Identification and characterisation of chloroplast 3C and 4C monocarboxylate and dicarboxylate transporters involved in the C_4 cycle



Oliver Mattinson

St. Hugh's College

Department of Biology

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Hilary 2023

Abstract

C_4 photosynthesis is a highly efficient form of photosynthesis that utilises a biochemical pump to concentrate CO_2 around rubisco. This biochemical pump is dependent on multiple transport steps to move metabolites between organelles and between cells. However, the transporters that carry out several of these transport steps have evaded identification or characterisation. This lack of molecular characterisation is a major gap in our understanding of terrestrial photosynthetic biology, and also represents a significant barrier to attempts to engineer C_4 photosynthesis into C_3 crop species to improve yield. This thesis aims to address this problem and identify and characterise some of the missing transporters of the C_4 pathway. The first half of this thesis is focussed on the characterisation of the HATCH1 transporter, demonstrating its function in *E. coli* and assessing the impact of the loss of this transporter (and its paralog) on C_4 photosynthesis in *Setaria viridis*. Ultimately, I demonstrate that HATCH1 is involved in, but not necessary for, C_4 photosynthesis and likely acts redundantly with other unknown transporter(s). The second half of this thesis concerns the identification and molecular characterisation of putative metabolite transporters of the C_4 cycle in *Zea mays*. Whilst HATCH1 and HATCH3.1 were identified as candidate C_4 cycle transporters, *in planta* data does not support this. *In vitro* and *in planta* data suggest a potential role for another candidate transporter in transport of C_4 cycle metabolites at the chloroplast envelope of the bundle sheath cell. Together the results from this thesis increase our understanding of intracellular metabolite transport in C_4 plants, and bring us closer to being able to engineer this trait into C_3 plants.

Acknowledgements

Doing a PhD has been a challenging experience, particularly when interrupted by a pandemic! I'm sure that I wouldn't have made it through to the end without the help and support of many people.

Firstly, thanks to everyone in the Kelly and Langdale labs who helped me throughout the process and made the lab a pleasant place to work. In particular I'd like to thank Michael Niklaus for showing me the ropes in the lab and getting the work on HATCH1 started, Dana Vlad for lots of help with molecular cloning and *Setaria* transformation, David Emms for bioinformatic help, Florian Hahn for invaluable CRISPR advice, Avishek Dey for amazing collaboration (and late nights!) developing bundle sheath strand isolation protocols, Julie and Lizzie for making sure the lab runs smoothly, VJ Ambekar for assisting with plant maintenance, and Pippa Sinclair for going through the whole process and experiencing the highs and lows at the same time as me! Of course, special thanks go to my supervisor, Steve Kelly, whose expertise led to the initial conception of the project, and whose advice and support has been crucial throughout. It has been a privilege to chip away at the frontier of knowledge together.

It has also been invaluable to me to have the help of several people outside of the lab: in particular my parents, who have always provided such a solid base for me, as well as growing quite adept at dealing with their perpetually stressed-out son. Finally, I'd like to give special thanks to my wonderful fiancé Alice, who had to deal with me on a daily basis, and somehow managed to complete her PhD at the same time. I couldn't have done it without you.

Table of contents

Abstract.....i

Acknowledgements.....ii

Chapter 1: Introduction.....1

1.1 C_3 photosynthesis, photorespiration, and carbon-concentrating mechanisms.....1

1.2 C_4 photosynthesis is highly dependent on intracellular metabolite transport.....8

1.3 A review of metabolite transporters involved in C_4 photosynthesis.....11

1.3.1 Chloroplastic dicarboxylate transport – the DCT family.....11

1.3.2 Chloroplastic monocarboxylate/phosphate transport – the TPT family.....18

1.3.3 Chloroplastic pyruvate transport – the BASS family.....28

1.3.4 Mitochondrial pyruvate transport – the MPC family.....34

1.3.5 Mitochondrial dicarboxylate transport – the MC family.....37

1.4 Current gaps in our knowledge of NADP-ME subtype C_4 transporters.....49

1.4.1 H^+-dependent mesophyll chloroplast pyruvate import.....50

1.4.2 Bundle sheath chloroplast malate import and pyruvate export.....50

Chapter 2: Characterisation of HATCH1, a bundle sheath chloroplast malate importer and pyruvate exporter in *Setaria*.....54

