginning of the growing culture (time zero). The MEP-*hSQS1* strain reached a titer of 90 mg/L and the MEP-*tSQS1* strain produced 43 mg/L of squalene. In all the evaluated scenarios, a greater accumulation of squalene was observed in the strain with the human squalene synthase gene compared with the gene of the cyanobacterium *Thermosynechococcus elongatus*.



**Figure 17.** Squalene production in the approach 1 with the MEP-hSQS and MEP-tSQS *E. coli* strains at different growth conditions: temperature at 30°C and 37°C; addition or not of glucose G or NG; IPTG induction (2mM) at start or after 4h of incubation 0h or 4h.

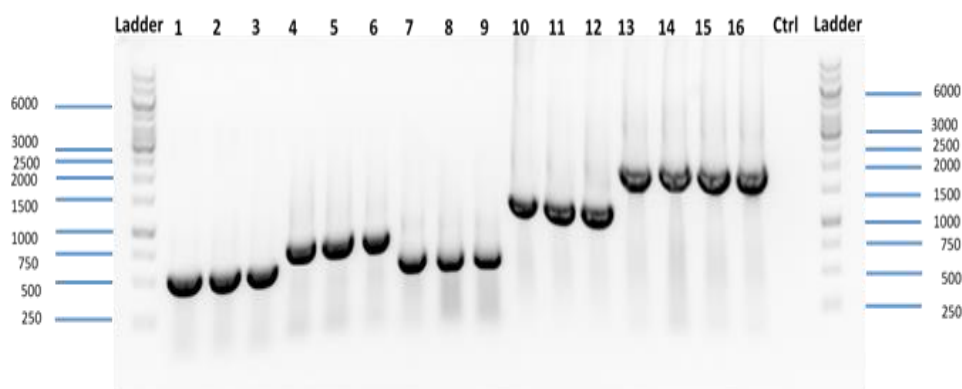
## 2.4.2 Approach 2: Generation of *E. coli* strains overexpressing endogenous MEP-biosynthetic genes and heterologous SQS genes.

### 2.4.2.1 Plasmid constructs

In order to build-up a squalene-producing strain overexpressing MEP pathway genes from *E. coli* mostly were subcloned in intermediate plasmids and then transferred into pET-32a(+) through a Gibson assembly reaction (**Table 13**). Then the four plasmid vectors genes from the MEP metabolic pathway were used for into the *E. coli* BL21(DE3) strain, resulting in the MEP-hSQS2 or MEP-tSQS2 strains. The presence of the genes in the plasmids was confirmed by *E. coli* colony PCR (**Figure 18**).

**Table 13.** Vector constructs for MEP pathway overexpression

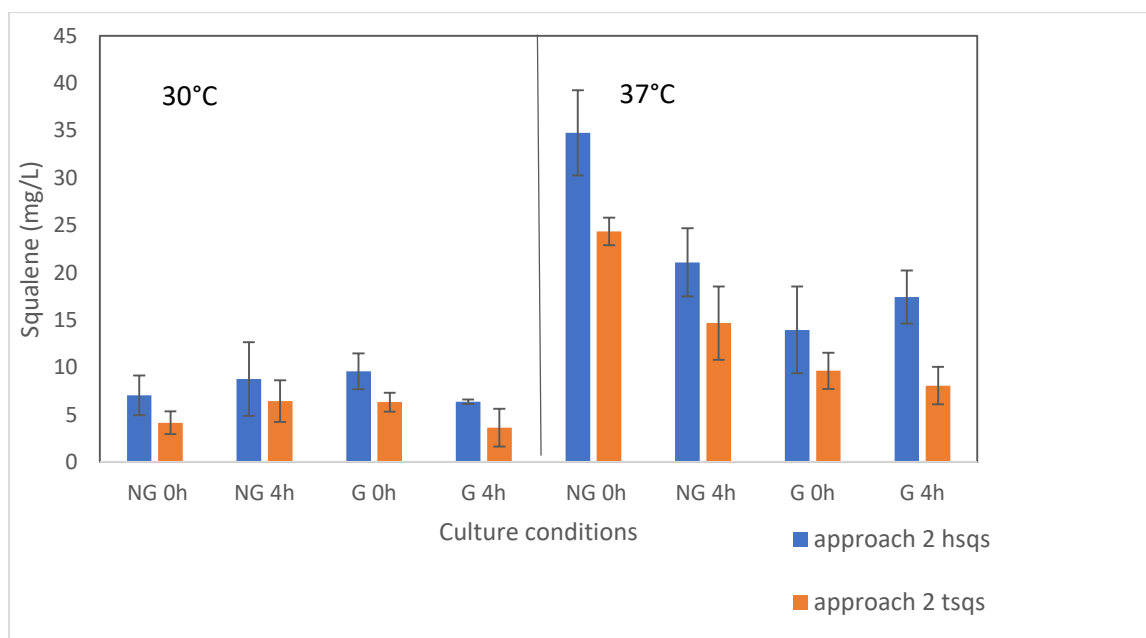
Construct	Genes	Intermediate plasmid	Restriction enzymes for cloning	Backbone vector for Gibson assembly
Vector 1	<i>EcDXS</i> <i>EcDXR</i>	p326 pHor-P	KpnI/EcoRI XbaI/SacI	
Vector 2	<i>EcispD</i> <i>EcispE</i> <i>EcispF</i>	pRP5 pHor-P p326	KpnI/EcoRI XbaI/SacI KpnI/EcoRI	
Vector 3	<i>EcispG</i> <i>EcispH</i> <i>EcFldA</i>	pRP5 pHor-P pHor-P	KpnI/EcoRI XbaI/SacI XbaI/EcoRI	pET-32a(+)
Vector 4	<i>ScERG20</i> <i>EcIDI</i> <i>hSQS</i> or <i>tSQS</i>	pUC8 pUC8 pUC18m-hsqs pJE404- <i>tsqs</i>	EcoRI/XhoI EcoRI/HindIII EcoRI/HindIII	



**Figure 18:** Colony PCR of the key genes inserted in the engineered *E. coli* BL21 (DE3) strain overexpressing MEP genes. Ladder 1kb; lane 1-3: *FLDA*; lane 4-6: *lspD*; lane 7-9: *lspF*; lane 10-12: *lspG*; lane 13-16: *DXS*

#### 2.4.2.2 Squalene production in *E. coli* strain overexpressing MEP genes and expressing SQS

Squalene content was measured in the MEP-hSQS2 or MEP-tSQS2 strains (**Figure 19**). The overexpression of the 9 endogenous genes from *E. coli*, assembled with *ERG20* from *S. cerevisiae* (that encodes FPS) and the squalene synthase, reached a titer of 34.7 mg/L with human SQS and 24.3 mg/L with *T. elongatus* SQS. After 24 hours of culture at 37°C, the highest values for squalene production were reached without glucose supplementation and with the addition of the inducer (2 mM IPTG) at the start of the bacterial growth (time 0), for both engineered strains (*MEP-hSQS2* or *MEP-tSQS2*). The data also shows that with the experiments carried out at a 30°C, lower values of squalene concentration were obtained. The same pattern was observed in approach 1.



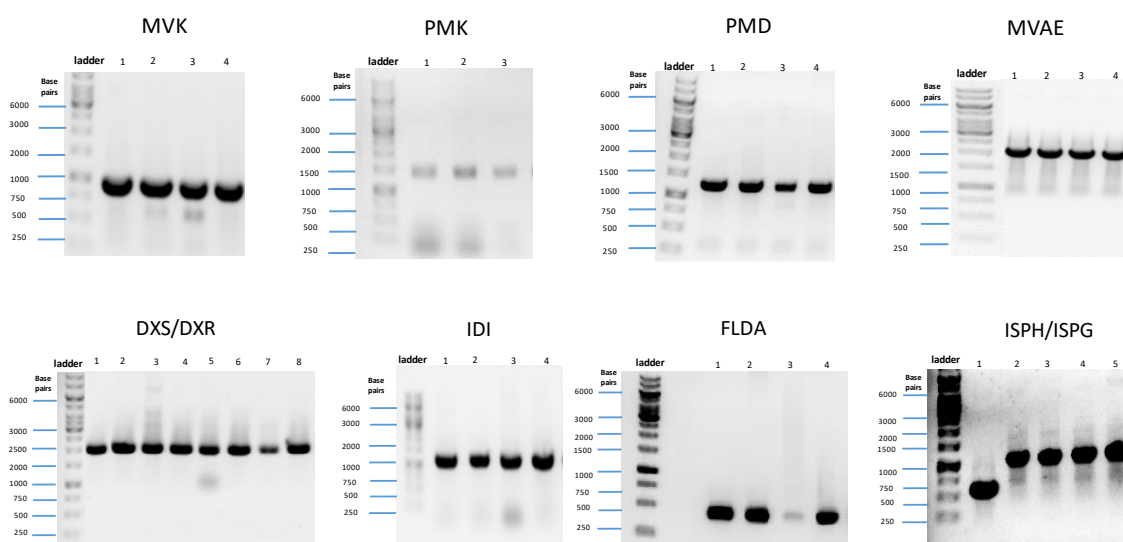
**Figure 19.** Squalene production in approach 2 from the engineered hSQS and tSQS strains (*MEP-hSQS2* and *MEP-tSQS2* stains) at different growth conditions: Temperature at 30°C and 37°C; addition or not of glucose; G or NG; Induction with IPTG at time 0 or after 4h of incubation; 0h or 4h.



### 2.4.3 Approach 3: Expression of heterologous squalene synthase genes in the *E. coli* strain engineered with the dual MEP/MVA pathway.

#### 2.4.3.1 Generation of the Dual-hSQS3 and Dual-SQS3 strains

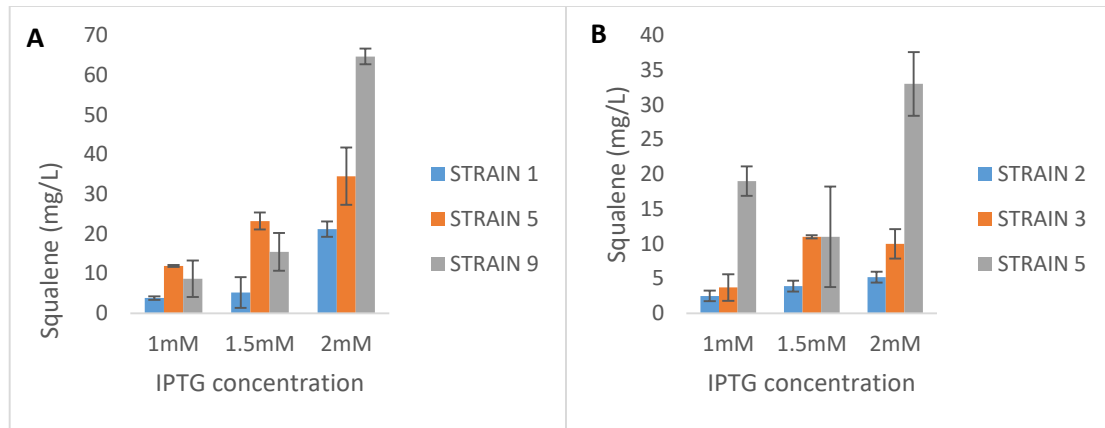
The strain 1440 expressing MEP/MVA pathway, called dual pathway, were constructed by Yang et al (2016). The dual pathway overexpresses the genes *DXS*, *DXR*, *ISPG*, *ISPH*, *FLDA* and *IDI* from the MEP pathway and *MVAE*, *MVAS*, *MK*, *PMK* and *PMD* from the MVA pathway. A PCR analysis of the dual pathway genes carried out in individual engineered *E. coli* colonies is shown in **Figure 20**. Following the confirmation of the presence of the key genes involved in terpenoid production, we performed a co-transformation with squalene synthase genes yielding to the strain Dual-hSQS3 or Dual-SQS3.



**Figure 20.** PCR analysis of the dual pathway genes carried out in individual *E. coli* colonies.

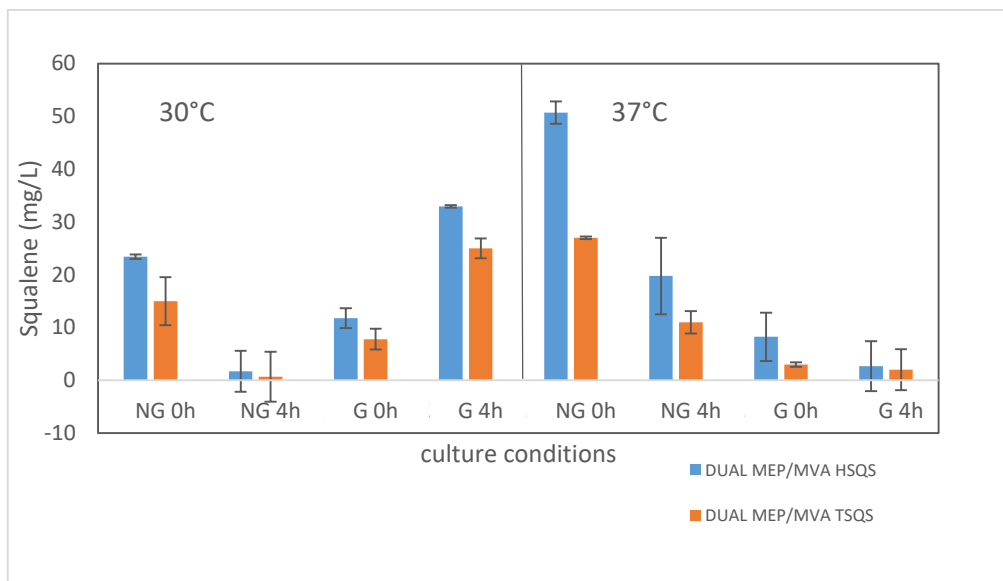
#### 2.4.3.2 Characterization of squalene production from the engineered *E. coli* dual MEP/MVA pathway with SQS genes

In order to choose the finest inducer concentration and the *E. coli* strain that expressed the dual MEP/MVA pathway and the two versions of squalene synthase, a preliminary characterization of squalene production was performed. We tested three different IPTG concentrations (1 mM, 1.5 mM and 2 mM) (**Figure 21**). The highest values of squalene accumulation were observed at an IPTG concentration of 2 mM, in the strain number 9 for Dual MEP/MVA *hSQS* and strain number 5 for Dual MEP/MVA *tSQS*.



**Figure 21.** Evaluation of the optimal IPTG concentration tested in three different *E. coli* strains containing the dual MEP/MVA pathway and a SQS gene. A) Squalene production in Dual MEP/MVA *hSQS* strain. B) Squalene production in Dual MEP/MVA *tSQS* strain.

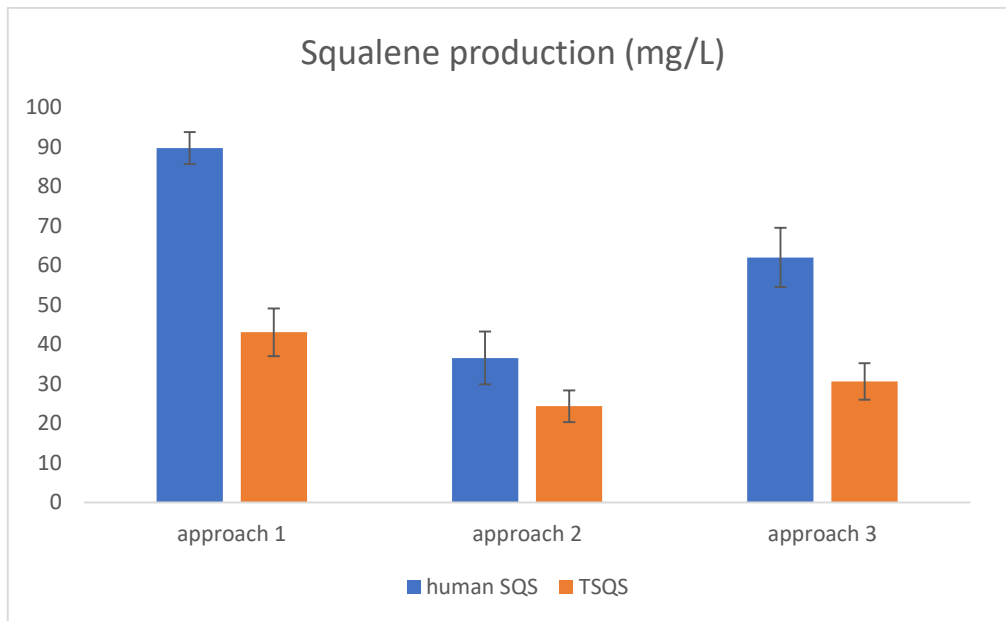
Furthermore, two different temperatures of incubation were assessed because some genes (*IDI*, *PMK* and *PMD*) come from *S. cerevisiae*, which has an optimal growth at 30°C. We determined the optimal temperature for the corresponding enzymes at 30°C and 37°C. Due to the presence of numerous genes involved in the Dual *hSQS* and Dual *tSQS* strain, the carbon source (glucose) and the transcriptional inducer (IPTG) concentration can be rate-limiting factors to take into account (Yang et al., 2016 and Katabami et al., 2015). For this reason, we performed cultures adding or not glucose and adding IPTG at the beginning or after 4 hours of incubation. The production of squalene in the *E. coli* Dual *hSQS* and Dual *tSQS* was measured after 24 hours of culture. **Figure 22**, shows the squalene production comparing different conditions. As occurred in the approach 1 and 2, the best culture conditions for squalene accumulation in the Dual MEP/MVA were reached at 37°C, without glucose supplementation and IPTG induction at time 0.



**Figure 22.** Squalene production in dual MEP/MVA *HSQS* or *TSQS* at different growth conditions: Temperature at 30°C and 37°C; addition or not of glucose G or NG; Time IPTG at start 0h or after 4h of incubation.

#### 2.4.3.3 Comparison of the squalene production in the three approaches

The squalene production was compared among different engineered *E. coli* strains achieved under three approaches (**Figure 23**). The data showed that in all strategies, the strains harboring the human squalene synthase gene accumulated more squalene than the ones harboring *T. elongatus* squalene synthase gene. The same result was reported by Katabami et al., 2015, but Qiao et al., 2019 reported that *tSQS* showed higher squalene content than *hSQS* in *E. coli*. We obtained the highest value of squalene production (90 mg/L) with the strain developed in the approach 1. Squalene content was lower with the strains from the approach 3 and 2, respectively, 62 mg/L and 35mg/L.



**Figure 23.** Squalene production of the *E. coli* strains tested with two different squalene synthases (*human SQS* and *tSQS*) designed through three different synthetic biology approaches. The bars indicate the average of three samples, and the error bars indicate the standard deviation.

## 2.5 Discussion

Different strategies for genetically engineering microorganisms in order to produce compounds of interest have been extensively reported in the literature (Hussain et al., 2022; Gohil et al., 2021; Ma et al., 2021; Rugbjerg et al., 2020; Nielsen and Keasling 2016). One of the most common strategies is to introduce genes belonging to different organisms (plants, animals, bacteria and fungi), combine them within a specific metabolic pathway and express them in a microbial cell host, in this way it is possible to produce a wide variety of compounds, from basic chemical products to pharmaceutical or nutraceutical products (Alonso-Gutiérrez et al., 2015). However, metabolic engineering to obtain specific compounds is not an easy task and requires different tools to achieve the desired yield. For that reason, in this work different strategies were addressed to design genetically an *E. coli* strain capable of producing squalene and serving as a platform for the production of other terpenes of interest.

In the first approach, an *E. coli* strain was designed harbouring MEP pathway genes from the plant plastid. As suggested by various authors (Burke and Croteau, 2002, Alonso-Gutiérrez et al 2015 and Katabami et al., 2015) for optimal enzyme expression in bacteria as host, plastid targeting must be eliminated. In this approach, the plastid-directed N-terminal sequences of the enzymes-encoding *DXS* (and its 1,2,3 isoforms) and *IDI 2* genes were removed. Also, the squalene synthase genes from human and *Thermosynechococcus elongatus* were used in their truncated form as described by Katabami et al., (2015).

As described in the results section, in approach 1 the *E. coli* strain was constructed through a plasmid-based system, the genes *DXS1*, *IDI1*, *FPS* (from plants) and *SQS* (human or *T. elongatus*) were transformed in the pET-32a(+) plasmid. This plasmid contains the T7 promoter, which is characterized by being a strong promoter widely used in different strategies in metabolic engineering in *E. coli* and especially for the production of isoprenoids (Koma et al., 2012; Alonso-Gutierrez et al., 2018).

Additionally, both in approach 1 and in the other two approaches, the accumulation of squalene with two different squalene synthases, from humans and from the bacterium *T. elongatus*, was evaluated. The results in flask culture showed that the human *SQS* produces more squalene than that of bacterial origin, 90 and 43 mg/L respectively, which agrees with the results reported in the work of Katabami et al., 2015, who designed a strain of *E. coli* expressing a chimeric metabolic pathway of MVA and MEP, which reached a production of 230 mg/L for *hSQS* and 150 mg/L for *tSQS*.

Regarding the second approach, a strain was designed by overexpressing 10 genes from the *E. coli* MEP pathway and one the two squalene synthases tested in the previous strategy. In this approach, the Gibson assembly tool was used, which allows to assembly up to 15 DNA fragments or up to 32 kb in single-tube reaction (Gibson et al., 2010). However, despite the ease of the procedure for vector construction, the results of

squalene accumulation in this approach were the lowest (35 mg/L and 25 mg/L for *hSQS* and *tSQS* respectively), compared to the other two approaches. This is possibly due to inconveniences in the growth and maintenance of the resulting strain. In this way, both the overexpression of native genes and the introduction of any heterologous DNA construct represent a challenge for the bacterial cell, since it affects the fine balance of the resources that they allocate for survival (Weber et al., 2011; Slusarczyk et al., 2012).

The above may be due to different factors as evidenced by various authors (Borkowsky et al., 2018; Alonso-Gutierrez et al., 2013; Shachrai et al., 2010; Roskov et al., 2004). Among them, it is worth highlighting the competition that exists between the endogenous genes that control the cellular processes in the host cell and the genes that encode proteins through synthetic constructs, due to the limited resources in the cell used for gene expression (Borkowsky et al., 2018). For example, those involved in transcription and translation (Borkowsky et al., 2016), in which the extra artificial load due to gene overexpression causes slow bacterial growth and low production of the desired protein, a phenomenon known as “burden” (Shachrai et al., 2010). Accordingly, the maintenance of multiple plasmids increases the metabolic burden of the cell from DNA, RNA, and protein synthesis, as well as the number of antibiotic resistance proteins that the cell must produce, leading to low production of the desired compound (Roskov et al., 2004). Moreover, having more plasmids, which are in the cell temporarily, can generate instability and be lost at the time of cell division (Gama et al., 2020; Csörgő et al., 2012; Kumar 1991). This phenomenon is known as loss-of-function mutation, in which the bacteria escape the metabolic load imposed by the expression of genes that are not essential for its normal function. This strategy allows the mutant cell to overcome and increase the growth rate in a short time (Sleight and Sauro 2013).