2.1 Chapter introduction and background.....54

2.1.1 Missing transport steps of the C_4 cycle in *S. viridis*.....54

2.1.2 HATCH1 is a putative bundle sheath-localised C_4 cycle transporter in *S. viridis*.....54

2.1.3 A bacterial system to characterise chloroplast-localised malate and pyruvate transporters.....57

2.1.4 HATCH1 has putative malate import and pyruvate export activity.....59

2.1.5 RNAi-mediated HATCH1 knock-down *S. viridis* show photosynthetic defects.....62

2.1.6 Chapter aims and outline.....64

2.2 Results and discussion.....64

2.2.1 HATCH1 is localised to the chloroplast inner envelope.....64

2.2.2 HATCH1 facilitates selective dicarboxylate import.....68

2.2.3 HATCH1 facilitates pyruvate export in the absence of external malate.....71

2.2.4 Computational prediction of HATCH1 complex structure.....75

2.2.5 HATCH1 carries out malate/pyruvate transport *in planta*.....83

2.2.6 Flux balance analysis modelling of HATCH1 knock-down plants recapitulates observed gas-exchange phenotype.....88

2.3 Chapter conclusion.....91

2.4 Materials and methods.....94

2.4.1 Generation of GFP fusion constructs for expression in *N. benthamiana*.....94

2.4.2 Transient protein over-expression in *N. benthamiana* leaves.....95

2.4.3 Confocal microscopy analysis of transiently transformed *N. benthamiana* leaves.....95

2.4.4 Generation of *S. italica* HATCH1 construct for expression in *E. coli*.....96

2.4.5 Experimental evaluation of dicarboxylate uptake specificity of HATCH1 in *ΔdctA E. coli*.....96

2.4.6 Expression of HATCH1 in IMW213 *E. coli* on minimal media containing glucose as a sole carbon source.....97

2.4.7 Pyruvate measurement in cell supernatant.....98

2.4.8 Whole genome sequencing and read mapping of AN387 and IMW213 *E. coli*.....98

2.4.9 YqhA and HATCH1 structure prediction using AlphaFold2.....99

2.4.10 *S. viridis* plant growth conditions.....99

2.4.11 *S. viridis* bundle sheath strand isolation.....100

2.4.12 *S. viridis* bundle sheath chloroplast isolation.....101

2.4.13 *S. viridis* bundle sheath strand transport assay.....101

2.4.14 Chlorophyll quantification.....102

2.4.15 Flux balance analysis modelling.....103

Chapter 3: Investigation into the role of HATCH1 in C_4 photosynthesis in *Setaria viridis*.....106

3.1 Chapter introduction and background.....106

3.1.1 Why did knockdown of a transporter of the C_4 cycle not produce a growth effect?.....106

3.1.2 Hypothesis 2: HATCH1 RNAi may be compensated for by the expression of a HATCH1 paralog in *S. viridis*.....107

3.1.3 Hypothesis 3: HATCH1 RNAi may be compensated for by the operation of auxiliary C_4 cycles in *S. viridis*.....108

3.1.4 The use of CRISPR/Cas9 to generate gene knock-outs in plants.....111

3.1.5 Chapter aims and outline.....113

3.2 Results and discussion.....114

3.2.1 Generation of *hatch1* single- and *hatch1,hatch3* double-knock-out lines in

S. viridis using CRISPR/Cas9.....114

3.2.2 *hatch1* single- and *hatch1,hatch3* double-knock-out lines do not show significantly different dry matter carbon isotope discrimination or carbon/nitrogen ratios.....118

3.2.3 *hatch1* single- and *hatch1,hatch3* double-knock-out lines show inhibited growth compared to wild-type.....119

3.2.4 *hatch1* single- and *hatch1,hatch3* double-knock-out lines show reduced CO_2 assimilation.....124

3.2.5 Transcriptomic analysis of *hatch1* single- and *hatch1,hatch3* double-knock-out lines reveals compensatory transcriptome changes but no evidence for upregulation of supplementary C_4 cycles.....134

3.3 Chapter conclusion.....146

3.4 Materials and methods.....150

3.4.1 *S. viridis* growth conditions.....150

3.4.2 Generating CRISPR/Cas9 constructs.....150

3.4.3 Transformation of *S. viridis* ME034V callus and plant regeneration.....152

3.4.4 Plant DNA extraction.....154

3.4.5 Genotyping and sequencing of transgenic plants.....154

3.4.6 Analysis of isotopic composition of *S. viridis*.....155

3.4.7 *S. viridis* growth measurements.....155

3.4.8 *S. viridis* photosynthetic gas-exchange and fluorescence measurements.....156

3.4.9 Photosynthetic parameter estimation from A-C_i data.....156

3.4.10 Total RNA extraction from *S. viridis* leaves.....158

3.4.11 RNA-seq data analysis.....159

Chapter 4: A novel screen for C_4 cycle transporters in *Zea mays*.....161

4.1 Chapter introduction and background.....161

4.1.1 Transporters missing from our understanding of the C_4 cycle in *Z. mays*..161

4.1.2 Using bacteria to screen for metabolite transport activity.....162

4.1.3 Pyruvate uptake systems in *E. coli*.....164

4.1.4 Chapter aims and outline.....168

4.2 Results and discussion.....169

4.2.1 A bioinformatic screen identifies candidate C_4 cycle transporters in *Z. mays*.....169

4.2.2 A bacterial screen for malate import reveals no candidates.....176

4.2.3 Generation of an *E. coli* pyruvate uptake null mutant.....178

4.2.4 A bacterial screen for pyruvate import reveals no candidates.....185

4.2.5 A bacterial screen for metabolite export reveals some candidates.....188

4.2.6 Metabolite export profiles of promising candidates.....191

4.2.7 Generation and preliminary analysis of maize *mep3* mutant lines.....194

4.3 Chapter conclusion.....197

4.4 Materials and methods.....201

4.4.1 A bioinformatic screen to identify candidate C_4 cycle transporters in *Z. mays*.....201

4.4.2 Construction of inducible vectors for over-expression of candidate proteins in *E. coli*.....202

4.4.3 Construction of inducible vectors for over-expression of *E. coli* transporters In *E. coli*.....203

4.4.4 Generation of an *E. coli* quintuple mutant using recombineering.....204

4.4.5 Whole genome sequencing and read mapping of quintuple mutant *E. coli*.....208

4.4.6 Expression of candidate proteins in *E. coli* grown on M9 minimal media and growth curve measurements.....208

4.4.7 Growth curve fitting.....209

4.4.8 Generation of knock-out lines in *Z. mays* using CRISPR/Cas9 by VIB.....210

Chapter 5: An investigation into the *HATCH* family across plants, with particular focus on *Zea mays*.....211

5.1 Chapter introduction and background.....211

5.1.1 HATCH family proteins may play a role in the C_4 cycle in maize.....211

5.1.2 Chapter aims and outline.....214

5.2 Results and discussion.....215

5.2.1 The *HATCH* gene family in plants.....215

5.2.2 Transport properties of representative HATCH family proteins exogenously expressed in *E. coli*.....218

5.2.3 HATCH3 is localised to the chloroplast inner envelope.....223

5.2.4 Generation of *hatch3.1* knock-out lines in *Z. mays* uing CRISPR/Cas9.....224

5.2.5 *hatch3.1* knock-out *Z. mays* do not show inhibited growth compared to

wild-type.....225

5.2.6 *hatch3.1* knock-out *Z. mays* do not show altered CO_2 assimilation compared to wild-type.....228

5.2.7 *hatch3.1* knock-out *Z. mays* do not show altered photosynthetic induction compared to wild-type.....231

5.3 Chapter conclusion.....233

5.4 Materials and methods.....235

5.4.1 Phylogenetic tree inference.....235

5.4.2 Phylogenetic tree branch length calculation.....236

5.4.3 Prediction of protein structural features.....236

5.4.4 Generation of GFP fusion constructs for expression in *N. benthamiana*...236

5.4.5 Transient protein over-expression in *N. benthamiana* leaves.....237

5.4.6 Confocal microscopy analysis of transiently transformed *N. benthamiana* leaves.....237

5.4.7 Generation of knock-out lines in *Z. mays* using CRISPR/Cas9 by VIB.....237

5.4.8 *Z. mays* growth conditions.....238

5.4.9 Plant DNA extraction.....238

5.4.10 PCR genotyping of transgenic plants.....239

5.4.11 *Z. mays* growth measurements.....239

5.4.12 *Z. mays* photosynthetic gas-exchange and fluorescence measurements.....239