According to studies carried out by Liu et al., 2017, the overexpression of the *DXS*, *DXR*, *ispD*, *ispF*, *ispH*, *ispE* and *IDI* genes decreased the production of squalene and also reduced the catalytic efficiency of the *DXS* and *IDI* enzymes in the production of squalene. On the other hand, Zhou et al., 2012, reports that the overexpression of the *DXS*, *IDI* and *ispF* genes can lead to the overproduction and accumulation of the compound methylerythritol cyclodiphosphate (MECPP) in the cell. In this investigation, it was discovered that MECPP effluxed when certain enzymes of the pathway were overexpressed, which demonstrated the existence of a new branch of the pathway that competes in the metabolism of MEP. Finally, several authors report that excessive overexpression of the *DXS*, *IDI*, *ispD*, and *ispF* genes can significantly inhibit the production of isoprenoids (Ajikumar et al., 2010; Yoon et al., 2006; Martin et al., 2003; Kim and Keasling et al., 2001).

In the third approach, the two squalene synthases (*hSQS* and *tSQS*) used in the previous two approaches were co-transformed independently with the previously designed dual MEP/MVA *E. coli* strain by Yang et al., 2016. This strain overexpresses six MEP pathway genes (*DXS*, *DXR*, *IspG*, *IspH*, *FLDA* and *IDI*) and heterologously expresses the complete

MVA pathway with 6 genes (*MvaE*, *MvaS*, *MK*, *PMK*, *PMD*, *IDI*). The *MK*, *PMK*, *PMD*, and *IDI* genes of the MVA pathway and the *DXS*, *DXR*, and *IDI* genes of the MEP pathway were integrated into the chromosome to ensure stability in the time of their expression. Chromosome integration generates an advantage compared to plasmid-based system approaches as they are not temporarily in the cell (Borkowsky et al., 2018, Alonso-Gutierrez et al., 2018).

The Dual MEP/MVA strain was originally used for isoprene production in *E. coli*, obtaining in that research a 20-fold increase in the yield of isoprene compared to the strains that overexpress only the MEP or MVA pathways (Yang et al., 2016). To the best of our knowledge, there are no reports of this genetically engineered strain used in other research than for the production of terpenes. Thus, this work reports for the first time, in the production of squalene, the use of an *E. coli* strain designed to completely overexpress the two metabolic pathways. As a result, the squalene accumulation obtained was 62 mg/L and 30 mg/L for *hSQS* and *tSQS*, respectively, lower than that obtained in approach 1, but higher than approach two. Those results showed that the use of the engineered dual MEP/MVA pathway strain is promising for the production of terpenes.

The studies by Yang et al., (2016) showed that the combined use of both routes supplies the energy needs of the MEP and the carbon needs of the MVA. Thus, the amount of the cofactor NAD(P)H that is required for isoprene synthesis via MEP can be provided by the MVA pathway. On the other hand, the authors speculated that the flow of carbon through the MEP pathway causes a reduction in the size of the intracellular pool of specific metabolites, which may alleviate the feedback inhibition of MVA pathway enzymes such as mevalonate kinase (MK). In addition, the aim of the dual MEP/MVA pathway is to enhance the precursor supply such as IPP (isopentenyl diphosphate) and DMAPP (dimethylallyl diphosphate). Both IPP and DMAPP are the necessary building blocks for the terpenoid production (Ma et al. al., 2021). However, due to its intrinsic complexity, few studies have been carried out using a strain that simultaneously overexpresses both metabolic pathways (Yang et al., 2016; Morrone et al., 2010, Yoon et al., 2007).

Some studies have shown that non-endogenous IPP production could be achieved in *E. coli* by optimizing and balancing the native MEP pathway and the exogenous MVA pathway (Chang et al., 2013; Hernandez-Arranz et al., 2019). This requires that a delicate balance can be achieved in the expression level of genes in the non-native metabolic pathway (Alonso-Gutiérrez et al., 2017) because any imbalance would cause a decrease in the production of the desired compound. Additionally, in the literature it is reported that the addition of a foreign metabolic pathway in *E. coli* can lead to a heavy metabolic load for the bacterial strain and consequently lead to a decrease in the cell growth rate (Grob et al., 2021; Yang et al., 2019; Alonso-Gutiérrez et al., 2018; Meng et al., 2011).

For the three strategies, a total of 24 different scenarios for squalene production in *E. coli* were generated and analysed. Two temperatures (30°C and 37°C), addition or not of glucose and induction of gene expression with IPTG (2mM) at the beginning of the culture and 4h later were evaluated. With these scenarios, the aim was to find the best culture condition that would allow the ideal environment for the engineered strains to grow and produce the desired compound. In all strategies, the highest squalene values were observed at a temperature of 37°C. These results agree with those obtained for similar strategies (2 and 3) carried out in other studies for squalene synthesis (Katabami et al., 2015; Furubayashi et al., 2014; Ghimire et al., 2009), but differ from what was observed by Liu et al., (2017), who only reported the accumulation of squalene at 30°C. In a study performed with heterologous expression of the MVA pathway in *E. coli*, it has been reported that the growth temperature of this microbe affects the production of limonene and bisabolene (Alonso-Gutierrez et al., 2015). In this way, at a lower temperature (30°C) of cultivation in shake flask culture and bioreactor, the production of certain terpenes is favoured. However, most of the work carried out on engineered *E. coli* for the production of squalene has been carried out at 37°C in shake flask cultures.

Studies accomplished by Alonso-Gutierrez et al., 2015 suggest that genetically engineered strains may be sensitive to the amount of expression inducer and the time at which the inducer is added to the culture in relation to culture cell density measured by optical density at 600 nm. Consequently, the inducer concentration and the time of the IPTG addition were correlated with the accumulation of squalene in the three approaches. After testing IPTG concentrations of 1, 1.5 and 2 mM, the highest squalene production was reached at a concentration of 2 mM IPTG, for all three approaches. In the literature it has been reported, a wide range of IPTG concentrations varying between 0 and 2 mM, used to induce gene expression under the control of lac promoters in the production of terpenes in *E. coli*, (Yin et al., 2021; Xu et al., 2019; Wu et al., 2018; Yang et al., 2016) depending on the metabolic engineering strategy used.

The foregoing may depend on the strategy used and the number of genes which are under the control of IPTG inducible promoters in the designed strains. However, some authors do not report the addition of IPTG to the culture, even using IPTG-inducible promoters (Katabami et al., 2015; Furubayashi et al., 2014). In another study by Dvorak et al., (2015) it is reported that IPTG is not harmless to *E. coli* and increases (exacerbates) the toxicity of a contaminating compound, TCP (1,2,3-trichloropropane) in a *E. coli* BL21 (DE3) strain, engineered to express 5 enzymes of a synthetic metabolic pathway and convert TCP to glycerol. The authors conclude that replacing IPTG with lactose, the natural inducer of the *lac* operon, can alleviate the metabolic burden of the engineered strain, and could be an economical alternative if the process is scaled up to an industrial level. In contrast, several authors point out the IPTG advantages as an inducer of gene expression in heterologous constructs. One of these advantages is that IPTG is synthetic and consequently is not metabolizable by the cell and can remain for a long time fulfilling



the function as an inducer (Fernández-Castané et al., 2012; Meng et al., 2020; Wu et al., 2018; Olaofe et al., 2010).

Regarding the time of addition of IPTG to the culture, our results suggest that at the beginning of the culture in a shake flask culture (0 hours), a greater accumulation of squalene is reached than an addition at 4h. These results are in accordance with what was evidenced by Vaccari et al., (2022) who observed that early induction (time 0) with IPTG 0.5 mM in the *E. coli* strain BL21 (DE3) presented a better result in the gene expression of *GFP* with respect to cell growth rate and yield of the protein than adding IPTG two hours later. Terol et al., (2021) also evidenced that an early induction with IPTG increased the expression of the protein. This seems to be related to the biomass of the cell culture since the greater the number of cells that consume the inducer, the lower the effective amount that each cell incorporates into its cytoplasm (Fernández-Castané 2012). Thus, a very precise balance is required between the active biomass and the amount of sufficient inducer, since the lack or excess of the inducer can affect the expression of heterologous genes and consequently the yield of the desired product (Donovan et al., 1996, Fernández-Castané 2012, Bhatwa et al., 2021).

In the three strategies were observed that the highest squalene production was achieved in strains grown in shake flask cultures without glucose addition. In other researches for squalene production such as those carried out with similar strategies to those presented in this work (Katabami et al., 2015, Liu et al., 2017, Furubayashi et al., 2014) there was not reported addition of glucose, even harbouring IPTG-inducible promoters. In contrast, the authors who designed the *E. coli* dual MEP/MVA pathway strain (Yang et al., 2016) added glucose in the culture medium for isoprene production both in shake flask and bioreactor. However, none of these studies compared culture conditions in the presence or absence of glucose getting the doubt about the effect of glucose on the heterologous expression of proteins for the production of isoprenoids. However, in the literature it has been described the mechanism of action of the *lac operon* at the cellular level at high glucose concentrations for foreign protein expression under the control of *lacI* promoters (Pierce 2014; Lewis 2013; Brown et al., 2008; Donovan 1996).

In this way, when the intracellular concentration of glucose is high, the levels of cyclic AMP (cAMP) decrease and the formation of the complex with the catabolite activating protein (CAP), cAMP-CAP does not occur. This complex is required to help to bind efficiently RNA polymerase with DNA, consequently, if this complex is not formed, it results in low rates of transcription of the structural genes *lacZ*, *lacY* and *lacA*, which encode the enzymes:  $\beta$ -galactosidase, permease, transacetylase (Pierce 2014; Lewis 2013). In several studies, the addition of glucose has been reported as a mechanism to reduce the rate of protein expression through its effect as a catabolic repressor on transcription (Lee and Keasling 2005; Lee et al., 2005; Grossman et al., 1998.; Donovan et al., 1996). In our strains designed using the three strategies, the addition of glucose

to the culture medium was not required to achieve a greater accumulation of squalene, which represents an advantage in terms of practicality and costs.

As mentioned above, the expression of a foreign protein which does not occur naturally in the host cell generates an extra burden on it, which translates into a decreased ability to grow under specific conditions (or in a controlled environment) (Borkowski et al., 2016). In agreement, our results suggest that as more metabolic load is present with heterologous genes inserted in one or several plasmids in the cell, as lower production of squalene is obtained. Thus, apart from the two squalene synthases (Human or *T. elongatus*) that are common in our three strategies, in approach 2 that overexpressed 10 genes, transiently in the cell, the lowest squalene values were obtained, followed by approach 3, which overexpressed 5 genes that were not integrated into the *E. coli* genome. The results showed that the approach 1, which only has 3 genes inserted from the plant MEP pathway (*DXS*, *IDI*, *FPS*), was the strategy in which the greatest accumulation of squalene occurred. This agrees with the observations made by (Gorochoowski et al., 2014), where it is evident that the metabolic load influences the production of compounds by the heterologous expression of proteins (Borkowski et al., 2016).

To the best of our knowledge, this work reports for the first time in *E. coli* the heterologous expression of the genes *DXS* and *IDI* from rice and *FPS* from *Gentiana lutea*. The engineered strain harbouring plant genes resulted in an accumulation of squalene of 90 and 43 mg/L for *hSQS* and *tSQS*, respectively. Although there are no reports in the literature on the heterologous insertion of genes from rice and *G. lutea* in *E. coli* for squalene production, our laboratory has been already successfully tested the efficiency of these genes for carotenoid production. (Zhu et al., 2021; Jin et al., 2021; Bai et al., 2014, 2016; Cervantes et al., 2006; Zhu et al., 2003).

As mentioned above, in strategy 2 was obtained an accumulation of squalene of 35 mg/L with *hSQS* and 25 mg/L with *tSQS* and in strategy 3 was obtained an accumulation of squalene of 62 mg/L and 30 mg/L for *hSQS* and *tSQS* respectively. These results are higher than that obtained in other studies with engineered *E. coli* strain, such as the one carried out by Choi et al., (2019) in which the *DXS*, *IDI*, *ispA* genes were overexpressed, resulting in an accumulation of squalene of 7.16 mg/L. Likewise, in the study carried out by Ghimire et al., (2009), a squalene concentration of 11.8 mg/L was obtained with the expression of the *hopA*, *hopB* and *hopD* genes and Furubayashi et al., 2014 reports a squalene concentration of 2.7 mg/L with the expression of *hSQS* (see Table 6).

In contrast, other studies with different approaches to the design of the microorganism obtained higher values of squalene production, reaching values up to 230 mg/L (Katabami et al., 2015) and 612 mg/L (Meng et al., 2020). In the study of Katabami et al., 2015, was used a chimeric metabolic pathway between MEP and MVA (previously designed by Peralta-Yahya et al. 2012), which is similar to approach 3 in terms of genetic

load. Also, it should be noted that the same measurements reported by Katabami et al (2015) indicate a high standard deviation in the concentration of squalene and that values close to 100 mg/L were also observed, which compares better with the results of the present study (90 mg/L) in approach 1.

It must be highlighted that not only is relevant the number of heterologous genes inserted in the engineered *E. coli* strain for the biosynthesis of compounds of interest such as squalene but other different factors must be considered. For example, the accumulation of intermediate products that can be toxic to the host cell, which lacks intrinsic mechanisms for the regulation of metabolites from non-native metabolic pathways (Ling et al., 2014; Chubukov et al., 2016). Without this regulation an imbalance is generated in the heterologous pathways, which leads to an accumulation and toxicity of metabolites (Jones et al., 2015) which can result in a decrease in the growth rate and in the production of the desired compound (Borkowsky et al., 2016; Dahl et al., 2013; George et al., 2018).

For instance, in the works carried out by Li et al., 2017 and George et al., 2018, which have similar strategies to the approaches 2 and 3 respectively, the toxicity of intermediate products has been reported. Li et al., (2018) evidenced the cytotoxicity in *E. coli* of the intermediate hydroxymethylbutenyl diphosphate (HMBPP) product by overexpressing the *IspG* gene of the MEP metabolic pathway, which negatively influences cell growth and the isoprene production such as  $\beta$ -carotene and lycopene. On the other hand, George et al., (2018) evaluated the toxicity of the accumulation of isopentenyl pyrophosphate (IPP) associated with the stress caused by assembling the non-native MVA pathway in *E. coli*. The results of their work showed that the accumulation of IPP is related to the decrease of cell viability, reduced nutrient intake and ATP levels and also affects nucleotide metabolism. This may explain why in approaches 2 and 3 lower concentrations of squalene were observed than those obtained in approach 1.

Another factor to consider is the formation of inclusion bodies in *E. coli*, which are aggregates of hydrophobic nature, generally composed of proteins that fold abnormally or that undergo inappropriate post-translational modifications, as a consequence of heterologous gene expression in the host cell (Bhatwa et al., 2021; Tsumoto et al., 2003). The formation of inclusion bodies is directly related to the cellular stress generated by the extra metabolic load in the bacteria due to the increased energy demand (Gill et al., 2000; Zeng and Yang, 2019). To face this challenge, various strategies have been developed to reduce the formation of inclusion bodies that may affect the production of the desired compound, some of these are: modifications of culture conditions that include growth temperature, inducer concentration and additives to the culture (Bhatwa et al., 2021). Also, a good strategy to avoid the formation of inclusion bodies could be the use of bioinformatics tools in order to predict the solubility or aggregation of proteins (Hebditch et al., 2017; Bhandari et al., 2020) and consequently, replace those

wild type proteins that are prone to form inclusion bodies by their mutants with truncated domains (Jung and Park, 2008; Bhatwa et al., 2021).

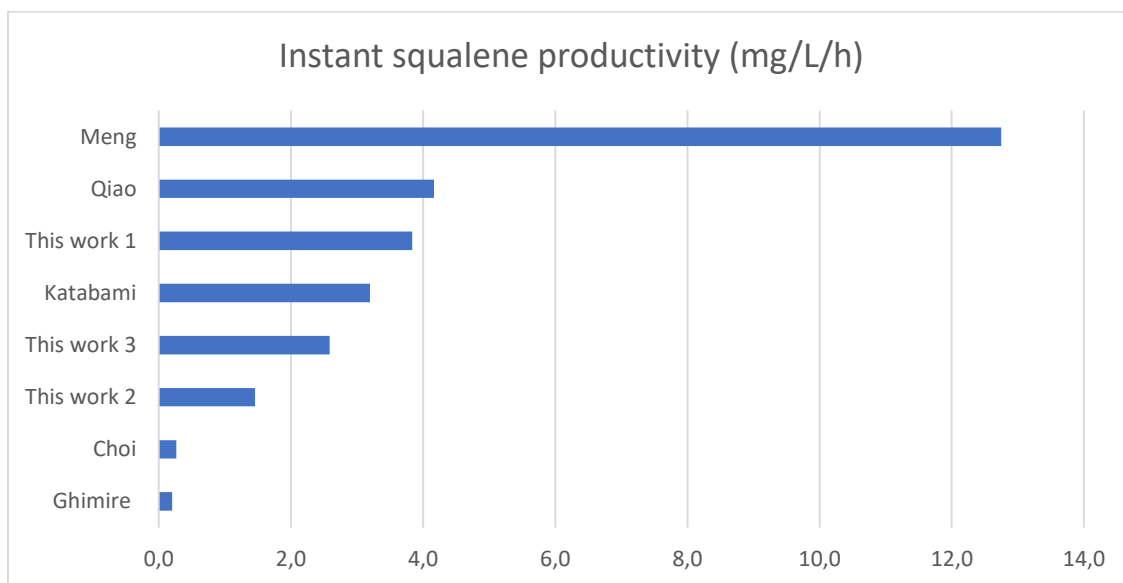
Many strategies are focused on overexpressing native and heterologous genes that lead to an increased precursor supply for the desired compound production and also adjustments are made to the culture conditions to generate the highest possible productivity. However, the bacteria *E. coli* is not physiologically "adapted" to store large amounts of lipophilic compounds inside since it does not have a specialized organelle for this purpose (Meng et al., 2020). Therefore, the production of compounds such as squalene is limited and increasing its production currently represents a challenge for metabolic engineering. With the strategies focused on the insertion of heterologous genes mentioned above, the highest squalene production reported in the literature for *E. coli* was 230 mg/L (Katabami et al., 2015).

However, this production remains uncompetitive against the global demand for squalene (note that higher contents are reported in Yeast, see table 6). In a recent study Meng et al., (2020) addressed this bottleneck, with the modification in the inner membrane of *E. coli* in the form of invaginations, simulating the mitochondrial cristae to increase the lipid storage area in the bacterial cell. Overexpressing some cell membrane proteins that induce invaginations, the authors reported a squalene concentration of 612 mg/L (Meng et al., 2020). In a similar strategy, metabolically engineering the Yeast peroxisome for squalene storage, a production of 11 g/L was achieved (Liu et al., 2020). This same research team (Zhu et al., 2021) reached the highest accumulation reported so far for a genetically engineered microbe (21.1 g/L), with the dual regulation and compartmentalization of the cytoplasmic and mitochondrial MVA pathway in yeast.

An interesting observation emerged when comparing our results with those reported in the literature regarding the time in hours of the flask shake culture required for squalene production in genetically engineered *E. coli* strains. Specifically, the *E. coli* strain designed by Ghimire et al., (2009) accumulated 11.8 mg/L of squalene at 48 hours, in the work of Katabami et al., (2015), it was required 72 hours at 37°C to reach a squalene titer of 230 mg/L, the strain designed by Choi et al., (2019) requires 120 hours of cultivation at 30°C to reach a squalene titer of 32 mg/L, the *E. coli* strain designed by Quiao et al., (2019), required 48 hours of culture to reach a squalene accumulation of 200 mg/L, while Meng et al., (2020) engineered the strain that has the highest squalene titer reported so far with 612 mg/L of squalene in 48 hours of cultivation at 30°C.

**Figure 24**, compares the instant productivities in mg/L/h of the different engineered *E. coli* strains belonging to the authors above mentioned. In this graph, it can be seen that the strain of *E. coli* designed with approach 1 "This work1", is found in the third position with an instant productivity of 3.8, very close to the second position of Quiao et al., (2019) with 4.2 and above the productivity achieved by Katabami et al., (2015) with 3.2 mg/L/h. Meng et al., (2020) had reported the highest instant productivity (12.75

mg/L/h). In consequence, it can be inferred that the approach 1 strain has broad possibilities of being used for the squalene production at industrial scale. On the other hand, the engineered *E. coli* strains made by approach 2 "this work 2" and approach 3 "this work 3" have instant productivities of 1.5 and 2.6, respectively, which are higher than the strains designed by Choi et al., (2019) (0.3 mg/L/h) and Ghimire et al., (2009) (0.2 mg/L/h). We speculate that with strategies focused to alleviate the "burden" or the metabolic load of the strains, in addition to adjusting their cultivation conditions and taking into account other considerations such as those previously mentioned, it could be achieved competitive productivity to cover the global demand of squalene market at an industrial level thanks to the use of synthetic biology and metabolic engineering tools.



**Figure 24.** Instant squalene productivity in mg/L/h of the different engineered *E. coli* strains obtained in this work compared to those reported in the literature.

Finally, it should be noted that for the first time in our laboratory this kind of research is being conducted in metabolic engineering and synthetic biology for the production of a compound with commercial interest in *E. coli*. In this way, with the exploration of the three different strategies, it was interesting to observe how each of them contributes to better understand the microbial cell factory for the production of squalene. The lessons learned will serve to use the engineered strains as a platform for the production of isoprenoids.

## 2.6 Conclusions

Our work for the first time reports metabolic engineering of genes from rice plants and *Gentiana lutea* (strategy 1) for the accumulation of squalene in a genetically engineered *E. coli* strain. With strategy 1, the highest instant productivity of squalene (3.8 mg/L/h) was reached, which is in the third position of the highest productivity reported so far in the literature. This engineered strain could have broad potential to be used for industrial-scale squalene production. The instantaneous productivities of strategies two and three (1.5 and 2.6 mg/L/h, respectively) could be improved by adjusting the culture conditions and alleviating the metabolic burden due to the number of genes inserted in plasmids.

## 2.7 Recommendations and prospects

In our particular case, to increase the production of squalene, optimization of the culture conditions in the Erlenmeyer flask is required. One of them is to sufficiently supply the oxygen needs required to cover the demand for reducing power and ATP produced by the MEP pathway. In addition to this, the stress to which the genetically engineered strain is subjected by simultaneous resistance to three different types of antibiotics in strategy three (chloramphenicol, ampicillin and spectinomycin) must be taken into account, which can lead to decreased cell growth (Fordjour et al., 2022)

Genome mining is a bioinformatics tool recently used to explore systematically genomes and protein sequence in robust databases and helps identify proteins of interest (Ouedraogo & Tsang 2021). This tool can contribute to find more efficient enzymes for the production of squalene. In this way, it is possible to identify the genes that encode the squalene synthase enzyme from the shark or a phylogenetically close organism and use them for designing new strains for the production of squalene.

The use of computational tools to identify and reject (if it is the case), unfavorable constructs that can generate a high metabolic load for the cell when implemented in a designed strain of *E. coli* (Borkowski et al., 2016). For example, the Cello software developed by Nielsen et al., (2016) is capable of designing constructs that incorporate considerations of metabolic load and other factors that influence cell viability. This tool works 90% of the time when implemented in vivo.

The use of genome editing tools such as CRISPR/Cas9, suitable to insert genes of interest into the chromosome at specific locations in the genome is required to achieve stable transformations over time, avoiding the segregational and spatial instability inherent in the plasmid transformation system (Wei, Y. et al., 2018; Englaender et al., 2017; Tyo et al., 2009). In our case, key genes from the squalene biosynthetic pathway could be inserted close to essential genes for cell maintenance in the bacteria to ensure that

these are transcribed and translated into proteins. This strategy could improve squalene production in *E. coli* and their stability in time.

## 2.8 References

Allison, A. C., & Byars, N. E. (1986). An adjuvant formulation that selectively elicits the formation of antibodies of protective isotypes and of cell-mediated immunity. *Journal of immunological methods*, *95*, 157-168.

Ajikumar, P. K., Xiao, W. H., Tyo, K. E., Wang, Y., Simeon, F., Leonard, E., & Stephanopoulos, G. (2010). Isoprenoid pathway optimization for Taxol precursor overproduction in *Escherichia coli*. *Science*, *330*, 70-74.

Alonso-Gutierrez, J., Chan, R., Batth, T. S., Adams, P. D., Keasling, J. D., Petzold, C. J., & Lee, T. S. (2013). Metabolic engineering of *Escherichia coli* for limonene and perillyl alcohol production. *Metabolic engineering*, *19*, 33-41.

Alonso-Gutierrez, J., Kim, E. M., Batth, T. S., Cho, N., Hu, Q., Chan, L. J. G., & Lee, T. S. (2015). Principal component analysis of proteomics (PCAP) as a tool to direct metabolic engineering. *Metabolic engineering*, *28*, 123-133.

Alonso-Gutierrez, J., Koma, D., Hu, Q., Yang, Y., Chan, L. J., Petzold, C. J., & Lee, T. S. (2018). Toward industrial production of isoprenoids in *Escherichia coli*: Lessons learned from CRISPR-Cas9 based optimization of a chromosomally integrated mevalonate pathway. *Biotechnology and bioengineering*, *115*, 1000-1013. Bhandari et al., 2020

ApE: A plasmid Editor (v3.1.0). (2022). [Plasmid Editor software]. M. Wayne Davis. <https://jorgensen.biology.utah.edu/wayned/ape/>

Bai, C., Capell, T., Berman, J., Medina, V., Sandmann, G., Christou, P., & Zhu, C. (2016). Bottlenecks in carotenoid biosynthesis and accumulation in rice endosperm are influenced by the precursor–product balance. *Plant Biotechnology Journal*, *14*, 195-205.

Bai, C., Rivera, S. M., Medina, V., Alves, R., Vilaprinyo, E., Sorribas, A., & Zhu, C. (2014). An in vitro system for the rapid functional characterization of genes involved in carotenoid biosynthesis and accumulation. *The Plant Journal*, *77*, 464-475.

Bhatwa, A., Wang, W., Hassan, Y. I., Abraham, N., Li, X. Z., & Zhou, T. (2021). Challenges associated with the formation of recombinant protein inclusion bodies in *Escherichia coli* and strategies to address them for industrial applications. *Frontiers in Bioengineering and Biotechnology*, *9*, 65.

Bhandari, B. K., Gardner, P. P., & Lim, C. S. (2020). Solubility-Weighted Index: fast and accurate prediction of protein solubility. *Bioinformatics*, *36*, 4691-4698.

- Bhilwade, N., H., Tatewaki, N., Nishida, H., & Konishi, T. (2010). Squalene as novel food factor. *Current Pharmaceutical Biotechnology*, *11*, 875-880.
- Bligh, E. G., & Dyer, W. J. (1959). A rapid method of total lipid extraction and purification. *Canadian journal of biochemistry and physiology*, *37*, 911-917.
- Borkowski, O., Bricio, C., Murgiano, M., Rothschild-Mancinelli, B., Stan, G. B., & Ellis, T. (2018). Cell-free prediction of protein expression costs for growing cells. *Nature communications*, *9*, 1-11.
- Borkowski, O., Ceroni, F., Stan, G. B., & Ellis, T. (2016). Overloaded and stressed: whole-cell considerations for bacterial synthetic biology. *Current opinion in microbiology*, *33*, 123-130.
- Brown, W., Ralston, A., & Shaw, K. (2008). Positive transcription control: The glucose effect. *Nature Education*, *1*, 202.
- Budiyanto, A., Ahmed, N. U., Wu, A., Bito, T., Nikaido, O., Osawa, T., & Ichihashi, M. (2000). Protective effect of topically applied olive oil against photocarcinogenesis following UVB exposure of mice. *Carcinogenesis*, *21*, 2085-2090.
- Burke, C., & Croteau, R. (2002). Geranyl diphosphate synthase from *Abies grandis*: cDNA isolation, functional expression, and characterization. *Archives of biochemistry and biophysics*, *405*, 130-136.
- Bunch, A. W., & Harris, R. E. (1986). The manipulation of micro-organisms for the production of secondary metabolites. *Biotechnology and genetic engineering reviews*, *4*, 117-144.
- Cervantes-Cervantes, M., Gallagher, C. E., Zhu, C., & Wurtzel, E. T. (2006). Maize cDNAs expressed in endosperm encode functional farnesyl diphosphate synthase with geranylgeranyl diphosphate synthase activity. *Plant physiology*, *141*, 220-231.
- Ciriminna, R., Pandarus, V., Béland, F., & Pagliaro, M. (2014). Catalytic hydrogenation of squalene to squalane. *Organic Process Research & Development*, *18*, 1110-1115.
- Chang, W. C., Song, H., Liu, H. W., & Liu, P. (2013). Current development in isoprenoid precursor biosynthesis and regulation. *Current opinion in chemical biology*, *17*, 571-579.
- Chen, G., Fan, K. W., Lu, F. P., Li, Q., Aki, T., Chen, F., & Jiang, Y. (2010). Optimization of nitrogen source for enhanced production of squalene from thraustochytrid *Aurantiochytrium* sp. *New biotechnology*, *27*, 382-389.
- Choi, B. H., Kim, J. H., Choi, S. Y., Han, S. J., & Lee, P. C. (2019). Redesign and reconstruction of a mevalonate pathway and its application in terpene production in *Escherichia coli*. *Bioresource Technology Reports*, *7*, 100291.



- Choi, S. Y., Lee, H. J., Choi, J., Kim, J., Sim, S. J., Um, Y., & Woo, H. M. (2016). Photosynthetic conversion of CO<sub>2</sub> to farnesyl diphosphate-derived phytochemicals (amorpha-4, 11-diene and squalene) by engineered cyanobacteria. *Biotechnology for biofuels*, *9*, 1-12.
- Chubukov, V., Mukhopadhyay, A., Petzold, C. J., Keasling, J. D., & Martín, H. G. (2016). Synthetic and systems biology for microbial production of commodity chemicals. *NPJ systems biology and applications*, *2*, 1-11.
- Crater, J. S., & Lievens, J. C. (2018). Scale-up of industrial microbial processes. *FEMS microbiology letters*, *365*(13), fny138.
- Csörgő, B., Fehér, T., Tímár, E., Blattner, F. R., & Pósfai, G. (2012). Low-mutation-rate, reduced-genome *Escherichia coli*: an improved host for faithful maintenance of engineered genetic constructs. *Microbial cell factories*, *11*(1), 1-13.
- Dahl, R. H., Zhang, F., Alonso-Gutierrez, J., Baidoo, E., Batth, T. S., Redding-Johanson, A. M., ... & Keasling, J. D. (2013). Engineering dynamic pathway regulation using stress-response promoters. *Nature biotechnology*, *31*(11), 1039-1046.
- Das, B., Yeger, H., Baruchel, H., Freedman, M. H., Koren, G., & Baruchel, S. (2003). In vitro cytoprotective activity of squalene on a bone marrow versus neuroblastoma model of cisplatin-induced toxicity: implications in cancer chemotherapy. *European Journal of Cancer*, *39*(17), 2556-2565.
- Donovan, R. S., Robinson, C. W., & Glick, B. R. (1996). Optimizing inducer and culture conditions for expression of foreign proteins under the control of the lac promoter. *Journal of industrial microbiology*, *16*(3), 145-154.
- Dumont, M. J., & Narine, S. S. (2007). Characterization of flax and soybean soapstocks, and soybean deodorizer distillate by GC-FID. *Journal of the American Oil Chemists' Society*, *84*(12), 1101-1105.
- Dvorak, P., Chrast, L., Nikel, P. I., Fedr, R., Soucek, K., Sedlackova, M., & Damborsky, J. (2015). Exacerbation of substrate toxicity by IPTG in *Escherichia coli* BL21 (DE3) carrying a synthetic metabolic pathway. *Microbial cell factories*, *14*(1), 1-15.
- Emanuelsson, O., Nielsen, H., Brunak, S., & Von Heijne, G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of molecular biology*, *300*(4), 1005-1016.
- Englaender, J. A., Jones, J. A., Cress, B. F., Kuhlman, T. E., Linhardt, R. J., & Koffas, M. A. (2017). Effect of genomic integration location on heterologous protein expression and metabolic engineering in *E. coli*. *ACS Synthetic Biology*, *6*(4), 710-720.

- Fan, K. W., Aki, T., Chen, F., & Jiang, Y. (2010). Enhanced production of squalene in the thraustochytrid *Aurantiochytrium mangrovei* by medium optimization and treatment with terbinafine. *World Journal of Microbiology and Biotechnology*, *26*(7), 1303-1309.
- Fernández-Castané, A., Caminal, G., & López-Santín, J. (2012). Direct measurements of IPTG enable analysis of the induction behavior of *E. coli* in high cell density cultures. *Microbial cell factories*, *11*(1), 1-9.
- Francis, D. M., & Page, R. (2010). Strategies to optimize protein expression in *E. coli*. *Current protocols in protein science*, *61*(1), 5-24.
- Fordjour, E., Mensah, E. O., Hao, Y., Yang, Y., Liu, X., Li, Y., & Bai, Z. (2022). Toward improved terpenoids biosynthesis: strategies to enhance the capabilities of cell factories. *Bioresources and Bioprocessing*, *9*(1), 1-33.
- Furubayashi, M., Li, L., Katabami, A., Saito, K., & Umeno, D. (2014). Construction of carotenoid biosynthetic pathways using squalene synthase. *FEBS letters*, *588*(3), 436-442.
- Gama, J. A., Zilhão, R., & Dionisio, F. (2020). Plasmid interactions can improve plasmid persistence in bacterial populations. *Frontiers in Microbiology*, 2033.
- George, K. W., Thompson, M. G., Kim, J., Baidoo, E. E., Wang, G., Benites, V. T., & Lee, T. S. (2018). Integrated analysis of isopentenyl pyrophosphate (IPP) toxicity in isoprenoid-producing *Escherichia coli*. *Metabolic engineering*, *47*, 60-72.
- Ghimire, G. P., Lee, H. C., & Sohng, J. K. (2009). Improved squalene production via modulation of the methylerythritol 4-phosphate pathway and heterologous expression of genes from *Streptomyces peucetius* ATCC 27952 in *Escherichia coli*. *Applied and environmental microbiology*, *75*, 7291-7293.
- Ghimire, G. P., Nguyen, H. T., Koirala, N., & Sohng, J. K. (2016). Advances in biochemistry and microbial production of squalene and its derivatives. *Journal of Microbiology and Biotechnology*, *26*, 441-451.
- Giacometti, J., & Milin, C. (2001). Composition and qualitative characteristics of virgin olive oils produced in northern Adriatic region, Republic of Croatia. *Grasas y aceites*, *52*, 397-402.
- Gibson, D. G., Glass, J. I., Lartigue, C., Noskov, V. N., Chuang, R. Y., Algire, M. A., & Venter, J. C. (2010). Creation of a bacterial cell controlled by a chemically synthesized genome. *Science*, *329*, 52-56.
- Gill, R. T., Valdes, J. J., & Bentley, W. E. (2000). A comparative study of global stress gene regulation in response to overexpression of recombinant proteins in *Escherichia coli*. *Metabolic engineering*, *2*, 178-189.

- Gohil, N., Bhattacharjee, G., Khambhati, K., Braddick, D., & Singh, V. (2019). Engineering strategies in microorganisms for the enhanced production of squalene: advances, challenges and opportunities. *Frontiers in bioengineering and biotechnology*, 7, 50.
- Gohil, N., Bhattacharjee, G., & Singh, V. (2021). An introduction to microbial cell factories for production of biomolecules. In *Microbial Cell Factories Engineering for Production of Biomolecules* (pp. 1-19). Academic Press.
- Gopakumar, K. (2012) Therapeutic applications of squalene - a review, *Fishery Technology*, 49, 1-9.
- Gorochofski, T. E., Van Den Berg, E., Kerkman, R., Roubos, J. A., & Bovenberg, R. A. (2014). Using synthetic biological parts and microbioreactors to explore the protein expression characteristics of Escherichia coli. *ACS synthetic biology*, 3, 129-139.
- Grob, A., Di Blasi, R., & Ceroni, F. (2021). Experimental tools to reduce the burden of bacterial synthetic biology. *Current Opinion in Systems Biology*, 28, 100393.
- Grossman, T. H., Kawasaki, E. S., Punreddy, S. R., & Osburne, M. S. (1998). Spontaneous cAMP-dependent derepression of gene expression in stationary phase plays a role in recombinant expression instability. *Gene*, 209, 95-103.
- Güneş, F. E. (2013). Medical Use of Squalene as a Natural Antioxidant. *Journal of Marmara University Institute of Health Sciences*, 3(4).
- Han, J. Y., Seo, S. H., Song, J. M., Lee, H., & Choi, E. S. (2018). High-level recombinant production of squalene using selected *Saccharomyces cerevisiae* strains. *Journal of Industrial Microbiology and Biotechnology*, 45, 239-251.
- Hebditch, M., Carballo-Amador, M. A., Charonis, S., Curtis, R., & Warwicker, J. (2017). Protein-Sol: a web tool for predicting protein solubility from sequence. *Bioinformatics*, 33, 3098-3100.
- Hernandez-Arranz, S., Perez-Gil, J., Marshall-Sabey, D., & Rodriguez-Concepcion, M. (2019). Engineering *Pseudomonas putida* for isoprenoid production by manipulating endogenous and shunt pathways supplying precursors. *Microbial cell factories*, 18, 1-14.
- Hill, P., Benjamin, K., Bhattacharjee, B., Garcia, F., Leng, J., Liu, C. L., & Kraft, C. (2020). Clean manufacturing powered by biology: how Amyris has deployed technology and aims to do it better. *Journal of Industrial Microbiology & Biotechnology: Official Journal of the Society for Industrial Microbiology and Biotechnology*, 47, 965-975.
- Huang, Z. R., Lin, Y. K., & Fang, J. Y. (2009). Biological and pharmacological activities of squalene and related compounds: potential uses in cosmetic dermatology. *Molecules*, 14, 540-554.

- Huang, Y. Y., Jian, X. X., Lv, Y. B., Nian, K. Q., Gao, Q., Chen, J., & Hua, Q. (2018). Enhanced squalene biosynthesis in *Yarrowia lipolytica* based on metabolically engineered acetyl-CoA metabolism. *Journal of biotechnology*, *281*, 106-114.
- Hussain, M. H., Mohsin, M. Z., Zaman, W. Q., Yu, J., Zhao, X., Wei, Y., & Guo, M. (2022). Multiscale engineering of microbial cell factories: A step forward towards sustainable natural products industry. *Synthetic and systems biotechnology*, *7*, 586-601.
- Jiang, W., Zhao, X., Gabrieli, T., Lou, C., Ebenstein, Y., & Zhu, T. F. (2015). Cas9-Assisted Targeting of CHromosome segments CATCH enables one-step targeted cloning of large gene clusters. *Nature communications*, *6*, 1-8.
- Jin, X., Baysal, C., Gao, L., Medina, V., Drapal, M., Ni, X., Sheng, Y., Shi, L., Capell, T., Fraser, P.D. Christou, P., & Zhu, C. (2020). The subcellular localization of two isopentenyl diphosphate isomerases in rice suggests a role for the endoplasmic reticulum in isoprenoid biosynthesis. *Plant Cell Reports*, *39*, 119-133.
- Jin, X., Baysal, C., Drapal, M., Sheng, Y., Huang, X., He, W., Shi, L., Capell, T., Fraser, P.D., Christou, P. & Zhu, C. (2021). The Coordinated upregulated expression of genes involved in MEP, chlorophyll, carotenoid and tocopherol pathways, mirrored the corresponding metabolite contents in rice leaves during de-etiolation. *Plants*, *10*, 1456.
- Jones, J. A., Toparlak, Ö. D., & Koffas, M. A. (2015). Metabolic pathway balancing and its role in the production of biofuels and chemicals. *Current opinion in biotechnology*, *33*, 52-59.
- Jung, S., & Park, S. (2008). Improving the expression yield of *Candida antarctica* lipase B in *Escherichia coli* by mutagenesis. *Biotechnology letters*, *30*, 717-722.
- Katabami, A., Li, L., Iwasaki, M., Furubayashi, M., Saito, K., & Umeno, D. (2015). Production of squalene by squalene synthases and their truncated mutants in *Escherichia coli*. *Journal of bioscience and bioengineering*, *119*, 165-171.
- Keasling, J. D. (2010). Manufacturing molecules through metabolic engineering. *Science*, *330*, 1355-1358.
- Kim, S. W., & Keasling, J. D. (2001). Metabolic engineering of the nonmevalonate isopentenyl diphosphate synthesis pathway in *Escherichia coli* enhances lycopene production. *Biotechnology and bioengineering*, *72*, 408-415.
- Kim, Y. J., Kim, T. W., Chung, H., Kwon, I. C., Sung, H. C., & Jeong, S. Y. (2003). The effects of serum on the stability and the transfection activity of the cationic lipid emulsion with various oils. *International journal of pharmaceuticals*, *252*, 241-252.
- Koma, D., Yamanaka, H., Moriyoshi, K., Ohmoto, T., & Sakai, K. (2012). A convenient method for multiple insertions of desired genes into target loci on the *Escherichia coli* chromosome. *Applied microbiology and biotechnology*, *93*, 815-829.

- Kumar, P. K. R., Maschke, H. E., Friehs, K., & Schügerl, K. (1991). Strategies for improving plasmid stability in genetically modified bacteria in bioreactors. *Trends in biotechnology*, 9, 279-284.
- Kuzuyama, T. (2002). Mevalonate and non-mevalonate pathways for the biosynthesis of isoprene units. *Bioscience, biotechnology, and biochemistry*, 66, 1619-1627.
- Kwak, S., Kim, S. R., Xu, H., Zhang, G. C., Lane, S., Kim, H., & Jin, Y. S. (2017). Enhanced isoprenoid production from xylose by engineered *Saccharomyces cerevisiae*. *Biotechnology and bioengineering*, 114, 2581-2591.
- Lewis, M. (2013). Allostery and the lac Operon. *Journal of molecular biology*, 425, 2309-2316.
- Lee, S. K., Newman, J. D., & Keasling, J. D. (2005). Catabolite repression of the propionate catabolic genes in *Escherichia coli* and *Salmonella enterica*: evidence for involvement of the cyclic AMP receptor protein. *Journal of bacteriology*, 187, 2793-2800.
- Lee, S. K., & Keasling, J. D. (2005). A propionate-inducible expression system for enteric bacteria. *Applied and environmental microbiology*, 71, 6856-6862.
- Lee, S., & Poulter, C. D. (2008). Cloning, solubilization, and characterization of squalene synthase from *Thermosynechococcus elongatus* BP-1. *Journal of Bacteriology*, 190, 3808-3816.
- Li, Q., Fan, F., Gao, X., Yang, C., Bi, C., Tang, J., Liu, T. & Zhang, X. (2017). Balanced activation of IspG and IspH to eliminate MEP intermediate accumulation and improve isoprenoids production in *Escherichia coli*. *Metabolic engineering*, 44, 13-21.
- Li, M., Nian, R., Xian, M., & Zhang, H. (2018). Metabolic engineering for the production of isoprene and isopentenol by *Escherichia coli*. *Applied microbiology and biotechnology*, 102(18), 7725-7738.
- Ling, H., Teo, W., Chen, B., Leong, S. S. J., & Chang, M. W. (2014). Microbial tolerance engineering toward biochemical production: from lignocellulose to products. *Current opinion in biotechnology*, 29, 99-106.
- Liu, H., Han, S., Xie, L., Pan, J., Zhang, W., Gong, G., & Hu, Y. (2017). Overexpression of key enzymes of the 2-C-methyl-D-erythritol-4-phosphate (MEP) pathway for improving squalene production in *Escherichia coli*. *African Journal of Biotechnology*, 16, 2307-2316.
- Liu, G. S., Li, T., Zhou, W., Jiang, M., Tao, X. Y., Liu, M., ... & Wei, D. Z. (2020). The yeast peroxisome: a dynamic storage depot and subcellular factory for squalene overproduction. *Metabolic engineering*, 57, 151-161.

- Lou-Bonafonte, J.M., Martínez-Beamonte, R., Sanclemente, T., Surra, J.C., Herrera-Marcos, L.V., Sanchez-Marco, J., Arnal, C. & Osada, J. (2018). Current insights into the biological action of squalene. *Molecular nutrition & food research*, *62*, 1800136.
- Ma, C., Zhang, K., Zhang, X., Liu, G., Zhu, T., Che, Q., & Zhang, G. (2021). Heterologous expression and metabolic engineering tools for improving terpenoids production. *Current Opinion in Biotechnology*, *69*, 281-289.
- Martin, V. J., Pitera, D. J., Withers, S. T., Newman, J. D., & Keasling, J. D. (2003). Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nature biotechnology*, *21*, 796-802.
- Meng, X., Yang, J., Cao, Y., Li, L., Jiang, X., Xu, X., & Zhang, Y. (2011). Increasing fatty acid production in *E. coli* by simulating the lipid accumulation of oleaginous microorganisms. *Journal of Industrial Microbiology and Biotechnology*, *38*, 919-925.
- Meng, Y., Shao, X., Wang, Y., Li, Y., Zheng, X., Wei, G., & Wang, C. (2020). Extension of cell membrane boosting squalene production in the engineered *Escherichia coli*. *Biotechnology and Bioengineering*, *117*, 3499-3507.
- Morrone, D., Lowry, L., Determan, M. K., Hershey, D. M., Xu, M., & Peters, R. J. (2010). Increasing diterpene yield with a modular metabolic engineering system in *E. coli*: comparison of MEV and MEP isoprenoid precursor pathway engineering. *Applied microbiology and biotechnology*, *85*, 1893-1906.
- Nakazawa, A., Matsuura, H., Kose, R., Kato, S., Honda, D., Inouye, I., & Watanabe, M. M. (2012). Optimization of culture conditions of the thraustochytrid *Aurantiochytrium sp.* strain 18W-13a for squalene production. *Bioresource Technology*, *109*, 287-291.
- Naz, S., Sherazi, S. T. H., Talpur, F. N., Kara, H., Uddin, S., & Khaskheli, A. R. (2014). Chemical Characterization of Canola and Sunflower oil deodorizer distillates. *Polish Journal of Food and Nutrition Sciences*, *64*(2).
- Naziri, E., Mantzouridou, F., & Tsimidou, M. Z. (2011). Squalene resources and uses point to the potential of biotechnology. *Lipid Technology*, *23*, 270-273.
- Nielsen, A. A., Der, B. S., Shin, J., Vaidyanathan, P., Paralanov, V., Strychalski, E. A., & Voigt, C. A. (2016). Genetic circuit design automation. *Science*, *352*, aac7341.
- Nielsen, J., & Keasling, J. D. (2016). Engineering cellular metabolism. *Cell*, *164*(6), 1185-1197.
- Olaofe, O. A., Burton, S. G., Cowan, D. A., & Harrison, S. T. (2010). Improving the production of a thermostable amidase through optimising IPTG induction in a highly dense culture of recombinant *Escherichia coli*. *Biochemical engineering journal*, *52*, 19-24.