Chapter 6: Discussion.....241

6.1.1 Characterisation of a C_4 cycle malate/pyruvate transporter in *Setaria*.....241

6.1.2 Identification of further candidate C_4 cycle transporters in maize.....245

6.1.3 Potential biotechnological applications of the expression of HATCH transporter proteins in *E. coli*.....247

6.1.4 Increasing our understanding of the C_4 cycle and potential biotechnological Applications.....248

Appendix.....253

A.1 Abbreviations.....253

A.2 Media.....257

A.3 Common protocols.....263

A.4 Golden Gate cloning of constructs for expression in *E. coli*.....269

A.5 Golden Gate cloning of GFP fusion constructs for expression in *Nicotiana benthamiana*.....293

A.6 *Setaria viridis* genotyping and sequencing.....302

A.7 Golden Gate cloning of CRISPR/Cas9 constructs.....303

A.8 *Setaria viridis* genotyping and sequencing.....309

A.9 *Setaria viridis* transcriptomics.....311

A.10 *E. coli* pyruvate transporter knock-out generation.....340

A.11 *Zea mays mep3* knock-out.....346

A.12 *Zea mays* genotyping.....347

A.13 *Zea mays mpa3.1* knock-out.....347

Bibliography.....348

Chapter 1: Introduction

1.1 C_3 photosynthesis, photorespiration, and carbon-concentrating mechanisms

Photosynthesis is the process by which plants utilize solar energy to fix atmospheric CO_2 into organic carbon compounds, providing the entry point for most carbon and energy into the biosphere. Most plants carry out a form of photosynthesis known as C_3 photosynthesis, in which CO_2 diffuses directly into the chloroplasts of mesophyll cells in the leaf, and is fixed into the 3-carbon compound 3-phosphoglycerate (3-PGA) by the enzyme ribulose-1,5-bis-phosphate carboxylase/oxygenase (rubisco) (Figure 1.1 A). However, the efficiency of C_3 photosynthesis is reduced by a competing reaction of rubisco with O_2, which produces 2-phosphoglycolate (2-PG), a toxic waste product that inhibits enzymes of the Calvin Benson Bassham Cycle and starch biosynthesis. To alleviate this inhibition, 2-PG is exported from the chloroplast and is recycled back to 3-PGA in a process known as photorespiration. This process also leads to the release of one molecule each of CO_2 and NH_4^+ that must then be re-assimilated or lost from the plant (Figure 1.1 B). Whilst the affinity of rubisco for O_2 is much lower than for CO_2, atmospheric concentration of O_2 is 525 times higher than that of CO_2, with the net result being the ratio for carboxylation:oxygenation in C_3 plants is approximately 3:1[1]. Thus, under normal atmospheric conditions, the direct energetic cost of photorespiration, as well as the indirect energetic costs of the re-assimilation of released CO_2 and NH_4^+, are predicted to result in approximately 50 % higher energy costs for the light-independent reactions of photosynthesis[2].

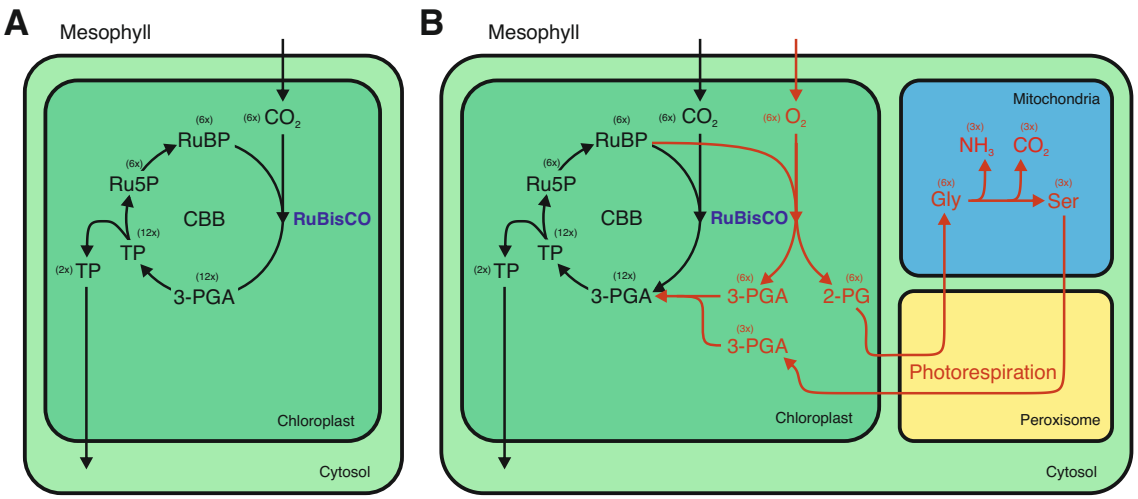


Figure 1.1: C_3 photosynthesis and photorespiration. Simplified diagrams showing **A**) C_3 photosynthesis, and **B**) C_3 photosynthesis in addition to photorespiration. Calvin cycle metabolites are shown in black. Photorespiratory metabolites are shown in red. The enzyme rubisco is shown in blue. Brackets show relative stoichiometries of each metabolite. For abbreviations see appendix A.1, table A.1.

Under many conditions this can cause a significant negative impact on plant growth and, in the case of C_3 crops, significantly reduced yield[3]. Even though anthropogenic CO_2 emissions might be expected to reduce rates of photorespiration in the future by increasing the carboxylation:oxygenation ratio of rubisco, the increase in average global temperature and aridity resulting from higher atmospheric CO_2 concentrations are predicted to have the opposite effect[4]. Thus, photorespiration is likely to cause significant energy and yield losses to plants even as atmospheric CO_2 levels rise.

In order to mitigate the cost of photorespiration, several plant lineages have evolved biochemical and anatomical adaptations that facilitate an increased concentration of CO_2 compared to O_2 around rubisco[5–8]. These lineages can be categorised into two

different photosynthetic types: CAM (Crassulacean Acid Metabolism) and C_4. Both types function by the same principle of initially fixing CO_2 by a different enzyme, phospho*enol*pyruvate carboxylase (PEPC), which is more selective for CO_2 than rubisco is[9], and then subsequently releasing and re-assimilating that CO_2 using rubisco. In both types, rubisco is kept in a high CO_2 environment in which it is protected from oxygen, and therefore the oxygenation reaction, and photorespiration, are significantly reduced.

In CAM plants, the separation of initial CO_2 fixation by PEPC and secondary fixation by rubisco is achieved temporally[10]. CO_2 is taken up through open stomata during the night, and is fixed by PEPC into 4-carbon organic acids, which are stored in the vacuole. During the day, these 4-carbon organic acids are released from the vacuole and are decarboxylated to produce CO_2 for secondary fixation by rubisco (Figure 1.2 A). Whilst CAM plants typically do not have fast photosynthetic rates[11], their inverse stomatal behaviour facilitates significantly reduced evapotranspiration. This allows for improved water-use efficiency (WUE), which confers an advantage for CAM plants in hot and dry environments.

In contrast to the temporal separation of CO_2 fixation in CAM plants, in C_4 plants, the separation of initial CO_2 fixation by PEPC and secondary fixation by rubisco is achieved spatially. Whilst some plants carry out C_4 photosynthesis within the context of a single cell[12], in most cases the C_4 pathway is split between two specialised cell types, known as the mesophyll and bundle sheath cells. A specialised leaf anatomy, known as Kranz anatomy, facilitates the ~1:1 ratio of mesophyll to bundle sheath cells that is required for optimal operation of the C_4 cycle. Unlike in C_3 plants, the mesophyll cells of C_4 plants

have high levels of PEPC expression, but do not express rubisco, which is instead found in the bundle sheath. Initial CO_2 fixation by PEPC into 4-carbon organic acids occurs in the mesophyll cells. The four-carbon organic acids then diffuse from the mesophyll to the bundle sheath through the plasmodesmata, where they are decarboxylated to produce CO_2 for secondary fixation by rubisco (Figure 1.2 B). A combination of a suberized bundle sheath cell wall (in some species) and cytosolic distance creates diffusive resistance between the site of initial carboxylation in the mesophyll, and decarboxylation and re-fixation in the bundle sheath[13], thus facilitating higher CO_2 and lower O_2 concentrations around rubisco. Compared to C_3 plants, C_4 plants therefore have decreased rates of photorespiration, increased photosynthetic efficiency, and increased maximal photosynthetic rates[14]. This is further associated with enhanced nitrogen-use efficiency (NUE) (as less nitrogen needs to be invested in rubisco to achieve the same rate of carbon fixation)[15] and WUE (as the carbon-concentrating effect allows photosynthetic rate to remain high despite low intercellular CO_2 concentration caused by stomatal closure)[16].