Ouedraogo, J. P., & Tsang, A. (2021). Production of native and recombinant enzymes by fungi for industrial applications. En O. Zaragoza & A. Casadevall (Eds.), *Encyclopedia of Mycology* (pp. 222-232). Elsevier.

Pan, J.J., Solbiati, J.O., Ramamoorthy, G., Hillerich, B.S., Seidel, R.D., Cronan, J.E., Almo, S.C. & Poulter, C.D. (2015). Biosynthesis of squalene from farnesyl diphosphate in bacteria: three steps catalyzed by three enzymes. *ACS central science*, *1*, 77-82.

Peng, S., Cao, F., Xia, Y., Gao, X. D., Dai, L., Yan, J., & Ma, G. (2020). Particulate Alum via Pickering Emulsion for an Enhanced COVID-19 Vaccine Adjuvant. *Advanced Materials*, *32*, 2004210.

Peralta-Yahya, P. P., Zhang, F., Del Cardayre, S. B., & Keasling, J. D. (2012). Microbial engineering for the production of advanced biofuels. *Nature*, *488*, 320-328.

Pérez, L., Alves, R., Perez-Fons, L., Albacete, A., Farré, G., Soto, E., Vilaprinçó, E., Martínez-Andújar, C., Basallo, O., Medina, V., Zhu, C., Capell, T., & Christou, P. (2022). Multilevel interactions between native and ectopic isoprenoid pathways affect global metabolism in rice. *Transgenic Research*, 1-20.

Pierce, Benjamin. (2014) *Genetics: A Conceptual Approach*, 5th ed. (New York: W. H. Freeman and Company), 446.

Popa, O., Băbeanu, N. E., Popa, I., Niță, S., & Dinu-Pârvu, C. E. (2015). Methods for obtaining and determination of squalene from natural sources. *BioMed research international*, 2015.

Prasad E. and Roy A. 2016. *Squalene market global opportunity analysis and industry forecast, 2014-2022*. (REPORT no. MA 161657) Allied Market Research Report. <https://www.researchandmarkets.com/reports/4033349/squalene-market-by-source-shark-liver>

Pretorius, I. S. (2017). Synthetic genome engineering forging new frontiers for wine yeast. *Critical reviews in biotechnology*, *37*, 112-136.

Qiao, W., Zhou, Z., Liang, Q., Mosongo, I., Li, C., & Zhang, Y. (2019). Improving lupeol production in yeast by recruiting pathway genes from different organisms. *Scientific reports*, *9*, 1-8.

Rabinovitch-Deere, C. A., Oliver, J. W., Rodriguez, G. M., & Atsumi, S. (2013). Synthetic biology and metabolic engineering approaches to produce biofuels. *Chemical reviews*, *113*, 4611-4632.

Rani, A., Meghana, R., & Kush, A. (2018). Squalene production in the cell suspension cultures of Indian sandalwood (*Santalum album* L.) in shake flasks and air lift bioreactor. *Plant Cell, Tissue and Organ Culture (PCTOC)*, *135*, 155-167.

- Rasool, A., Ahmed, M. S., & Li, C. (2016). Overproduction of squalene synergistically downregulates ethanol production in *Saccharomyces cerevisiae*. *Chemical Engineering Science*, *152*, 370-380.
- Reddy, L. H., & Couvreur, P. (2009). Squalene: A natural triterpene for use in disease management and therapy. *Advanced drug delivery reviews*, *61*, 1412-1426.
- Ronco, A. L., & De Stéfani, E. (2013). Squalene: a multi-task link in the crossroads of cancer and aging. *Functional Foods in Health and Disease*, *3*, 462-476.
- Rosales-García, T., Jimenez-Martinez, C., & Dávila-Ortiz, G. (2017). Squalene extraction: biological sources and extraction methods. *International Journal of Environment, Agriculture and Biotechnology*, *2*, 238838.
- Rozkov, A., Avignone-Rossa, C. A., Ertl, P. F., Jones, P., O'Kennedy, R. D., Smith, J. J., & Bushell, M. E. (2004). Characterization of the metabolic burden on *Escherichia coli* DH1 cells imposed by the presence of a plasmid containing a gene therapy sequence. *Biotechnology and bioengineering*, *88*, 909-915.
- Shachrai, I., Zaslaver, A., Alon, U., & Dekel, E. (2010). Cost of unneeded proteins in *E. coli* is reduced after several generations in exponential growth. *Molecular cell*, *38*, 758-767.
- Sherazi, S. T. H., & Mahesar, S. A. (2016). Vegetable oil deodorizer distillate: a rich source of the natural bioactive components. *Journal of oleo science*, ess16125.
- Shepelin, D., Hansen, A. S. L., Lennen, R., Luo, H., & Herrgård, M. J. (2018). Selecting the best: evolutionary engineering of chemical production in microbes. *Genes*, *9*, 249.
- Singh, V. (2014). Recent advancements in synthetic biology: current status and challenges. *Gene*, *535*, 1-11.
- Singh, V., Chaudhary, D. K., Mani, I., & Dhar, P. K. (2016). Recent advances and challenges of the use of cyanobacteria towards the production of biofuels. *Renewable and Sustainable Energy Reviews*, *60*, 1-10.
- Singh, V., Braddick, D., & Dhar, P. K. (2017). Exploring the potential of genome editing CRISPR-Cas9 technology. *Gene*, *599*, 1-18.
- Singh, V., Gohil, N., Ramirez Garcia, R., Braddick, D., & Fofié, C. K. (2018). Recent advances in CRISPR-Cas9 genome editing technology for biological and biomedical investigations. *Journal of cellular biochemistry*, *119*, 81-94.
- Sleight, S. C., & Sauro, H. M. (2013). Visualization of evolutionary stability dynamics and competitive fitness of *Escherichia coli* engineered with randomized multigene circuits. *ACS synthetic biology*, *2*, 519-528.
- Slusarczyk, A. L., Lin, A., & Weiss, R. (2012). Foundations for the design and implementation of synthetic genetic circuits. *Nature Reviews Genetics*, *13*, 406-420.



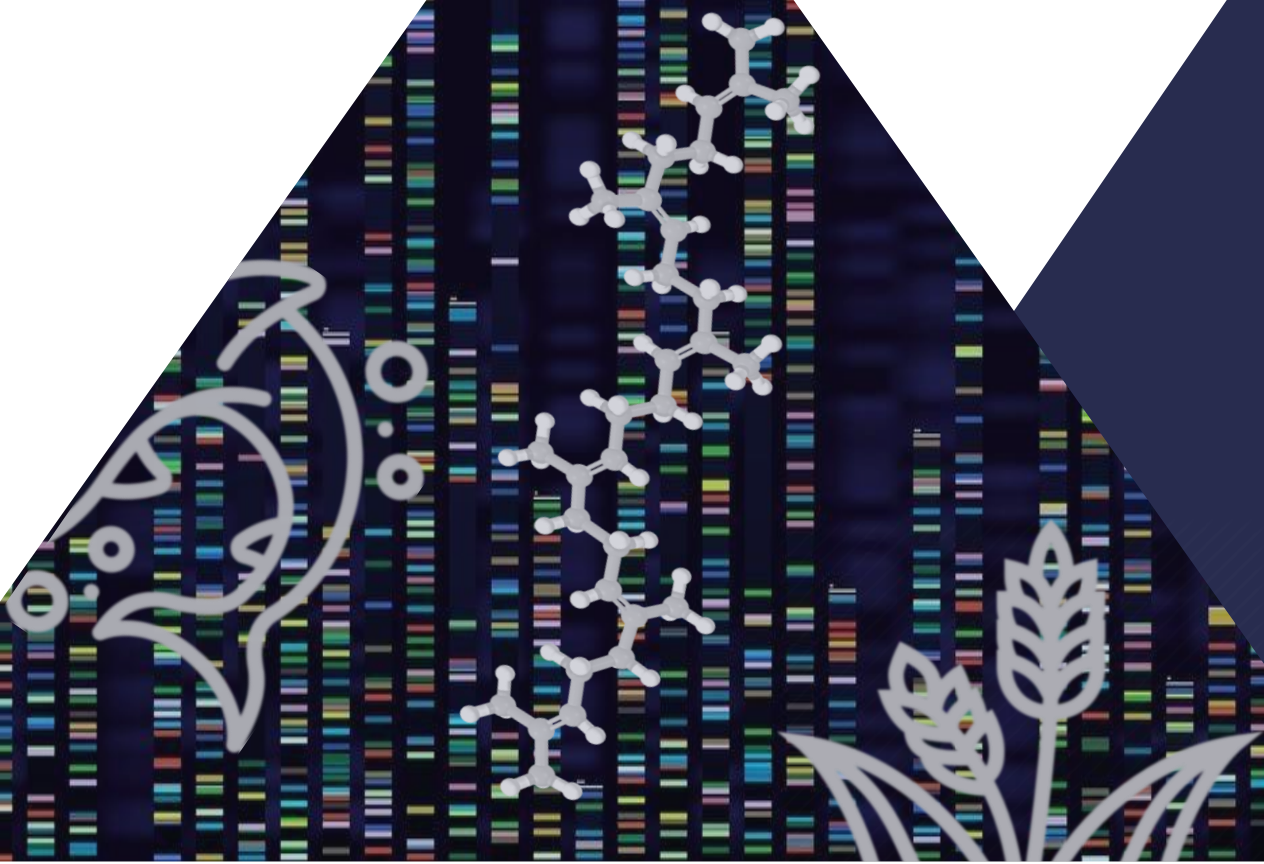
- Spanova, M., & Daum, G. (2011). Squalene–biochemistry, molecular biology, process biotechnology, and applications. *European journal of lipid science and technology*, *113*, 1299-1320.
- Strandberg, T. E., Tilvis, R. S., & Miettinen, T. A. (1990). Metabolic variables of cholesterol during squalene feeding in humans: comparison with cholestyramine treatment. *Journal of lipid research*, *31*, 1637-1643.
- Terol, G. L., Gallego-Jara, J., Martínez, R. A. S., Vivancos, A. M., Díaz, M. C., & de Diego Puente, T. (2021). Impact of the expression system on recombinant protein production in *Escherichia coli* BL21. *Frontiers in Microbiology*, *12*.
- Thompson, J. F., Danley, D. E., Mazzalupo, S., Milos, P. M., Lira, M. E., & Harwood Jr, H. J. (1998). Truncation of human squalene synthase yields active, crystallizable protein. *Archives of biochemistry and Biophysics*, *350*, 283-290.
- Tsoi, K. H., Chan, S. Y., Lee, Y. C., Ip, B. H. Y., & Cheang, C. C. (2016). Shark conservation: an educational approach based on children’s knowledge and perceptions toward sharks. *PLoS one*, *11*, e0163406.
- Tsujimoto, M. (1916). A highly unsaturated hydrocarbon in shark liver oil. *Industrial & Engineering Chemistry*, *8*, 889-896.
- Tsumoto, K., Ejima, D., Kumagai, I., & Arakawa, T. (2003). Practical considerations in refolding proteins from inclusion bodies. *Protein expression and purification*, *28*, 1-8.
- Tyo, K. E., Ajikumar, P. K., & Stephanopoulos, G. (2009). Stabilized gene duplication enables long-term selection-free heterologous pathway expression. *Nature biotechnology*, *27*, 760-765.
- United States Department of Agriculture, 2018 US Department of Agriculture, Agricultural Research Service, Nutrient Data Laboratory. USDA Branded Food Products Database. Version Current: July 2018. Internet: <http://www.ars.usda.gov/nutrientdata> (Accessed July 9, 2020).
- Vaccari, N., & Guerra, D. G. (2022). Influences of induction time and additional *laci* on the growth rate and yield of the t7 expression system in three *Escherichia coli* strains. *World Journal of Microbiology and Biotechnology* DOI:10.21203/rs.3.rs-1220741/v1
- Valachovič, M., & Hapala, I. (2017). Biosynthetic approaches to squalene production: the case of yeast. In *Vaccine Adjuvants* (pp. 95-106). Humana Press, New York, NY.
- Wang, J. W., Wang, A., Li, K., Wang, B., Jin, S., Reiser, M., & Lockey, R. F. (2015). CRISPR/Cas9 nuclease cleavage combined with Gibson assembly for seamless cloning. *BioTechniques*, *58*, 161-170.

- Wang, C., Liwei, M., Park, J. B., Jeong, S. H., Wei, G., Wang, Y., & Kim, S. W. (2018). Microbial platform for terpenoid production: *Escherichia coli* and yeast. *Frontiers in microbiology*, 2460.
- Weber, E., Engler, C., Gruetzner, R., Werner, S., & Marillonnet, S. (2011). A modular cloning system for standardized assembly of multigene constructs. *PLoS one*, 6, e16765.
- Wei, Y., Mohsin, A., Hong, Q., Guo, M., & Fang, H. (2018). Enhanced production of biosynthesized lycopene via heterogenous MVA pathway based on chromosomal multiple position integration strategy plus plasmid systems in *Escherichia coli*. *Bioresource technology*, 250, 382-389.
- Wei, L. J., Kwak, S., Liu, J. J., Lane, S., Hua, Q., Kweon, D. H., & Jin, Y. S. (2018). Improved squalene production through increasing lipid contents in *Saccharomyces cerevisiae*. *Biotechnology and bioengineering*, 115, 1793-1800.
- Wejnerowska, G., Heinrich, P., & Gaca, J. (2013). Separation of squalene and oil from *Amaranthus* seeds by supercritical carbon dioxide. *Separation and Purification Technology*, 110, 39-43.
- Weissmann, F., & Peters, J. M. (2018). Expressing multi-subunit complexes using biGBac. In *Protein Complex Assembly* (pp. 329-343). Humana Press, New York, NY.
- Wu, W., Liu, F., & Davis, R. W. (2018). Engineering *Escherichia coli* for the production of terpene mixture enriched in caryophyllene and caryophyllene alcohol as potential aviation fuel compounds. *Metabolic Engineering Communications*, 6, 13-21.
- Xu, W., Ma, X., & Wang, Y. (2016). Production of squalene by microbes: an update. *World Journal of Microbiology and Biotechnology*, 32, 1-8.
- Xu, W., Yao, J., Liu, L., Ma, X., Li, W., Sun, X., & Wang, Y. (2019). Improving squalene production by enhancing the NADPH/NADP<sup>+</sup> ratio, modifying the isoprenoid-feeding module and blocking the menaquinone pathway in *Escherichia coli*. *Biotechnology for biofuels*, 12, 1-9.
- Yang, C., Gao, X., Jiang, Y., Sun, B., Gao, F., & Yang, S. (2016). Synergy between methylerythritol phosphate pathway and mevalonate pathway for isoprene production in *Escherichia coli*. *Metabolic engineering*, 37, 79-91.
- Yin, H., Chen, H., Yan, M., Li, Z., Yang, R., Li, Y., Wang, Y., Guan J., Mao, H., Wang, Y., & Zhang, Y. (2021). Efficient Bioproduction of Indigo and Indirubin by Optimizing a Novel Terpenoid Cyclase Xial in *Escherichia coli*. *ACS omega*, 6, 20569-20576.
- Yoon, S. H., Lee, Y. M., Kim, J. E., Lee, S. H., Lee, J. H., Kim, J. Y., Jung, K.H, Shin, Y.C., Keasling, J.D., & Kim, S. W. (2006). Enhanced lycopene production in *Escherichia coli* engineered to synthesize isopentenyl diphosphate and dimethylallyl diphosphate from mevalonate. *Biotechnology and bioengineering*, 94, 1025-1032.

- Yoon, S.H., Park, H.M., Kim, J.E., Lee, S.H., Choi, M.S., Kim, J.Y., Oh, D.K., Keasling, J.D. & Kim, S.W. (2007). Increased  $\beta$ -carotene production in recombinant *Escherichia coli* harboring an engineered isoprenoid precursor pathway with mevalonate addition. *Biotechnology progress*, 23, 599-605.
- Zeng, H., & Yang, A. (2019). Quantification of proteomic and metabolic burdens predicts growth retardation and overflow metabolism in recombinant *Escherichia coli*. *Biotechnology and Bioengineering*, 116, 1484-1495.
- Zhang, J., Zeng, H., Gu, J., Li, H., Zheng, L., & Zou, Q. (2020). Progress and prospects on vaccine development against SARS-CoV-2. *Vaccines*, 8, 153.
- Zhou, K., Zou, R., Stephanopoulos, G., & Too, H. P. (2012). Metabolite profiling identified methylerythritol cyclodiphosphate efflux as a limiting step in microbial isoprenoid production. *PLoS One*, 7, e47513.
- Zhu, C., Yamamura, S., Nishihara, M., Koiwa, H., & Sandmann, G. (2003). cDNAs for the synthesis of cyclic carotenoids in petals of *Gentiana lutea* and their regulation during flower development. *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression*, 1625, 305-308.
- Zhu, Z.T., Du, M.M., Gao, B., Tao, X.Y., Zhao, M., Ren, Y.H., Wang, F.Q. & Wei, D.Z. (2021). Metabolic compartmentalization in yeast mitochondria: Burden and solution for squalene overproduction. *Metabolic Engineering*, 68, 232-245.
- Zhuang, X., & Chappell, J. (2015). Building terpene production platforms in yeast. *Biotechnology and bioengineering*, 112, 1854-1864.



**Genome editing in rice APL2  
gene using CRISPR/Cas9**





## Chapter III. Genome editing in rice *APL2* gene using CRISPR/Cas9

### 3.0 Abstract

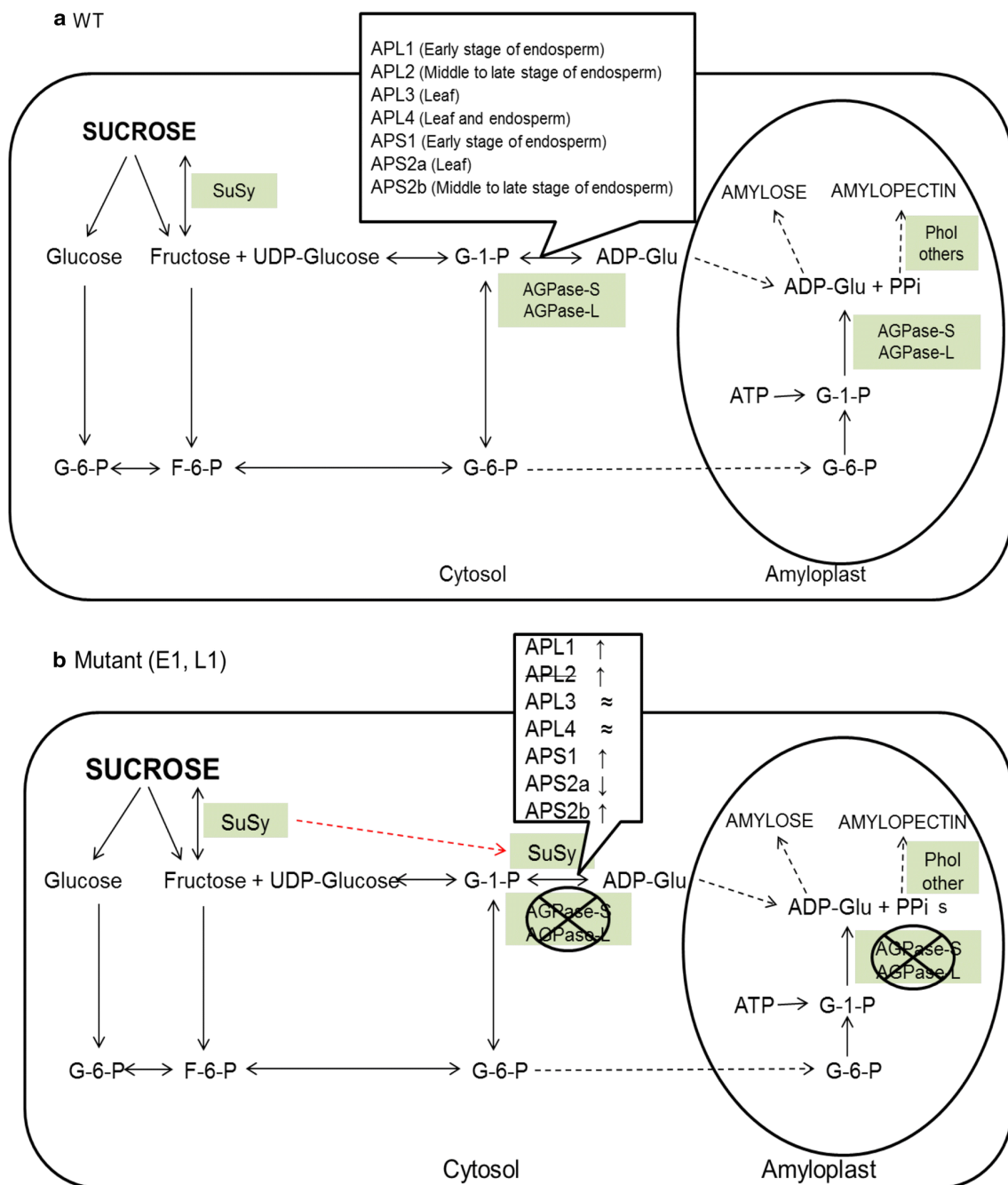
The first committed step in the endosperm starch biosynthetic pathway is catalyzed by the cytosolic glucose-1-phosphate adenylyl transferase (AGPase) comprising large and small subunits encoded by the *OsAPL2* and *OsAPS2b* genes, respectively. *OsAPL2* is expressed solely in the endosperm so we hypothesized 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.

### 3.1 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  $\alpha$ -(1,4)-linked glucose residues, whereas amylopectin contains additional  $\alpha$ -(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  $\alpha$ -(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), Phol (plastidial  $\alpha$ -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).