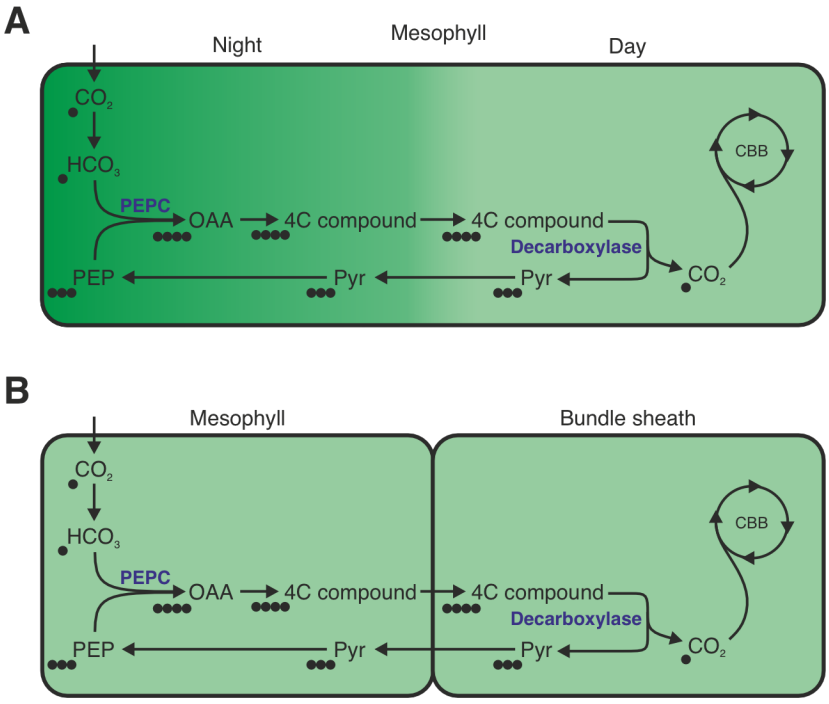


Figure 1.2: Simplified cartoon of the two most abundant carbon concentrating mechanisms in plants. Simplified diagrams showing **A**) CAM photosynthesis, and **B**) C_4 photosynthesis. Other, less common carbon-concentrating mechanisms found in plants, such as single-cell C_4 cycles and hornwort pyrenoids are not shown. In C_4 photosynthesis the 4-carbon compound may be malate or aspartate, whilst the 3-carbon compound is usually pyruvate but may be converted to alanine for transfer between cells. Metabolites are shown in black. Black circles indicate number of carbon atoms per metabolite. Enzymes are shown in blue. For abbreviations see appendix A.1, table A.1.

C_4 photosynthesis has evolved at least 61 times independently during the evolution of land plants, with the highest density of origins occurring within the grasses[8]. It is not surprising, given the large number of independent evolutionary origins, that the trait encompasses significant morphological and biochemical variation between species. Biochemically, C_4 plants have traditionally been divided into three subtypes, based on the decarboxylation enzyme with the highest activity in the leaf tissue. These are

chloroplastic NADP-malic enzyme (NADP-ME), mitochondrial NAD-ME, or cytosolic phospho*enol*pyruvate carboxykinase (PCK)[17]. All 3 subtypes share similar metabolic adaptations in the mesophyll cell. Pyruvate entering from the bundle sheath cell is converted to phospho*enol*pyruvate (PEP) by chloroplastic pyruvate, phosphate dikinase (PPDK). Carbonic anhydrase converts CO_2 into HCO_3^-, and PEPC then generates oxaloacetate (OAA) from HCO_3^- and PEP. OAA is short-lived and spontaneously decomposes back to pyruvate and CO_2. Therefore, to stabilise the newly synthesised 4-carbon molecule before it can diffuse back to the bundle sheath, OAA is imported into the mesophyll chloroplast where it is converted to either malate (by NADP-malate dehydrogenase (NADP-MDH)), or aspartate (by aspartate aminotransferase (AspAT)), depending on the C_4 subtype. If it is stabilised as malate, the malate is exported from the mesophyll cell chloroplast and diffuses to the bundle sheath cell where it is imported into the bundle sheath cell chloroplast and decarboxylated to produce pyruvate and CO_2 by chloroplastic NADP-ME (Figure 1.3 A). If OAA is stabilised as aspartate, then the aspartate is exported from the mesophyll cell chloroplast and diffuses to the bundle sheath cell. Upon arrival in the bundle sheath cell there are two potential routes depending on the primary decarboxylase that is used. In NAD-ME type C_4 photosynthesis the aspartate is imported into the bundle sheath cell mitochondrion, converted back to OAA (by mitochondrial AspAT), re-stabilised as malate (by mitochondrial NAD-MDH) and decarboxylated to produce pyruvate and CO_2 by mitochondrial NAD-ME (Figure 1.3 B). In PCK type C_4 photosynthesis, the aspartate is converted to OAA (by cytosolic AspAT) and then directly decarboxylated to produce PEP and CO_2 by cytosolic PCK (Figure 1.3 C).