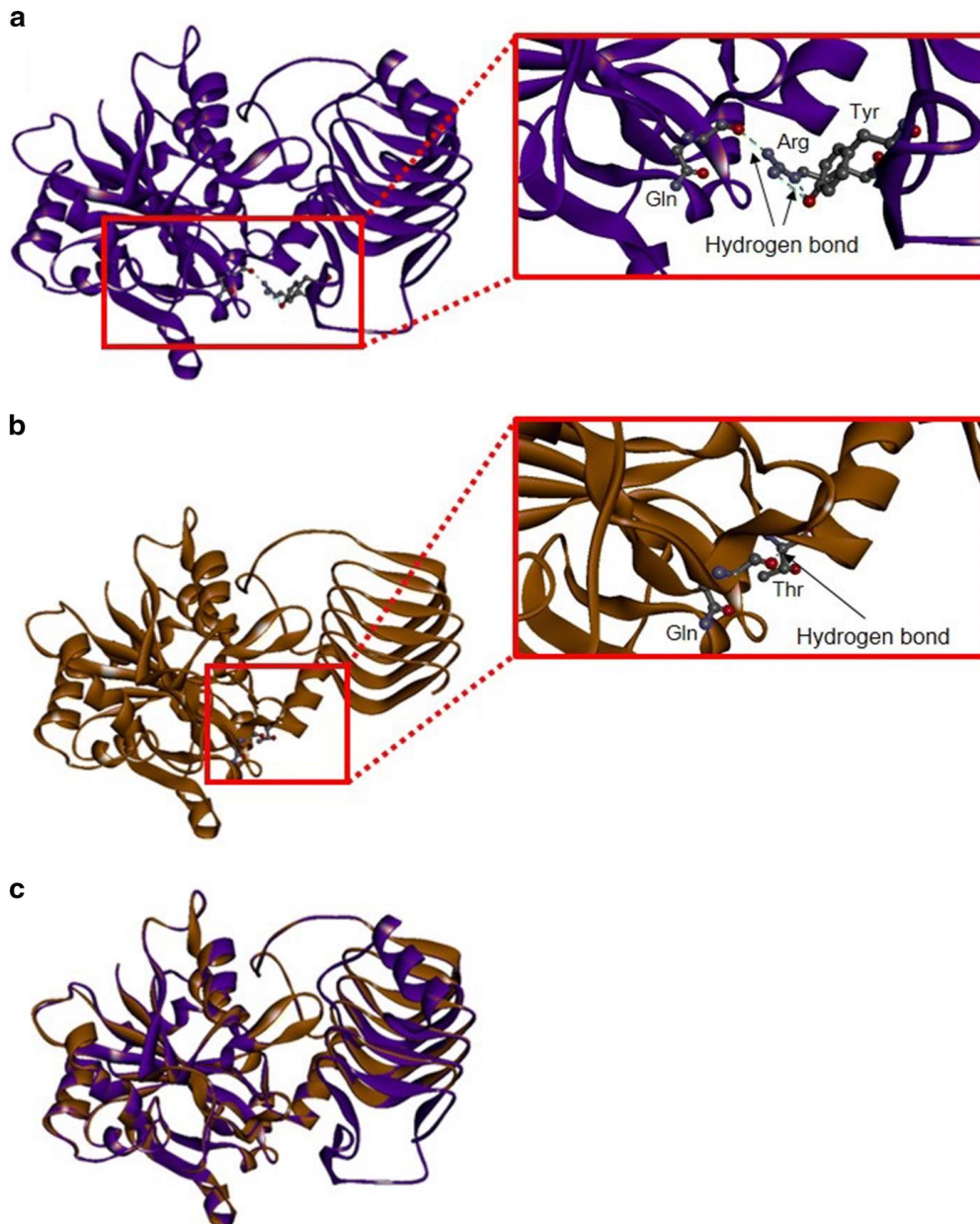
Conventional mutagenesis such as irradiation, chemical mutagenesis and T-DNA/transposon insertional mutagenesis generate random lesions in DNA sequences and require the screening of large populations to isolate useful mutants. These techniques have been largely supplanted by targeted mutagenesis using 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. 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 *OsSBEIIb* mutants accumulate higher proportion of amylose and debranched amylopectin in the seeds than normal (Sun et al. 2017). *OsSBEIIb* has also been targeted using two different sgRNAs with different activity scores and different degrees of conservation with the paralogous gene *OsSBEIIa* 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).

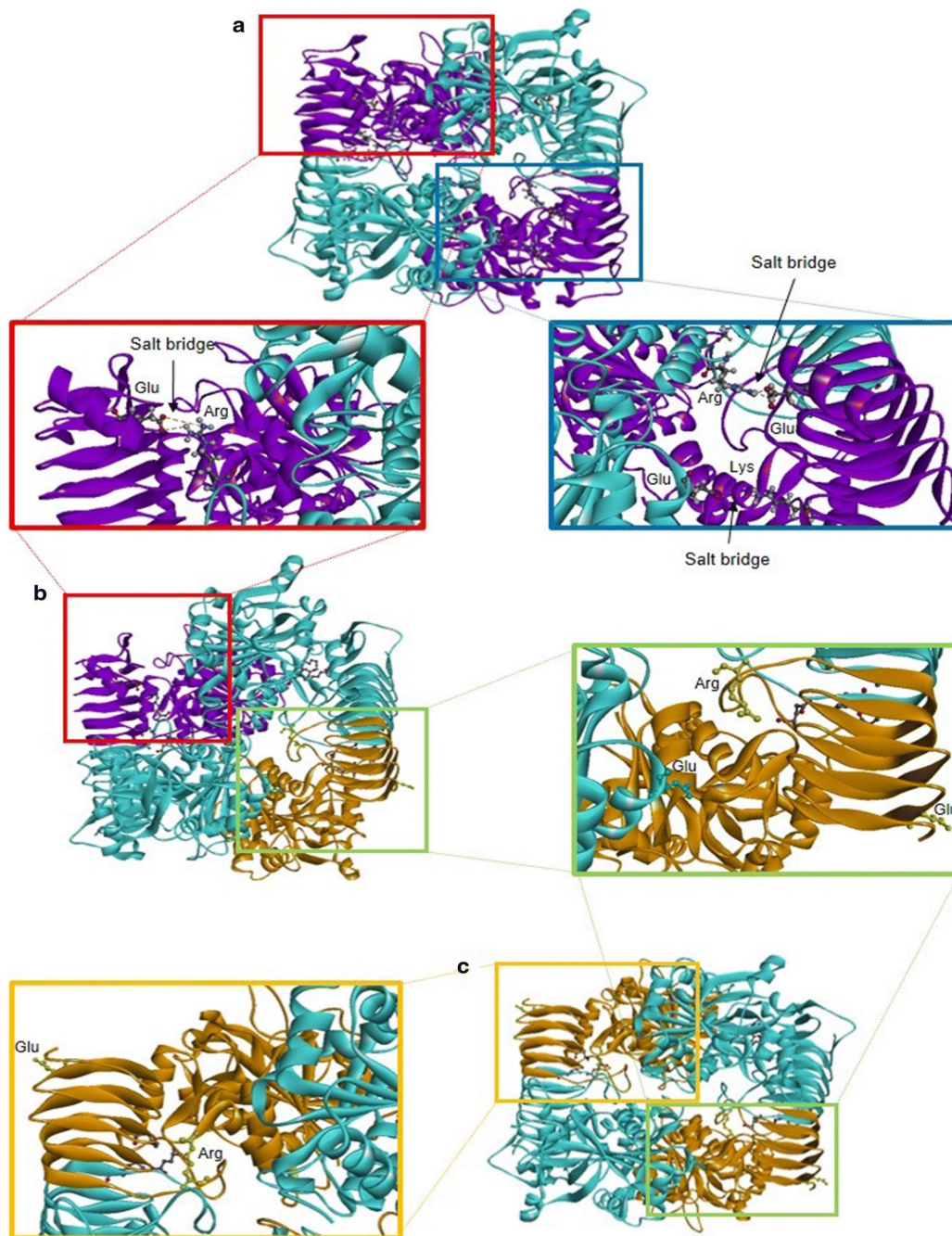
Although the CRISPR/Cas9 system has been used to target GBSS and SBE, it has not yet been used to target AGPase, which catalyzes the first step in the starch biosynthesis pathway (Tang et al. 2016; Lee et al. 2016a, b). Rather than altering the ratio of amylose and amylopectin, blocking AGPase would 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  $\alpha$ -helices and  $\beta$ -sheets that form a substrate binding cleft. A further  $\alpha$ -helix and  $\beta$ -sheet outside the catalytic center is required to maintain the correct tertiary conformation and bind with the other subunit, *APS2b*, to form the tetrameric structure (**Figures 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.



**Figure 2.** Comparisons of APL2 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 modelling 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.** Wild type and mutant comparisons of 3-D 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).

## 3.2 Materials and methods

### 3.2.1 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 1a) catalytic efficiencies were designed to minimize off-targets. The catalytic efficiency of the sgRNAs was predicted using gRNA scorer (Chari et al. 2015).

### 3.2.2 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 pJIT163-2NSCas9D10A 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 35S 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).

### 3.2.3 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 et 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 not transformed (i.e. they did not contain Cas9WT/D10A, *hpt* and sgRNA). Negative controls behaved exactly like wild type plants.

### 3.2.4 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'.

### 3.2.5 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 *E. 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 analysis.

	APS1-RT-Forward	5'-GTGCCACTTAAAGGCACCATT-3'
	APS1-RT-Reverse	5'-CCCACATTTTCAGACACGGTTT-3'
	APS2a-RT- Forward	5'-ACTCCAAGAGCTCGCAGACC-3'
	APS2a-RT- Reverse	5'-GCCTGTAGTTGGCACCCAGA-3'
	APS2b-RT- Forward	5'-AACAAATCGAAGCGCGAGAAA-3'
	APS2b-RT- Reverse	5'-GCCTGTAGTTGGCACCCAGA-3'
	APL1-RT- Forward	5'-GGAAGACGGATGATCGAGAAAAG-3'
RT-qPCR for	APL1-RT- Reverse	5'-CACATGAGATGCACCAACGA-3'
expression	APL2-RT- Forward	5'-AGTTCGATTCAAGACGGATAGC-3'
pattern	APL2-RT- Reverse	5'-CGACTTCCACAGGCAGCTTATT-3'
analysis	APL3-RT-Forward	5'-AAGCCAGCCATGACCATTG-3'
	APL3-RT- Reverse	5'-CACACGGTAGATTCACGAGACAA-3'
	APL4-RT- Forward	5'-TCAACGTCGATGCAGCAAAT-3'
	APL4-RT- Reverse	5'-ATCCCTCAGTTCTAGCCTCATT-3'
	Pho1-RT- Forward	5'-TTGGCAGGAAGGTTTCGCT-3'
	Pho1-RT- Reverse	5'-CGAAGCCTGAAGTGAACCTTGCT-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'

### 3.2.6 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 VisualizerEnzymatic activity and carbohydrate levels.

### 3.2.7 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).

### 3.2.8 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 Quantitect® 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.

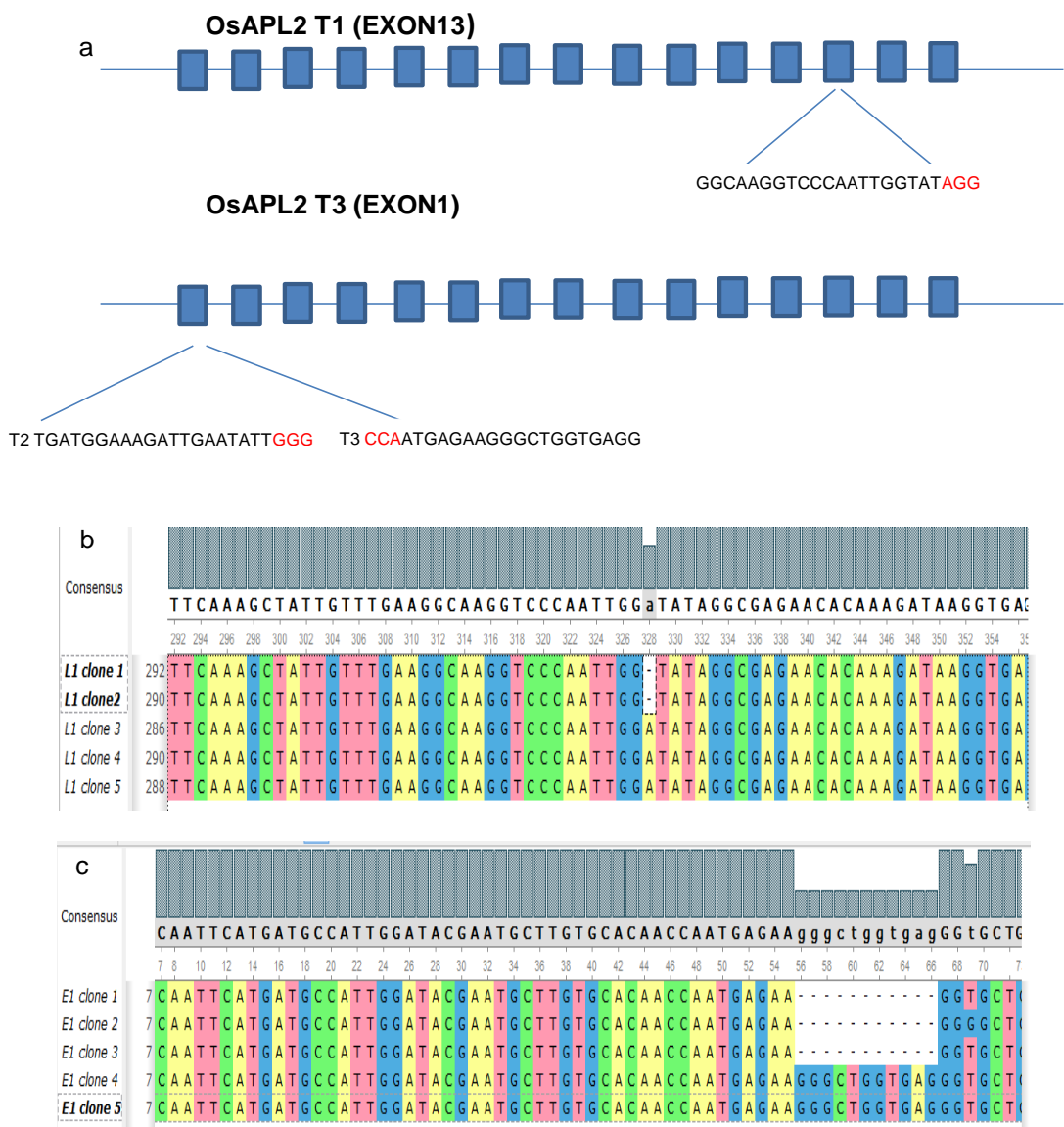
The *OsAPL1*, *OsAPL3*, *OsAPL4*, *OsAPS1*, *OsAPS2a/b*, *OsAPL2* and *OsPho1* transcripts were amplified using the primers listed in Table 1, as described by Tang et al. (2016) Primers at the end of *OsAPL2* were designed to amplify the non-common region between the WT and mutants E1 and L1 (Table 1). 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  $\Delta C_t$  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.

## 3.3 Results

### 3.3.1 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 4**. 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.





**Figure 4.** gRNA sites of the CRISPR/Cas9 system and sequencing results in the rice mutant lines.

### 3.3.2 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.

### 3.3.3 Structural comparisons

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; 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

MQFMPLDTNACAQPMRRAGEGAGTERLMERLNIGGMTQEALRKRCFGDGVGTARCV  
FTSDADRDTPHLRTQSSRKNYADASHVSAVILGGGTGVQLFPLTSTRATPAVPVGGCYRLIDI  
PMSNCFNSGINKIFVMTQFNASLNRHIIHHTYLGGGINFTDGSVQVLAATQMPDEPAGWFQ  
GTADAIKFMWILEDHYNQNNIEHVILCGDQLYRMNYMELVQKHVDDNADITISCAPIDGSR  
ASDYGLVKFDDSGRVIQFLEKPEGADLESMKVDTSFSLYAIDDKQKYPYIASMGIYVLKDVDL  
LDILKSKY AHLQDFGSEILPRAVLEHNVKACVFTEYWEDIGTIKSFDDANLALTEQPPKFEFYD  
PKTPFFTSRYLPPARLEKCKIKDAIISDGCSFSECTIEHSVIGISSRVSIGCELKDTMMMGAD  
QYETEEETSKLLFEGKVP**IGENTKIRNCIIMNARIGRNVIANTQGVQESDHPEEGYIRSG**  
**IVVILKNATIKDGTVI**

b Mutated L1 protein sequence

MQFMPLDTNACAQPMRRAGEGAGTERLMERLNIGGMTQEALRKRCFGDGVGTARCV  
FTSDADRDTPHLRTQSSRKNYADASHVSAVILGGGTGVQLFPLTSTRATPAVPVGGCYRLIDI  
PMSNCFNSGINKIFVMTQFNASLNRHIIHHTYLGGGINFTDGSVQVLAATQMPDEPAGWFQ  
GTADAIKFMWILEDHYNQNNIEHVILCGDQLYRMNYMELVQKHVDDNADITISCAPIDGSR  
ASDYGLVKFDDSGRVIQFLEKPEGADLESMKVDTSFSLYAIDDKQKYPYIASMGIYVLKDVDL  
LDILKSKY AHLQDFGSEILPRAVLEHNVKACVFTEYWEDIGTIKSFDDANLALTEQPPKFEFYD  
PKTPFFTSRYLPPARLEKCKIKDAIISDGCSFSECTIEHSVIGISSRVSIGCELKDTMMMGAD  
QYETEEETSKLLFEGKVP**IGYRREHKDKELYRHEC**

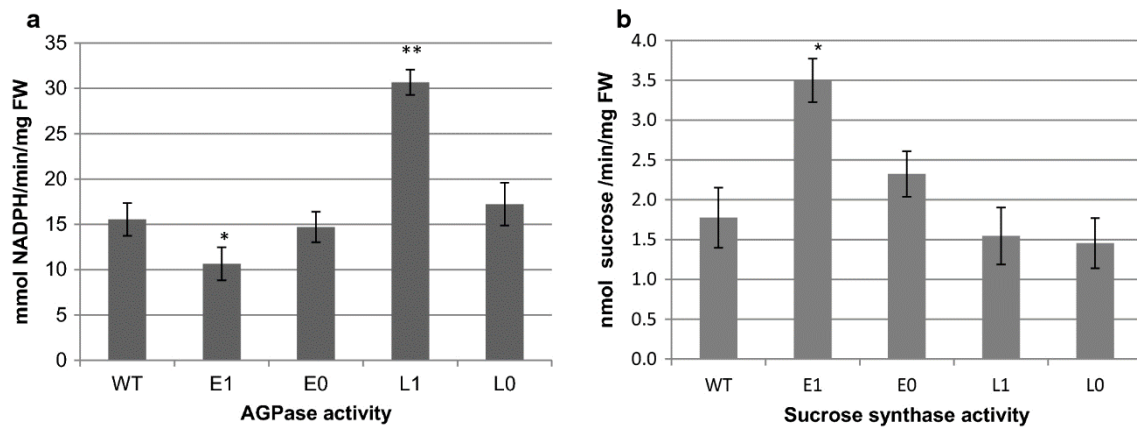
c Mutated E1 protein sequence

MQFMPLDTNACAQPMRRCWD**Stop**

**Figure 5.** Expected APL2 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.

### 3.3.4 Analysis of AGPase and sucrose synthase activity

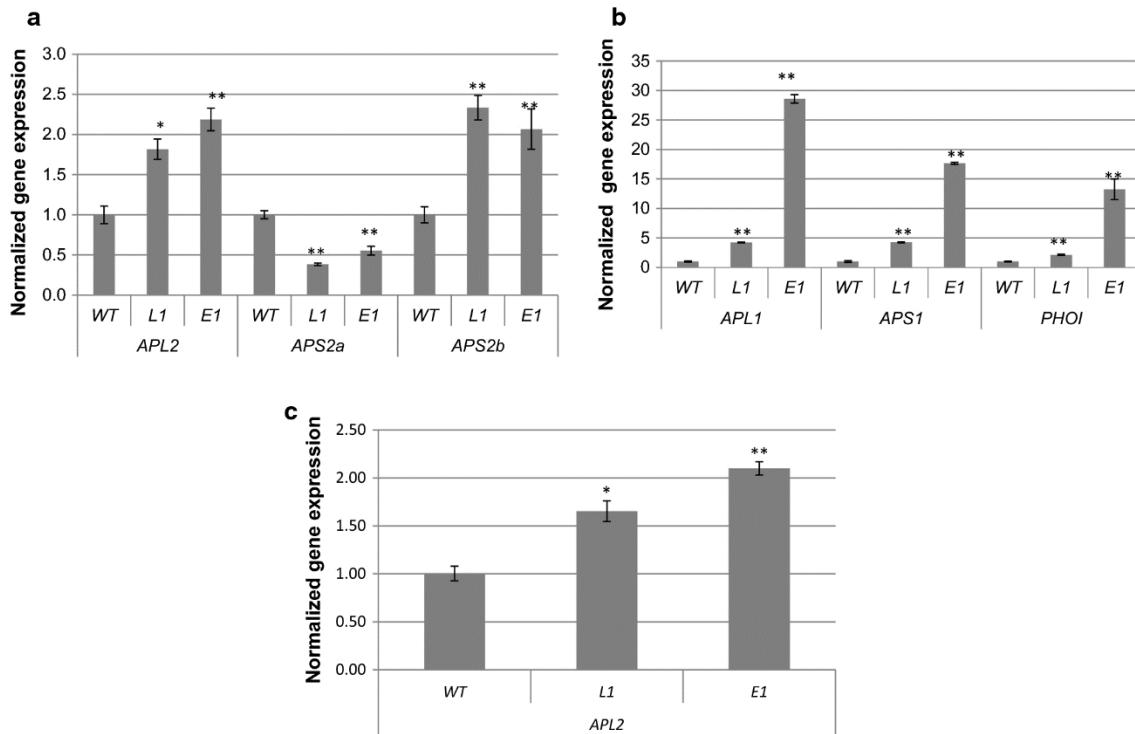
We determined the direct impact of each mutation on AGPase activity by comparing enzyme activity in the flag leaves of the mutant and wild-type plants as described by Nishi et al. (2001), using three biological replicates and two technical replicates for each biological sample (**Figure 6**). No analysis was possible in seed as both mutants were infertile. AGPase activity in mutant E1 was low ( $10.6 \pm 1.8$  nmol NADPH/min/mg FW) whereas activity in mutant L1 was much higher ( $30.7 \pm 1.4$  nmol NADPH/min/mg FW) than wild-type plants ( $15.5 \pm 1.8$  nmol NADPH/min/mg FW). AGPase activity in the corresponding negative controls was similar to wild-type levels, as expected (**Figure 6a**). 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 ( $3.5 \pm 0.27$  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.** Enzymatic activity in rice flag leaves. (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  $\pm$  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 (\* $P < 0.05$ , \*\* $P < 0.001$ ).

### 3.3.5 Analysis of AGPase family gene expression

Next, we measured the expression of *OsAPL2* 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 (**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.



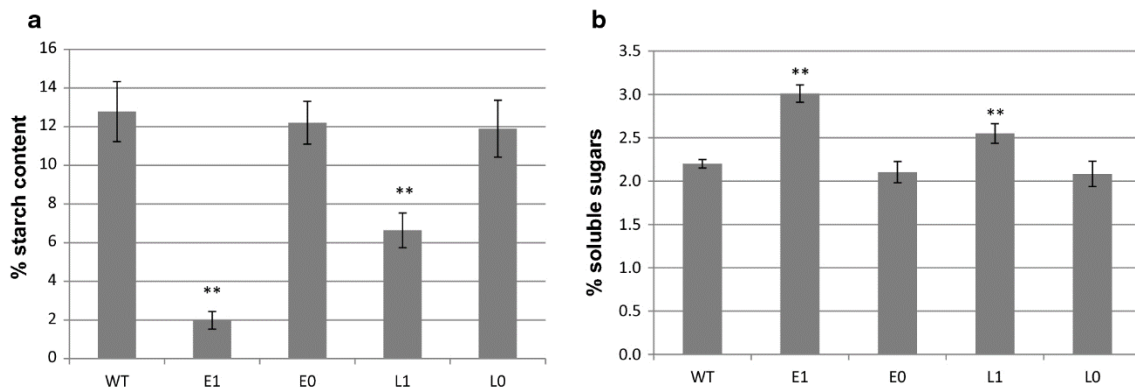
**Figure 7.** Relative expression levels of rice AGPase family genes in the flag leaves of wild-type and mutants. (a) *OsAPL2*, *OsAPS2a* and *OsAPS2b*. (b) *OsAPL1*, *OsAPS1* and *OsPHOI*. (c) *OsAPL2* amplifying C-terminal end of the protein (non-common between WT and mutants). Values are means  $\pm$  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**).