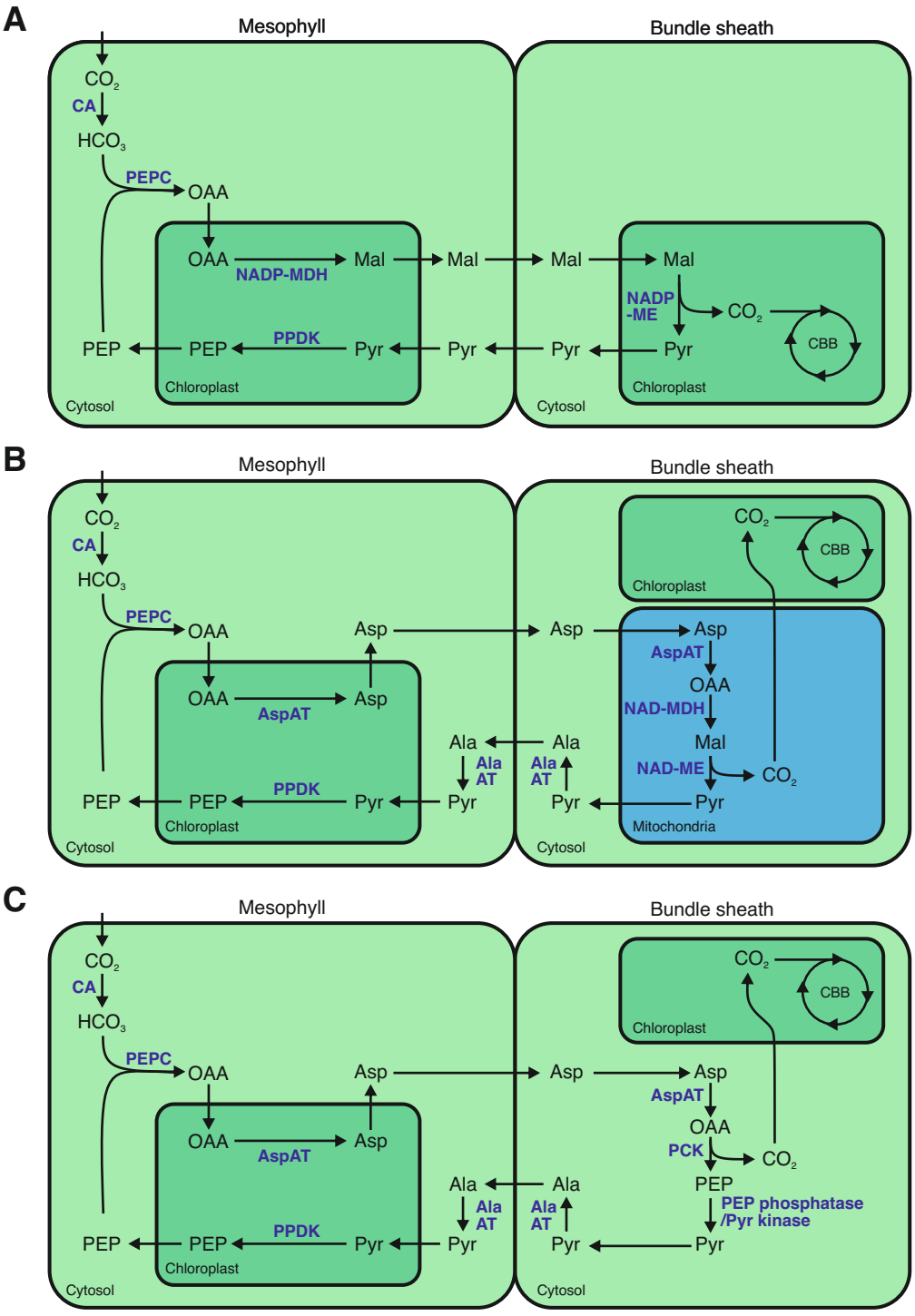


Figure 1.3: C_4 photosynthesis subtypes. Diagrams showing **A**) NADP-ME subtype, **B**) NAD-ME subtype, and **C**) PCK subtype C_4 photosynthesis. Metabolites are shown in black. Enzymes are shown in blue. For abbreviations see appendix A.1, table A.1.

Although three separate subtypes (NADP-ME, NAD-ME and PCK) have been described, a strict division into distinct subtypes is now considered to be an over-simplification, with many C_4 plants possessing mixed C_4 cycles utilising more than one decarboxylase[18,19]. For example, *Zea mays* (maize) has traditionally been classified as an NADP-ME subtype C_4 plant, but has now been shown to carry out an auxiliary PCK pathway[20]. Such mixed cycles may prove advantageous by reducing the diffusion gradient requirement for any one transfer acid in order to sustain metabolite movement between cells, as well as allowing for finer tuning of energy balancing between mesophyll and bundle sheath cells[19].

1.2 C_4 photosynthesis is highly dependent on intracellular metabolite transport

Chloroplasts and mitochondria, reflecting their bacterial ancestry, are double membrane-bound organelles[21]. In both cases the outer membrane contains pores that allow for the non-selective uptake of low molecular weight solutes, whilst the inner membrane contains more specific transporters that control metabolite flux into and out of the organelle[22]. One striking difference between C_3 and C_4 photosynthesis is that, thanks to the spatial compartmentalisation of different steps of the C_4 cycle amongst different organelles (chloroplast and mitochondria) in two separate cell types, C_4 photosynthesis is much more dependent on intracellular metabolite transport than C_3 photosynthesis. In C_3 photosynthesis, the assimilation of three molecules of CO_2 into a molecule of triose phosphate (TP) requires only one intracellular metabolite transport