### 3.3.6 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  $\pm$  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 (\* $P < 0.05$ ; \*\* $P < 0.01$ ).

### 3.4 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 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 (Sato et al. 2008; Steup 1990; Yu et al. 2001). There are two types of phosphorylases that differ in terms of structure, kinetics, expression patterns and subcellular location (Sato 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; Sato 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 *shrunk* 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 *phol* 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 *Phol* 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. 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-terminal 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-terminal 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  $\beta$ -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). *OsPho1* was also induced in L1 and E1 perhaps reflecting its key role in the initiation of starch biosynthesis (Sato 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).

### **3.5 Conclusion**

In conclusion, 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.

### 3.6 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, Dauvillè D, Moulle G et al (1999) Genetic and biochemical evidence for the involvement of  $\alpha$ -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  $p$ -hydroxyphenylpyruvate dioxygenase shift tocopherol synthesis from the  $\gamma$  to the  $\alpha$  isoform without increasing absolute tocopherol levels. *Transgenic Res* 21:1093-1097
- Fausser 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* 337:816-821

- 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, Takaiwa F, 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 C3 reductive cycle and carbohydrate synthesis. In: Tolbert NE, Preiss J (eds) *Regulation of atmospheric CO<sub>2</sub> and O<sub>2</sub> 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: Pessaraki 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, 3rd 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 4

**Genome editing in rice  
Waxy/GBSSI gene using  
CRISPR/Cas9**



## Chapter IV. Genome editing in rice *Waxy* gene using CRISPR/CAS9

### 4.0 Abstract

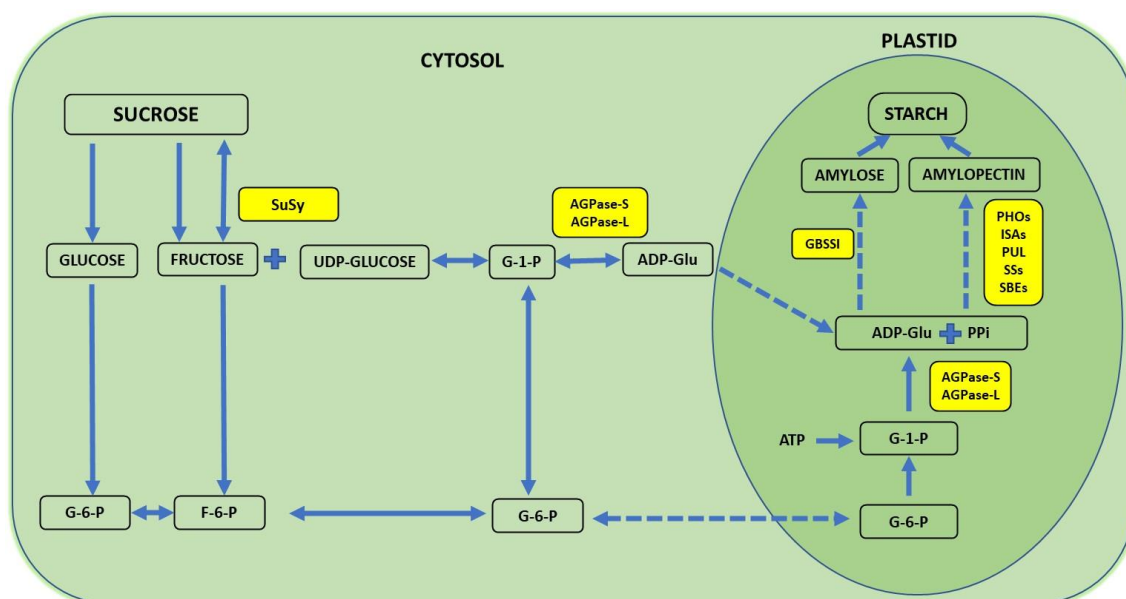
The mutation of genes in the starch biosynthesis pathway has a profound effect on starch quality and quantity and is an important target for plant breeders. Mutations in endosperm starch biosynthetic genes may impact starch metabolism in vegetative tissues such as leaves in unexpected ways due to the 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 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.

### 4.1 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  $\alpha$ -(1,4)-linked glucose monomers whereas amylopectin contains additional  $\alpha$ -(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). Differences in enzyme activities may thus induce changes in starch composition.

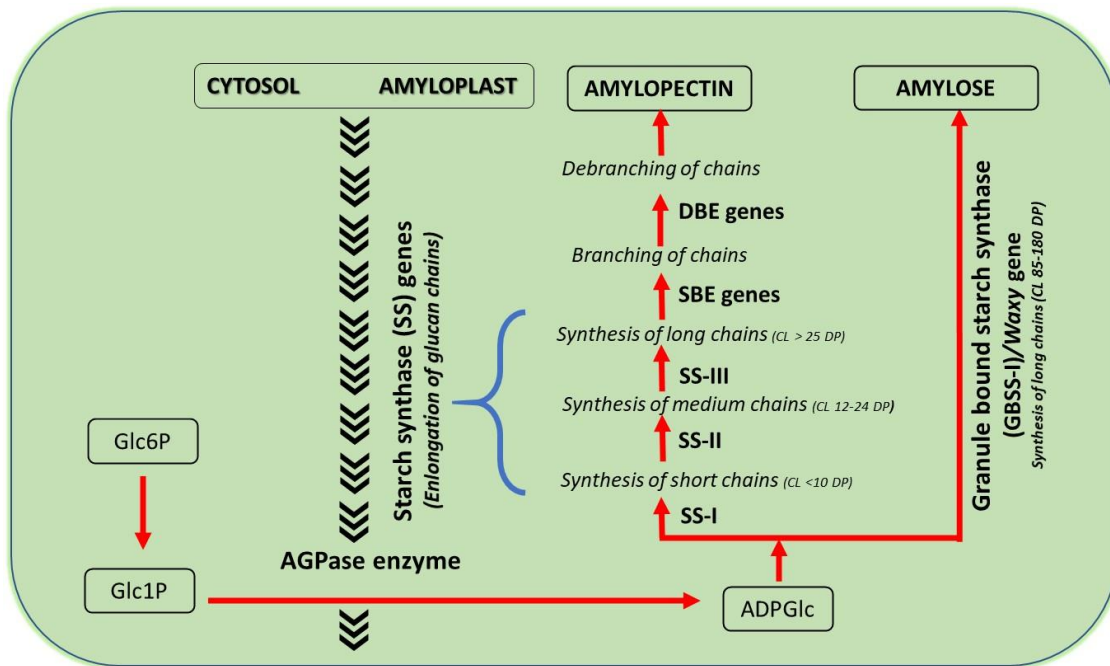
Starch synthesis in plants begins with the conversion of glucose 1-phosphate to ADP-glucose, catalyzed by ADP-glucose pyrophosphorylase (AGPase). This is followed by the polymerization of ADP-glucose to form amylose and amylopectin via the coordinated activity of AGPase and starch synthases (SSs), which form the  $\alpha$ -(1,4)-linked glycosidic bonds of both molecules, and starch-branching enzymes (SBEs), which form the  $\alpha$ -(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  $\alpha$ -(1,4)-linked glucan primer (Tetlow et al. 2004) whereas SBEs cleave internal amylose  $\alpha$ -(1,4) glycosidic bonds and transfer the reducing ends to C6 hydroxyls to form

$\alpha$ -(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). Disproportionating enzyme (DPE) and starch phosphorylase (PHO) also play a role in the initiation and elongation of starch polymers (Sato 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). 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  $\alpha$ -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 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  $\alpha$ -helices and  $\beta$ -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 Fujimoto 2012).



**Figure 2.** Steps in the starch biosynthesis pathway that generate the different components of starch found in rice endosperm. DBE: Debranching enzyme; SBE: starch branching enzyme

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 indica subspecies, and has strong GBSSI activity, and thus more amylose in the endosperm. *Wxb* has weaker GBSSI activity and the amount of amylose and amylopectin in the endosperm is more variable (Hirano and Sano 1998; Umemoto and Terashima 2002). Because of its impact on grain quality, *Wx* is an important 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 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 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 (Ma et al. 2015; 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. In one report (Ma et al. 2015), wild-type Cas9 (Cas9WT) was used to target three different sites simultaneously of GBSSI gene, 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). The starch

pathway is tightly regulated by the ratio of triose-phosphate to inorganic phosphate, and the disruption of this balance may lead to further changes in the expression of other genes involved in starch metabolism (Preiss 1982). In a *Wx* mutant generated by irradiation (*GM077*), the loss of GBSSI activity and lower starch and amylose levels induced the expression of GBSSII, AGPases, starch synthases and isoamylases (Zhang et al. 2012). Given that the mutation of endosperm-specific *GBSSI* is viewed as a good strategy to modulate amylose production in the endosperm without affecting amylose metabolism in vegetative tissues, impact on GBSSII expression (affecting starch biosynthesis in the leaves) and on enzymes related to amylopectin biosynthesis and starch degradation needs to be investigated in detail.

We hypothesized that at the protein level GBSSI enzyme activity might be lost in part or entirely depending on the nature of the mutation. In the former case we anticipated endosperm amylose levels to be reduced with no further impact on starch metabolism, as reported in other studies (Ma et al. 2015; Zhang et al. 2018). We, therefore, used the CRISPR/Cas9 system to generate truncated (nonfunctional) or partially active GBSSI mutants in rice endosperm and carried out a comprehensive analysis of starch and soluble sugar levels, grain morphology and analyzed the expression of a number of genes involved in starch metabolism. The overarching objective was to establish a mechanistic basis to support further targeted interventions to generate rice grains with varying starch content and composition for various applications.

## 4.2 Materials and methods

### 4.2.1 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' (D10A CRISPR/Cas9, targeted nucleotides 1337-1359), TS2 = 5'-GCTGCTCCGCCACGGGT↓TCCAGG-3' (D10A CRISPR/Cas9, targeted nucleotides 1377-1399) and TS3 = 5'-CCGGCTTCGGCATCGCC↓GACAGG-3' (WT CRISPR/Cas9, targeted nucleotides 5082-5104), where the arrows indicate the expected site of the DSB.

### 4.2.2 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 35S 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).

#### **4.2.3 Rice transformation and recovery of transgenic plants**

Seven-day-old mature zygotic embryos (*Oryza sativa* ssp. Japonica 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 hours 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 (Sudhakar et al. 1998; Valdez et al. 1998). The embryos were returned to osmotic medium for 12 hours 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. We compared two *Wx* *Oryza sativa* ssp. Japonica mutants [KUR and Mushashimochi (Musa)] with our mutant lines.

#### **4.2.4 Confirmation of the presence of cas9 and gRNA in regenerated rice plants**

Genomic DNA was isolated from 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'.

#### **4.2.5 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 *E. 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.

#### **4.2.5 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.

#### **4.2.6 Enzymatic activity assays**

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

The GBSS activity in 10 pooled 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<sub>2</sub>, 2 mM EDTA, 50 mM 2-mercaptoethanol, 12.5% (v/v) glycerol and 5% (w/v) polyvinylpyrrolidone 40. Thirty (30) µL of the homogenate were added to 1.8 mL of the same buffer and centrifuged at 2000×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 40 mM phosphoenolpyruvate, 50 mM MgCl<sub>2</sub>, 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.

#### **4.2.7 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 discarded transparent seeds of the mutated lines in the T1 generation as they represented the wild-type phenotype. All remaining seeds were opaque. We pooled 20 such seeds and processed them to powder for all subsequent analyses. 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.

#### **4.2.8 RNA extraction and real-time quantitative RT-PCR analysis**

Total leaf/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 *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  $\Delta C_t$  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) gene was used as an internal control.

#### **4.2.9 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 Reichert 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, Hanau, Germany). Grains were cut through the center to expose the endosperm. Two 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. They were then processed for scanning electron microscopy (SEM) by dehydration at 60°C for 24 hours 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 Jeol JSM-6300. Seed length, width and thickness were

measure using Digimatic Caliper CD-6"CX. Seed volume was calculated multiplying length, width and thickness of seeds.

#### **4.2.10 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.

#### **4.2.11 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)).

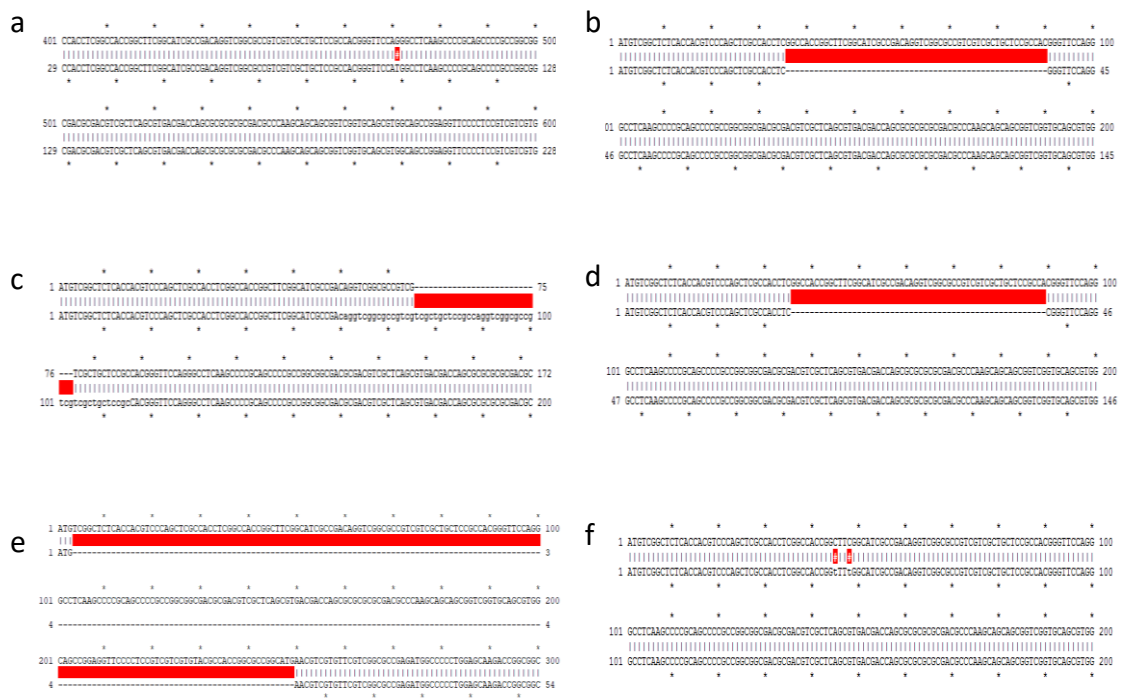
## 4.3 Results

### 4.3.1 Recovery and characterization of GBSSI mutants

We analyzed six induced mutant lines in detail firstly in T1 seeds (**Table 1**). PCR analysis revealed that line 1 featured the substitution of a single nucleotide at site (**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. In T1 plants, all the lines contained heterozygous mutations except line 2, in which the mutation was homozygous. In T2 plants, all the lines contained homozygous mutations. 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-target sites 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 1.** 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

Rice 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



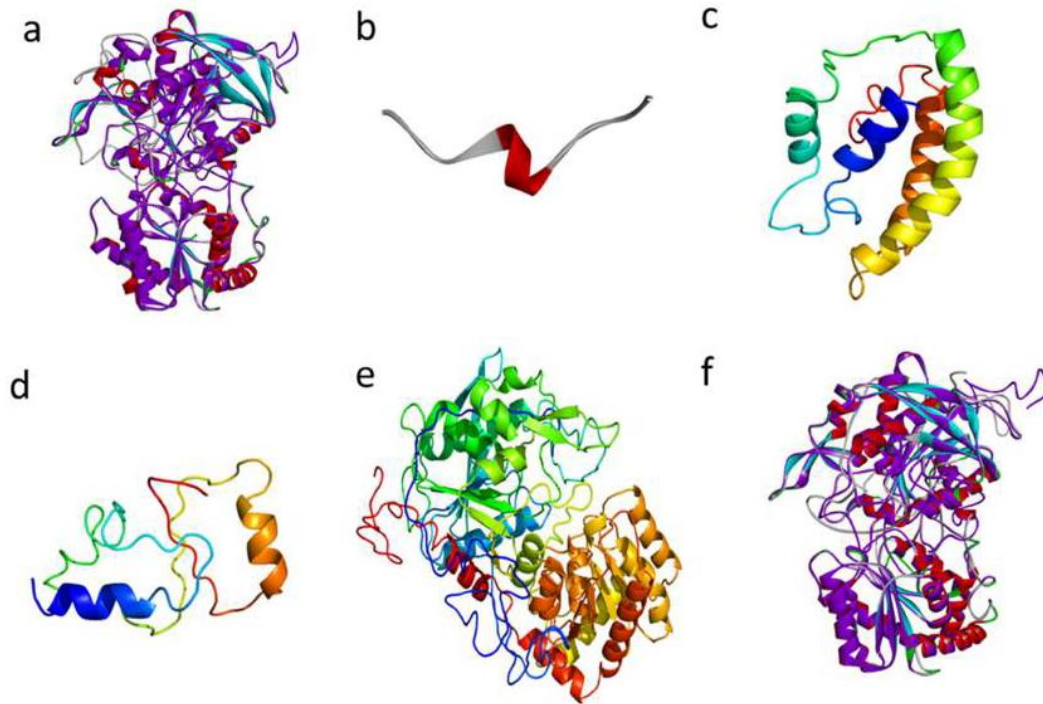
**Figure 3.** The gRNA target sites and sequencing results showing the nature of the six *OsWaxy* mutation in the lines 1 to 6.

#### 4.3.2 Structural comparisons and phylogenetic analysis of protein sequence

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) MSALTTSQLATSATGFGIADRSAPSSLRHGFQGLKPRSPAGGDATLSVTTTSAR  
ATPKQQRSVQRGSRFRFSSVVVYATGAGMNVVFGAEMAPWSKTGGLGDVLG  
GLPPAMAANGHRVMVISPRYDQYKDAWDTSVVAEIKVADRYERVRFHFCYK  
RGVDRVFDHPSFLEKVVWGTGTEKIYGPDTGVDYKDNQMRFSLLCQAALAPR  
ILNLNPNPYFKGTYGEDVVFVNCNDWHTGPLASYLKNNYQPNGIYRNAKVAFCI  
HNISYQGRFAFEDYPELNLSEFRSSDFIDGYDTPVEGRKINWMKAGILEADR  
VLTVSPYYAEELISGIARGCELDNIMRLTGITGIVNGMDVSEWDPKDKYITAK  
YDATTAEAKALNKEALQAEAGLPVDRKIPLIAFIGRLEEQKGPVMAAAIPEL  
MQEDVQIVLLGTGKKKFEKLLKSMEEKYPGKVRVVKFNAPLAHLIMAGADV  
LAVPSRFPCGLIQLQGMRYGTPCACASTGGLVDTVIEGKTGFHMGRLSVDCR  
VVEPSDVKKVAATLKRAIKVVGTPAYEEMVRNCMNQDLSWKGPAKNWENVL  
LGLGVAGSAPGIEGDEIAPLAKENVAAP
- b) MSALTTSQLATSATGFGIADRSAPSSLRHGFHGLKPRSPAGGDATLSVTTTSAR  
ATPKQQRSVQRGSRFRFSSVVVYATGAGMNVVFGAEMAPWSKTGGLGDVLG  
GLPPAMAANGHRVMVISPRYDQYKDAWDTSVVAEIKVADRYERVRFHFCYK  
RGVDRVFDHPSFLEKVVWGTGTEKIYGPDTGVDYKDNQMRFSLLCQAALAPR  
ILNLNPNPYFKGTYGEDVVFVNCNDWHTGPLASYLKNNYQPNGIYRNAKVAFCI  
HNISYQGRFAFEDYPELNLSEFRSSDFIDGYDTPVEGRKINWMKAGILEADR  
VLTVSPYYAEELISGIARGCELDNIMRLTGITGIVNGMDVSEWDPKDKYITAK  
YDATTAEAKALNKEALQAEAGLPVDRKIPLIAFIGRLEEQKGPVMAAAIPEL  
MQEDVQIVLLGTGKKKFEKLLKSMEEKYPGKVRVVKFNAPLAHLIMAGADV  
LAVPSRFPCGLIQLQGMRYGTPCACASTGGLVDTVIEGKTGFHMGRLSVDCR  
VVEPSDVKKVAATLKRAIKVVGTPAYEEMVRNCMNQDLSWKGPAKNWENVL  
LGLGVAGSAPGIEGDEIAPLAKENVAAP
- c) MSALTTSQLATSGSRASSPAAPPAATRRRSA
- d) MSALTTSQLATSATGFGIADRSAPSSLLKQVAVVAAPPRVPGPQAPQPRRRR  
DVAQRDDQQRADAQAAA VGAAWQPEVPLRRRVRRRRHERRVRRRDGPLE  
QDRRPR
- e) MSALTTSQLATSVPGPQAPQPRRRRDVAQRDDQQRADAQAAAVGAAWQPE  
VPLRRRVRRRRHERRVRRRDGPLEQDRPR
- f) MSALTTSQLATSATGFGIADRSAPSSLRHGFQGLKPRSPAGGDATLSVTTTSAR  
ATPKQQRSVQRGSRFRFSSVVVYATGAGM  
MNVVFGAEMAPWSKTGGLGDVGGGLPPAMAANGHRVMVISPRYDQYKDA  
WDTSVVAEIKVADRYERVRFHFCYKRGVDRVFDHPSFLEKVVWGTGTEKIYGP  
DTGVDYKDNQMRFSLLCQAALAPRILNLNPNPYFKGTYGEDVVFVNCNDWHI  
GPLASYLKNNYQPNGIYRNAKVAFCIHNISYQGRFAFEDYPELNLSEFRSSDFI  
IDGYDTPVEGRKINWMKAGILEADRVLTVSPYYAEELISGIARGCELDNIMRLT  
GITGIVNGMDVSEWDPKDKYITAKYDATTAEAKALNKEALQAEAGLPVDRK  
IPLIAFIGRLEEQKGPVMAAAIPELMQEDVQIVLLGTGKKKFEKLLKSMEEKY  
PGKVRVVKFNAPLAHLIMAGADVAVPSRFPCGLIQLQGMRYGTPCACAST  
GGLVDTVIEGKTGFHMGRLSVDCR VVEPSDVKKVAATLKRAIKVVGTPAYEE  
MVRNCMNQDLSWKGPAKNWENVLLGLGVAGSAPGIEGDEIAPLAKENVAAP

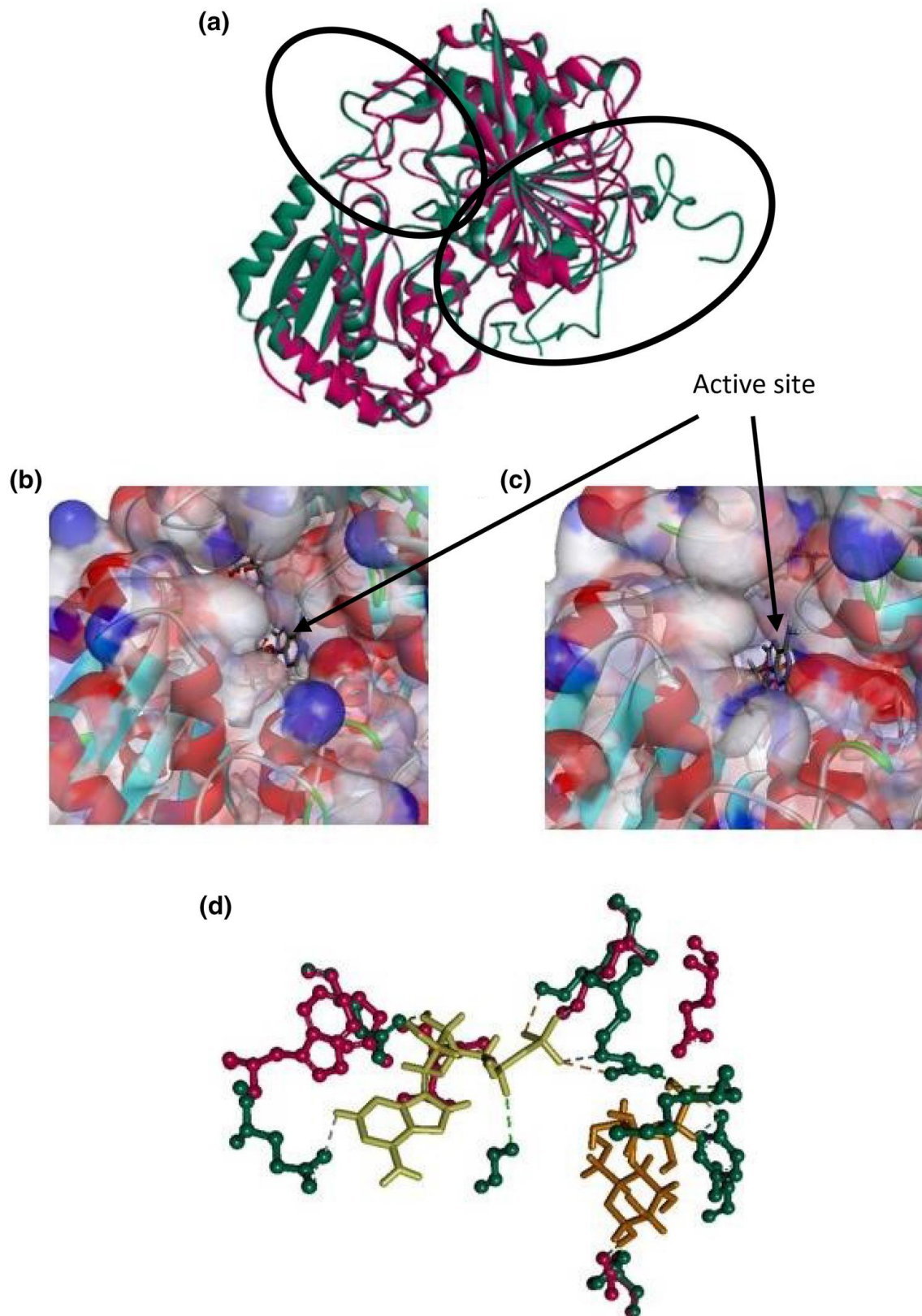
**Figure 4.** GBSSI predicted protein sequences encoded by each of the six mutated *Wx* alleles. a) Wild-type Waxy protein sequence, the amino acid that changes respect to line 1 is highlighted in yellow. Line 6 has the same protein sequence as wild-type protein; b) Waxy expected protein sequence in Line 1, with the changed amino acid in yellow; c) Waxy expected protein sequence in Line 2, with the changed amino acids in yellow; d) Waxy expected protein sequence in Line 3, with the changed amino acids in yellow; e) Waxy expected protein sequence in Line 4 in yellow deletion between these two aminoacids; f) Waxy expected protein sequence in Line 5, with the deleted amino acids in yellow.



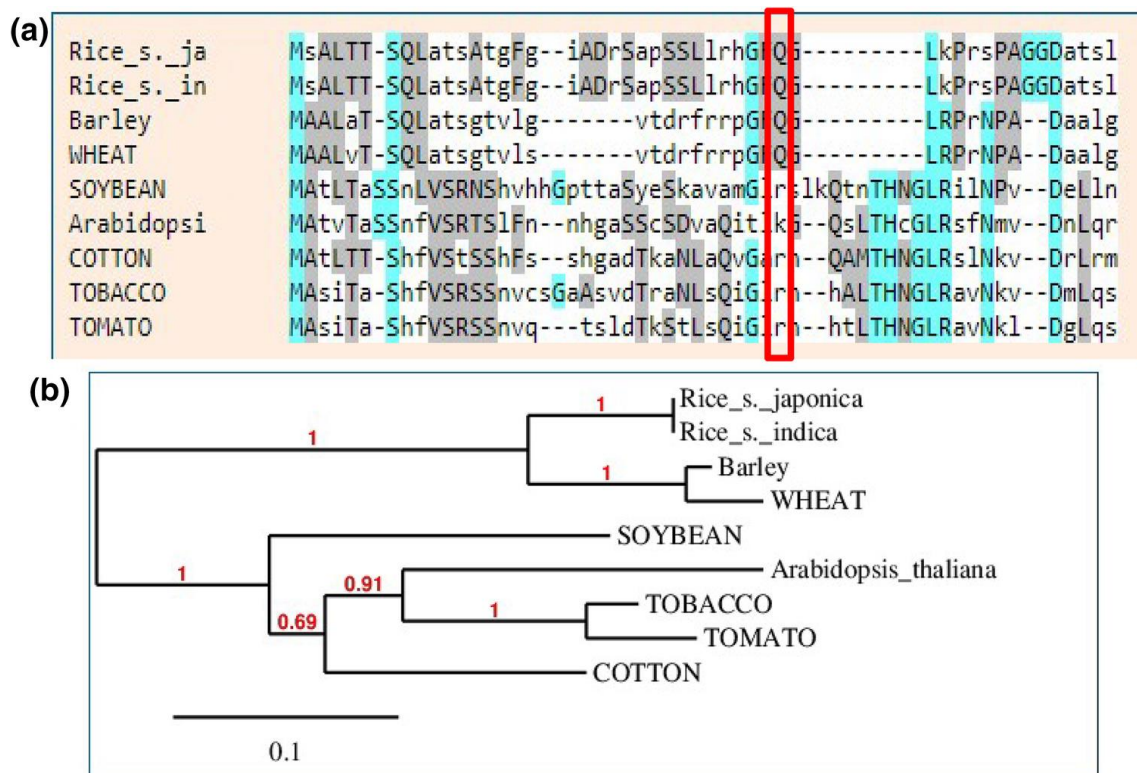
**Figure 5.** GBSSI predicted protein structures encoded by each of the six mutated *Wx* alleles, superimposed over the wild-type structure. a) line 1 superimposed with WT b) line 2 c) line 3 d) line 4 e) line 5 f) line 6 superimposed with WT

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 (indicated with black arrows) 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. A phylogenetic tree was constructed from the GBSSI sequences of the japonica and indica rice subspecies, 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 into starch granules in cereals and that the line 1 mutant may, therefore, suffer from the inefficient import of the enzyme into starch granules.





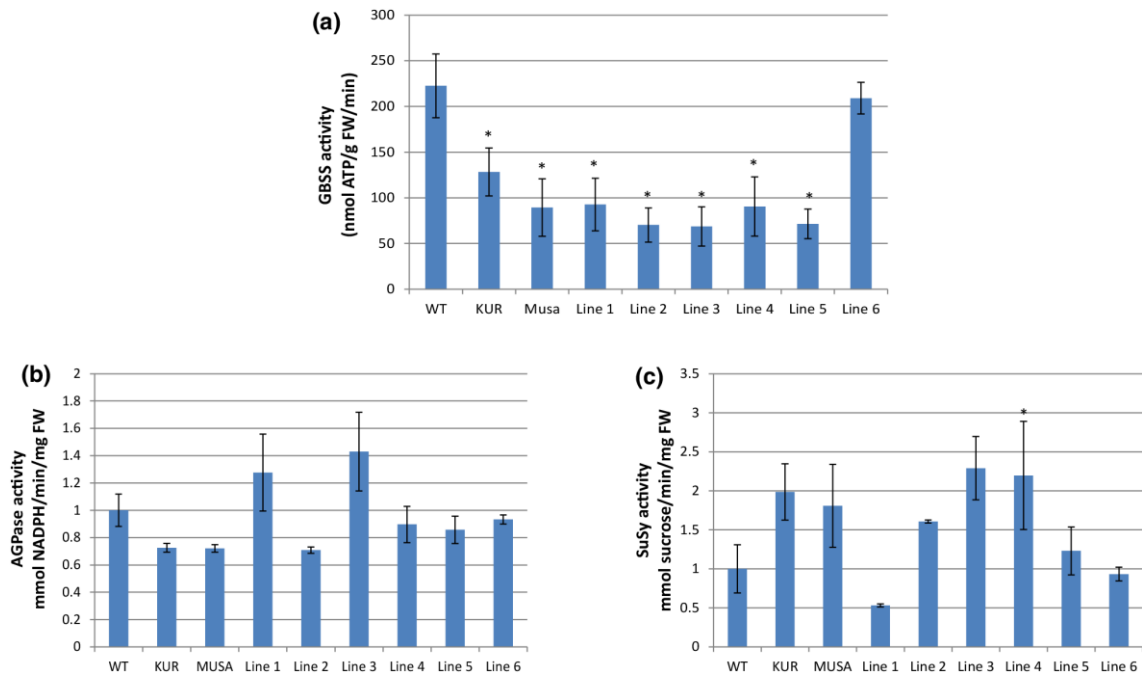
**Figure 6.** Structure of GBSSI in wild-type rice and mutant line 1. (a) Superimposition of wild-type GBSSI (pink) and the line 1 mutant (green). The principal structural differences are encircled by dashed lines. (b) Wild-type surface model with ADP and malto-oligosaccharide in the catalytic site indicated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

#### 4.3.3 Changes in enzymatic activity of GBSS by the loss of GBSSI 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 Oryza sativa* ssp. Japonica mutants, namely KUR generated by exposure to neutrons (Yatou and Amano 1991) and Musashimochi generated by exposure to gamma rays (Itoh et al. 1997). Consistent with the observation that line 6 did not exhibit any changes at the protein level, 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**).

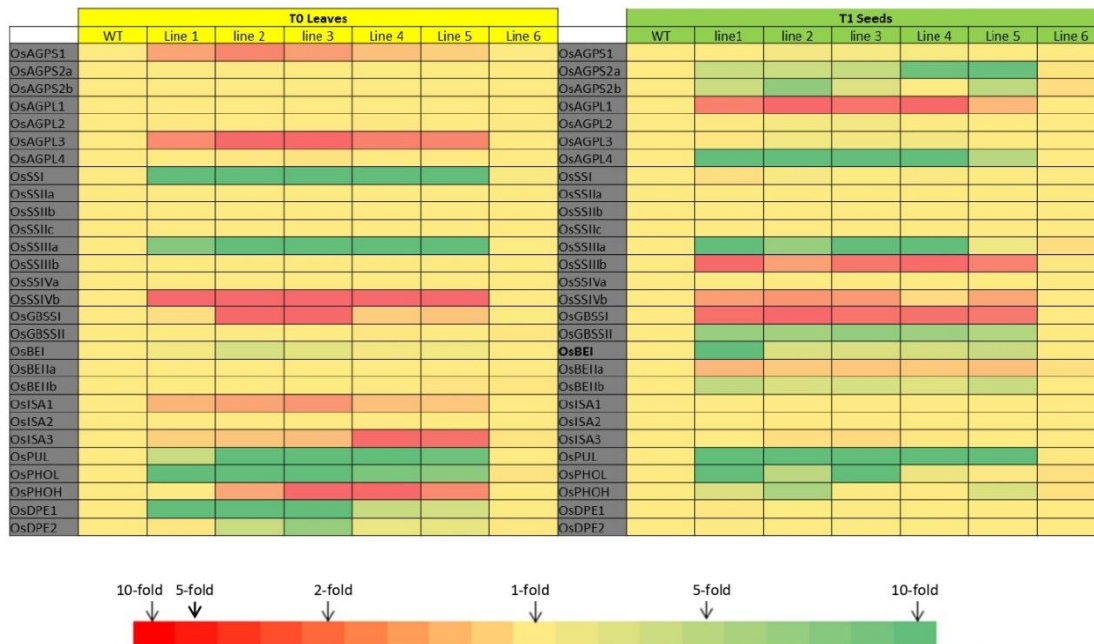


**Figure 8.** Enzyme activity in wild-type and mutant rice plants. (a) GBSS1 activity in T2 seeds of wild-type (WT) plants, the Wx mutant lines and the two Wx irradiation mutants, KUR and Musa. (b) AGPase activity in the T0 flag leaves of WT and mutant lines. (c) Sucrose synthase activity in the T0 flag leaves of WT and mutant lines. Values are means  $\pm$  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 in T2 seeds. 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%; line 6 showed similar AGPase/SuSy activity compared to the wild-type. Although the trends were clear, these differences were not statistically significant. Conversely, almost all mutant 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 modest 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. 8b, c**).

#### 4.3.4 Deregulation of starch-related family gene expression induced by GBSSI mutations

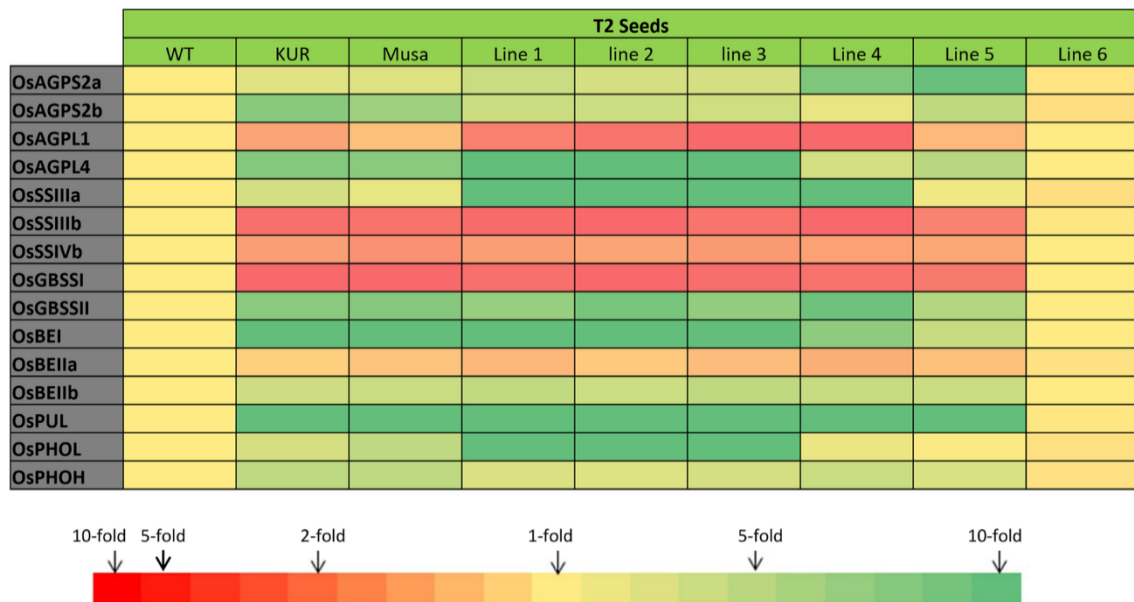
Next, we measured the expression of the rice *Wx* gene in T0 flag leaves and T1 (Figure 9) and T2 (Figure 10) endosperm and compared the expression of other genes involved in starch biosynthesis (*OsAPL1*, *OsAPL3*, *OsAPL4*, *OsAPS1*, *OsAPS2a/b*, *OsAPL2*, *OsSSI*, *OsSSIIa*, *OsSSIIb*, *OsSSIIc*, *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 ( $\geq 10$ -fold) downregulation compared to wild-type and the red gradient represents moderate ( $< 10$ -fold but  $\geq 1.12$ -fold) downregulation compared to wild-type. The intense green color represents extreme ( $\geq$  tenfold) upregulation compared to wild-type and the green gradient represents moderate ( $<$  tenfold but  $\geq$  twofold) upregulation compared to the wild-type. All other values close to onefold (no change) are colored yellow.



**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. 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.

In T0 leaves (Figure 9), 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 10). Genes with no changes in levels of expression in T2 with respect to wild-type were not included in the heat map. T2 seeds exhibited a

similar expression profile with the, KUR and Musa mutants and also with T1 seeds (Figure 10).



**Figure 10.** Heat map showing fold-changes in 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.

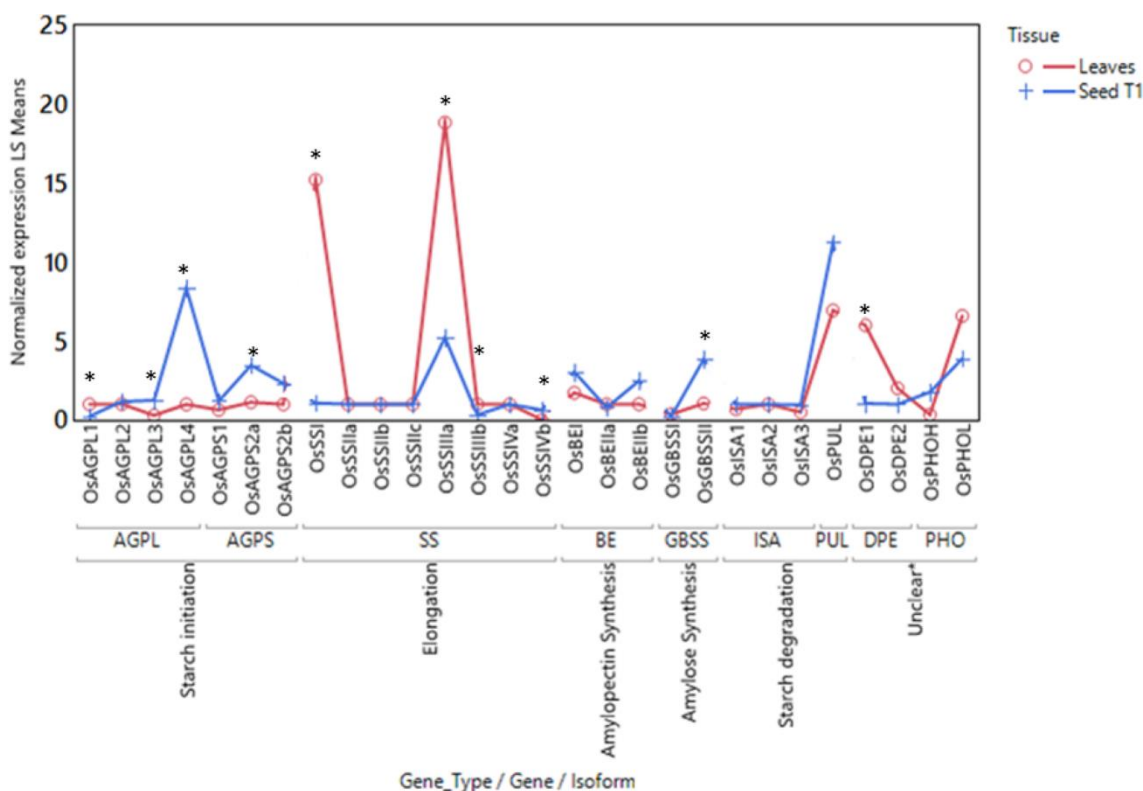
An analysis of variance for normalized expression on log-transformed data of T0 leaves and T1 seeds was carried out using Tissue (leave/seed); Genotype (Wild-type and separate mutant lines); Gene (Family of genes); Gene\_Type (Function of genes) and isoform (different isoforms of the same gene) as independent 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 2.** 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 semi-squares partial R <sup>2</sup> (%)		Mean square	F ratio	Prob > F
C. total	391	779.27				
Model	145	558.56	71.7	3.85	4.29	< 0.0001*
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	< 0.0001*
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*	30	13.31	2.4	0.44	0.49	0.9885
Gene_Type						
Gene[Gene_Type]	3	61.05	10.9	20.35	22.68	< 0.0001*
Gene*Tissue	3	11.47	2.1	3.82	4.26	0.0059*
[Gene_Type]						
Gene*Genotype	18	30.38	5.4	1.69	1.88	0.0180*
[Gene_Type]						
Isoform [Gene_Type,Gene]	19	251.65	45.1	13.24	14.76	< 0.0001*
Isoform*Tissue	19	113.05	20.2	5.95	6.63	< 0.0001*
[Gene_Type,Gene]						
Error	246	220.71	28.3	0.90		

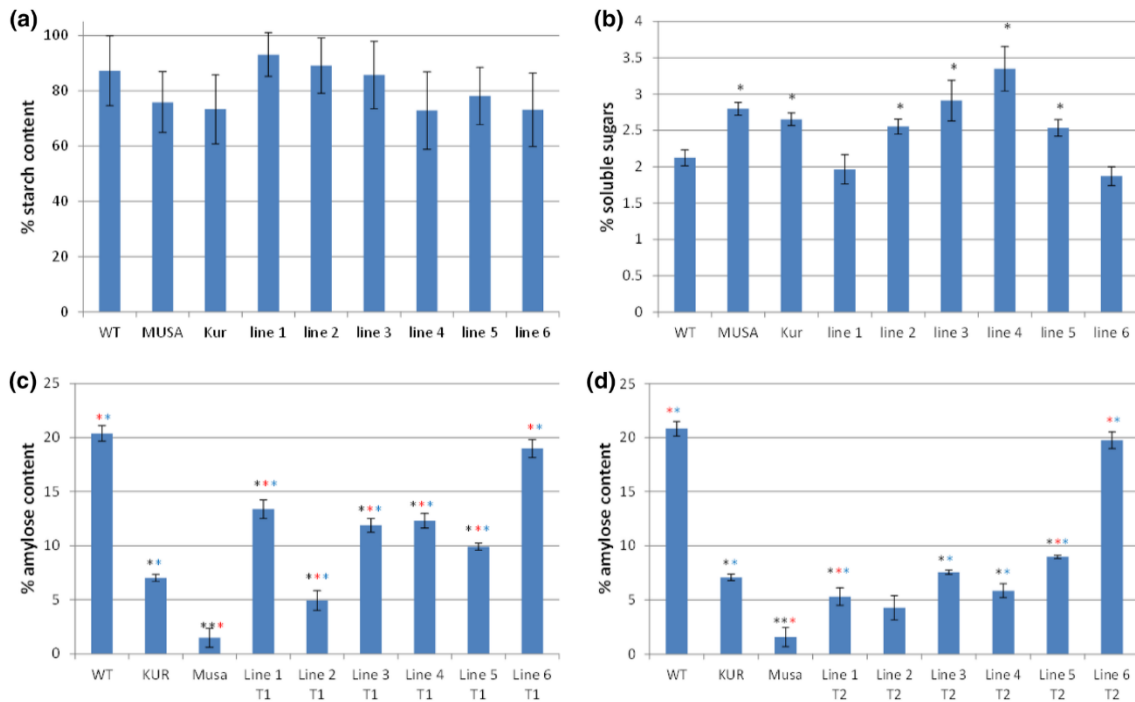
Asterisks indicate a statistically significant difference, as determined by t test (\*P < 0.05) on log-transformed data.



**Figure 11.** Mean normalized expression correlating different plant tissues with the different starch pathway gene isoforms depending on gene family (Gene) and the enzymatic function (Gene\_Type). Asterisks indicate a statistically significant difference, as determined by t test (\*P < 0.05) on log-transformed data.

### 4.3.5 Changes in starch, amylose and soluble sugar levels in GBSSI mutants

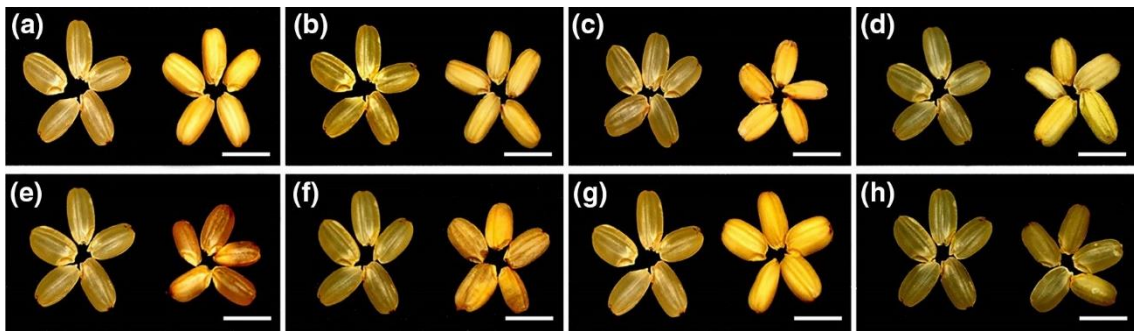
The starch content of the T2 seeds was similar to wild-type levels in all six mutant lines and in the KUR and Musa mutants (**Figure 12a**). In contrast, the soluble sugar content in T2 seeds was significantly (23-57%) higher than wild-type levels in all the mutant lines except lines 1 and 6, the latter 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 mutant 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.** Seed carbohydrate content in wild-type and mutant rice plants. (a) Total starch content of T2 seeds from wild-type (WT) plants, the *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  $\pm$  SD ( $n = 3$  biological replicates, 2 technical replicates for each biological replicate). Asterisks indicate a statistically significant difference, as determined by Student's t test (\* $P < 0.05$ ; \*\* $P < 0.01$ ). Asterisks in black indicate a statistically significant difference between wild-type and mutant lines. Asterisks in blue indicate a statistically difference between KUR and the other lines. Asterisks in red indicate a statistically significant difference between Musa and the other lines.

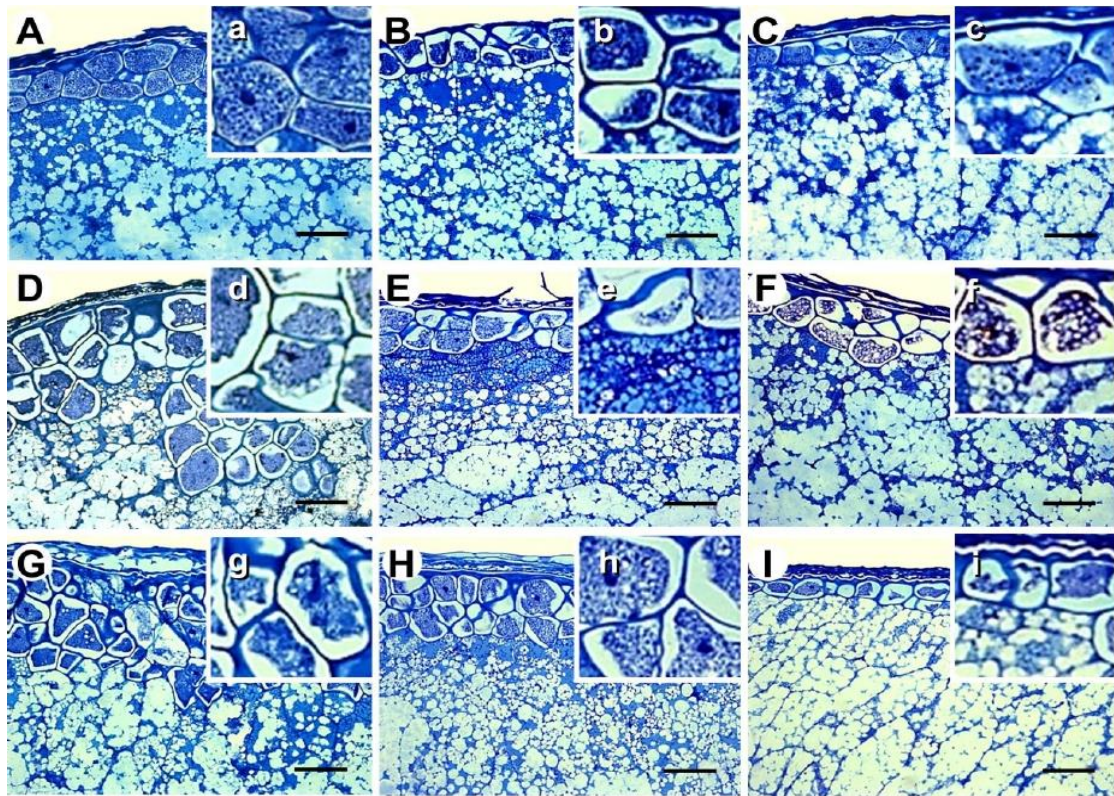
#### 4.3.6 Phenotype and microscopy changes produced by GBSSI mutations

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**).

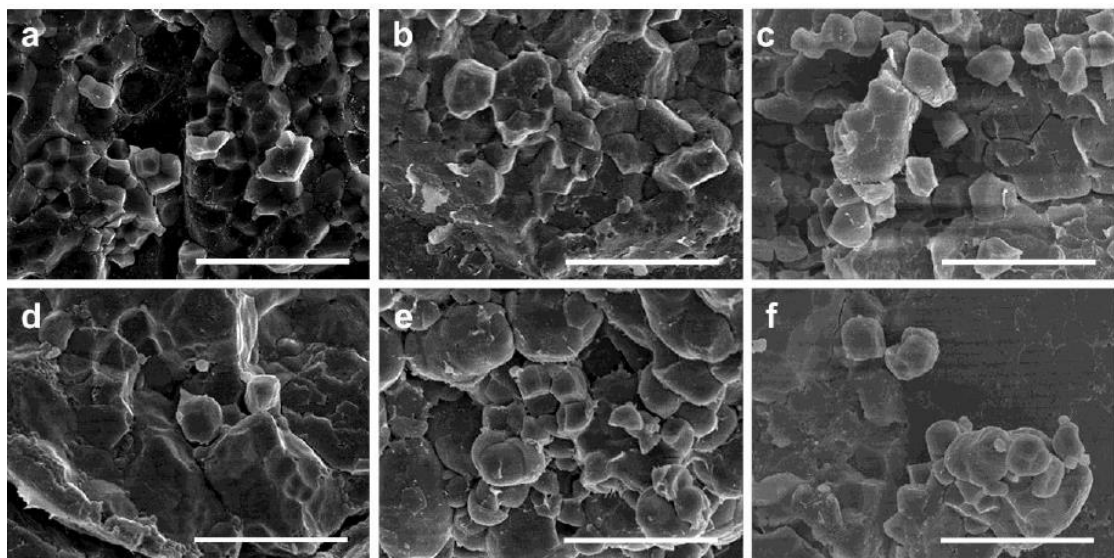


**Figure 13.** Seed phenotypes of wild-type (WT) plants, the *Wx* mutant lines and the two *Wx* irradiation mutants KUR and Musa. Scale bar, 5 mm. (a) Wild-type vs KUR. (b) Wild-type vs Musa. (c) Wild-type vs line 1. (d) Wild-type vs line 2. (e) Wild-type vs line 3. (f) Wild-type vs line 4. (g) Wild-type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





**Figure 14.** Optical microscopy 40x showing the structure of the aleurone layer in wild-type and mutant rice seeds. Scale bar is 50  $\mu\text{m}$ . a) WT, b) *OsWaxy* cultivar Musashimochi, c) *OsWaxy* cultivar KUR, d) line 1, e) line 2, f) line 3, g) line 4, h) line 5 and i) Line 6. Stained with Richardson Blue to observe the cell's structure and organization. Where cells of aleurone layer are disorganized, with a greater number of cells and rounder shape.



**Figure 15.** Scanning electron microscopy showing the structure of starch granules in wild-type and mutant rice seeds. Scale bar is 20  $\mu\text{m}$ . Wild-type (a) Musashimochi (b), mutated line 1 (c), mutated line 2 (d), mutated line 3 (e), and mutated line 5 (f). Where small changes in the cell structure (WT seeds showing sharp edges while CRISPR-waxy seeds were more irregular structure).

#### 4.4 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). In rice, there are two GBSS isoforms: GBSSI is mainly 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 mainly 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.

*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% compared to 20.8%, 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 *Wx* sites simultaneously. 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 (Baysal et al. 2016; Pérez et al. 2018). 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 2 was a homozygous mutant) and all six lines were homozygous mutants in the T2 generation.

The major consequence of mutating the *Wx* gene described earlier was the modification of the relative abundance of amylose and amylopectin without changing the overall starch content (Zhang et al. 2012, 2018). 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. Consistently with the homozygous GBSSI mutation in the T2 generation, 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 T2 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. Our data agree with earlier studies which reported a direct correlation between starch and soluble sugars (Preiss 1982). The phenotypes described above were concordant with the GBSS activity of the seeds. Our mutants and the two irradiation mutants showed only 42.4-69.16% of the GBSS activity compared to 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 *SBE1* result in grains with a chalky appearance (Ryoo et al. 2007; Zhang et al. 2011), mutations in *SBE1b* 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). 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. Changes in the length, width, thickness and volume of grains in starch pathway gene mutants were reported earlier (Wang et al. 2013; Tang et al. 2016). Statistical analysis for these parameters in grains of lines 1, 2 and 3 showed significant changes with respect to wild-type; lines 4 and 5 showed moderate changes which were still significant.

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 Musa mutant 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).

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 non-functional. 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  $\alpha$ -helices and  $\beta$ -sheets that form two substrate-binding clefts, one for ADP-glucose and other for malto-oligosaccharide 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. Line 6 was essentially identical to wild-type plants in all respects and this line was used as a second control. In wild-type T0 leaves, the major AGPases are *APL1*, *APL3* and *APS2a* and the major SEB is *SBE1* (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 *SBE1* 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 T0 leaves. Although *GBSSI* is low 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 T1/T2 seeds compared to T0 leaves. In T1/T2 seeds, *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 contain < 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) in T2 seeds, which provides an alternative pathway for starch synthesis in the absence of AGPase (Li et al. 2013). We found that the opposing changes in the expression levels of different AGPase subunits did not result 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). The ratio of SuSy/AGPase exhibited an inverse relationship with activity levels of the two enzymes as was hypothesized earlier (Li et al. 2013).

These collective results provide a basis to suggest that the GBSS activity levels in the mutant lines result from overexpression of GBSSII to compensate for the loss of GBSSI activity. GBSS activity in lines 2, 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.

#### 4.5 Conclusions

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 in the endosperm, 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 needed for import into starch granules is sufficient to block GBSSI activity without affecting the catalytic center in any other way. Our results provide critical mechanistic

insight into the complex feedback relationships among genes, enzymes, intermediates and end products in the starch biosynthesis and degradation pathways. This mechanistic understanding provides a basis for more targeted interventions to modulate starch biosynthesis in plants in a more precise manner to generate plants with altered starch content and composition for various applications.

#### 4.6 References

Ball SG, Morell MK (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 FJ, Li J, Bahaji A, Almagro G, Montero M, Etxeberria E, Hidalgo M, Sesma T, 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, 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 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, 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, Church GM (2015) Unraveling CRISPR-Cas9 genome engineering parameters via a library-onlibrary approach. *Nat Methods* 12:823

Christou P, Ford TL, 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

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 NF, Matsushima R, Hayashi M, Tetlow IJ, Emes MJ, Nakamura Y, 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 JF, Guindon S, Lefort V, Claverie JM, 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, Zhu C (2012) Transgenic rice grains expressing a heterologous  $\rho$ -hydroxyphenylpyruvate dioxygenase shift tocopherol synthesis from the  $\gamma$  to the  $\alpha$  isoform without increasing absolute tocopherol levels. *Transgenic Res* 21:1093–1097
- Fujita N, Yoshida M, Asakura N, Ohdan T, Miyao A, Hirochika H, Nakamura Y (2006) Function and characterization of starch synthase I using mutants in rice. *Plant Physiol* 140:1070–1084
- Heigwer F, Kerr G, Boutros M (2014) E-CRISP: fast CRISPR target site identification. *Nat Methods* 11:122 Hirano HY (1993) Genetic variation and gene regulation at the wx locus in rice. *Gamma Field Symp* 24:63–79
- Hirano HY, 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 Itoh K, Nakajima M, 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, 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 JS, Ryou N, Hahn TR, Walia H, Nakamura Y (2010) Starch biosynthesis in cereal endosperm. *Plant Physiol Biochem* 48:383–392
- Jiang D, Cao WX, Dai TB, 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, Weeks DP (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 BO (1971) A simplified assay for milled-rice amylose. *Cereal Sci Today* 16:334–336
- Kang TJ, Yang MS (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, Nakamura Y (2005) Roles of isoamylase and ADP-glucose pyrophosphorylase in starch granule synthesis in rice endosperm. *Plant J* 42:164–174
- Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protov* 10:845
- Li J, Baroja-Fernández E, Bahaji A, Muñoz FJ, Ovecka M, Montero M (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



- Li C, Powell PO, Gilbert RG (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, 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, Liu YG (2015) A robust CRISPR/Cas9 system for convenient, high-efficiency multiplex genome editing in monocot and dicot plants. *Mol Plant* 8:1274–1284
- Maddelein ML, Libessart N, Bellanger F, Delrue B, D’Hulst C, Van den Koornhuyse N, Fontaine T, Wieruszkeski JM, Decq A, Ball S (1994) Toward an understanding of the biogenesis of the starch granule. Determination of granule-bound and soluble starch synthase functions in amylopectin synthesis. *J Biol Chem* 269:25150–25157
- Martin C, Smith AM (1995) Starch biosynthesis. *Plant Cell* 7:971–985
- Momma M, 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 SY, Ohya T (1989) Carbohydrate metabolism in the developing endosperm of rice grains. *Plant Cell Physiol* 30:833–839
- Ohdan T, Francisco Jr PB, 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
- 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 Res* 27:423–439
- Preiss J (1982) Regulation of the biosynthesis and degradation of starch. *Annu Rev Plant Physiol* 33:431–454
- 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
- Ryoo N, Yu C, Park CS, Baik MY, Park IM, Cho MH, Bhoo SH, An G, Hahn TR, Jeon JS (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

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

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

Seung D, Soyk S, Coiro M, Maier BA, Eicke S, Zeeman SC (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, Gao C (2014) Genome editing in rice and wheat using the CRISPR/Cas system. *Nature Protoc* 9:2395

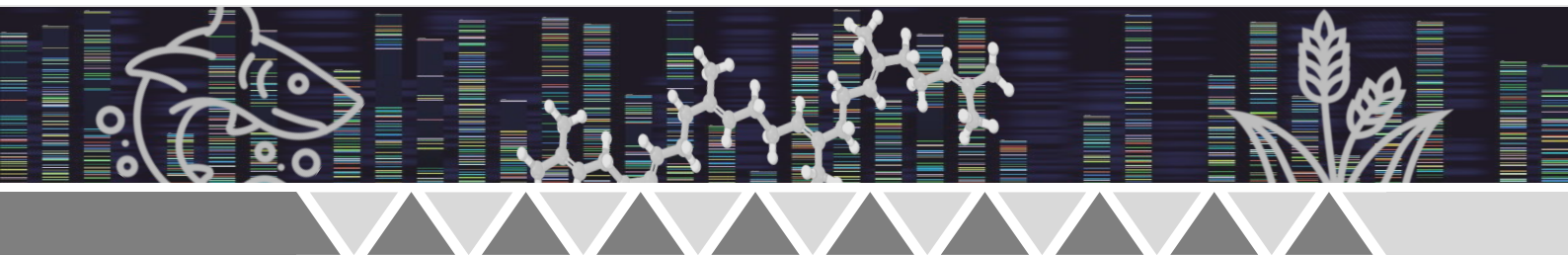
Sudhakar D, Bong BB, Tinjuangjun P, Maqbool SB, Valdez M, Jefferson R, Christou P (1998) An efficient rice transformation system utilizing mature seed-derived explants and a portable, inexpensive particle bombardment device. *Transgenic Res* 7:289–294

Sun Y, Jiao G, Liu Z, Zhang X, Li J, Guo X, Du J, Francis F, Zhao Y, Xia L (2017) Generation of high-amylose rice through CRISPR/Cas9-mediated targeted mutagenesis of starch branching enzymes. *Front Plant Sci* 8:298





# General Discussion





## General discussion

For thousands of years agriculture has been a fundamental part of civilizations and the genetic characteristics of plants and animals have been modified by humans, carried out selective breeding to obtain desired characteristics and eliminate or reduce the non-favorable ones. The elucidation of the 3D-structure of DNA by James Watson, Francis Crick and Rosalind Frankling in 1953 and its recognition as the molecule that carried the heritable information, opened the doors to the development of powerful technologies for manipulate the genetic code in an increasingly precise way, altering both genotype and phenotype of organisms from one generation to another (Watson and Crick 1953; Franklin and Gosling, 1953).

Over the past decades, scientists have been bringing together the approaches on engineering and biology to increase the capability to manipulate genes. With the development of synthetic techniques, using bioinformatics tools and simple chemical compounds it was possible to design genes and reconstruct DNA pieces *de Novo* (Bio fab group 2006; Brown et al., 2015; Deplazes 2009).

In this research, different metabolic engineering approaches had been explored to produce squalene in *E. coli*, the most widely microbial chassis used as experimental model in molecular biology. On the other hand, a CRISPR/Cas9 genome editing tool has been used in rice (*Oryza sativa*), one of the most important food crops in the world, to knock-out two key genes that are involved in starch metabolism.

Three different strategies have been used for the production of squalene in *E. coli*. In the first strategy genes from plants were inserted, in the second strategy native *E. coli* genes were overexpressed and in the third strategy an *E. coli* strain containing two metabolic pathways (MEP/MVA) was used. Of those three strategies, the first one resulted in the highest production of squalene and the one that imposed the least metabolic load on the bacteria. The combination of genetic engineering strategies and adjustments to the microbial cell culture conditions could help alleviate bacterial stress due to metabolic load and thereby improve the squalene production of strategies two and three.

In conclusion, a squalene-producing *E. coli* strain has been engineered with genes from rice (*Oryza sativa*) and *Gentiana lutea* (approach one), creating a new synthetic microorganism, reported for the first time in this thesis. With this strain an instant productivity of squalene of 3.8 mg/L/h was achieved, being the third-highest reported so far in the open literature.

In order to know the function of the *APL2* gene in the starch metabolism in rice, the CRISPR/Cas9 genome editing tool has been used to achieve a knockout of this gene and observe how it affects the plant fertility, since starch is accumulated in the grain endosperm as an energy reserve for germination. In effect, it was determined that mutations in the *APL2* gene resulted in the loss of function of the enzyme AGPase and

consequently, the fertility of the plant was severely affected due to decrease of the starch content in rice grains.

In the same way, CRISPR/Cas9 induced mutations has been introduced in *Waxy* gene encoding GBSSI resulted in a partial loss of the activity of GBSS enzyme. In consequence, it was observed a loss of amylose and an accumulation of amylopectin in the endosperm. Unexpectedly, the expression of other downstream genes involved in the starch metabolism were observed.

The effects of the CRISPR/Cas9 induced mutations on the function of *APL2* and *Waxy* genes involved in starch metabolism in rice were tested. Genome editing offers the opportunity of create mutations that allow reprogramming the metabolism machinery in plants. These findings contribute to a better understanding of starch biosynthesis in rice and provide useful evidence of the effects of genome editing to create rice varieties with better nutritional properties in the near future.

With this research and findings, it is expected to increase the knowledge in biotechnology using metabolic engineering and synthetic biology tools, as they are key to facing the challenges posed by the Sustainable Development Goals (SDG). Responsible consumption and production (SDG 12), life under water (SDG 14), and zero hunger (SDG 2) can be addressed by the creation of alternative sources for sustainable squalene production to avoid shark's uncontrolled slaughter and improving rice crops by genome editing.

## References

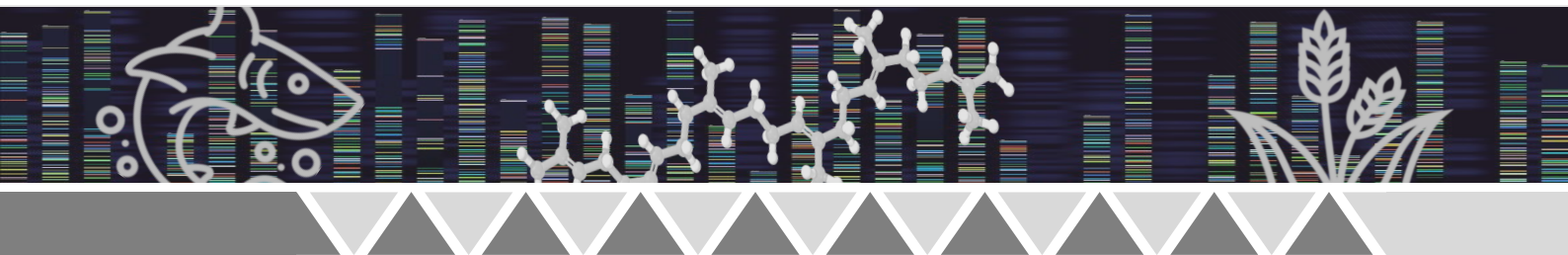
- Bio FAB Group. (2006). Engineering life: building a fab for biology. *Scientific American*, 294, 44-51.
- Brown, S., Clastre, M., Courdavault, V., & O'Connor, S. E. (2015). De novo production of the plant-derived alkaloid strictosidine in yeast. *Proceedings of the National Academy of Sciences*, 112, 3205-3210.
- Deplazes, A. (2009). Piecing together a puzzle: An exposition of synthetic biology. *EMBO reports*, 10, 428-432.
- Franklin, R. E., & Gosling, R. G. (1953). Molecular configuration in sodium thymonucleate. *Nature*, 171, 740-741.
- Watson, J. D., & Crick, F. H. (1953). Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature*, 171, 737-738.







# General Conclusions





## General conclusions

With respect to objective 1 and 2 the main conclusions are:

1. Three squalene-producing *E. coli* strains were designed under three different approaches, between them with strategy 1, the highest instant productivity of squalene (3.8 mg/L/h) was reached.
2. This work for the first time reports metabolic engineering of genes from rice plants and *Gentiana lutea* (strategy 1) for the accumulation of squalene in a genetically engineered *E. coli* strain.
3. The productivities of strategies two and three could be improved by adjusting the culture conditions and alleviating the metabolic burden due to the number of genes inserted in plasmids.

With respect to objective 3 the main conclusions are:

4. Mutating one allele of *OsAPL2*, resulted in the unexpected expression of both *OsAPL2* and *OsAPS2b* in the leaves, the latter encoding the only cytosolic small subunit.
5. 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.
6. 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.

With respect to objective 4 was concluded:

7. 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 in the endosperm, but also caused the unexpected expression of other downstream starch pathway genes, partly to deal with abundant intermediates and unusual starch structures.
8. The increase in GBSSII expression did not compensate for the loss of GBSSI, reflected in the amylose content of the mutant lines.
9. The mechanistic understanding of the complex feedback relationships among genes, enzymes, intermediates and end products in the starch biosynthesis and degradation pathways, provides a basis for more targeted interventions to modulate starch biosynthesis in plants.