step: the export of TP from the chloroplast via the triose phosphate translocator (TPT) (Figure 1.4 A). In contrast, the assimilation of the same number of molecules of CO_2 using the NADP-ME C_4 pathway requires at least 31 intracellular metabolite transport steps[23] (Figure 1.4 B). Like C_3 plants, the final transport step is the export of TP from the chloroplast (although in C_4 plants this occurs in the bundle sheath). Prior to this, each run of the C_4 cycle, three of which are needed to generate three molecules of re-released CO_2 in the bundle sheath, requires at least ten intracellular metabolite transport steps. In a conventional NADP-ME C4 pathway these transmembrane transport steps are: the import of OAA and export of malate from the mesophyll chloroplast, the import of malate and export of pyruvate from the bundle sheath chloroplast, the import of pyruvate and export of PEP from the mesophyll chloroplast, and the exchange of TP and 3-PGA across the envelopes of both mesophyll and bundle sheath chloroplasts. The latter, whilst not a part of the core C_4 cycle, is a necessary feature of C_4 plants containing an NADP-ME cycle, and is proposed to be related to the downregulation of photosystem II (PSII) in the bundle sheath of these species[24,25]. This leads to cyclic electron transport around PSI, which produces ATP but not NADPH, both of which would normally be produced by linear electron transport in the presence of PSII. The absence of NADPH production by photosynthetic electron transport in the bundle sheath chloroplast may assist in kinetically favouring the forward (NADPH-producing) reaction of NADP-ME[26]. However, the reduction stage of the Calvin-Benson-Bassham cycle requires NADPH, and stoichiometrically only half of this can be provided by NADP-ME through decarboxylation of malate. Therefore, half the 3-PGA generated in the bundle sheath chloroplast must be transported to the mesophyll chloroplast to be converted to TP,

through the actions of phosphoglycerate kinase (PGK) and glyceraldehyde 3-phosphate dehygrogenase (GAPDH), the latter of which utilises NADPH produced by linear electron transport in the mesophyll chloroplast. TP must then be returned to the bundle sheath chloroplast for the completion of the Calvin cycle.

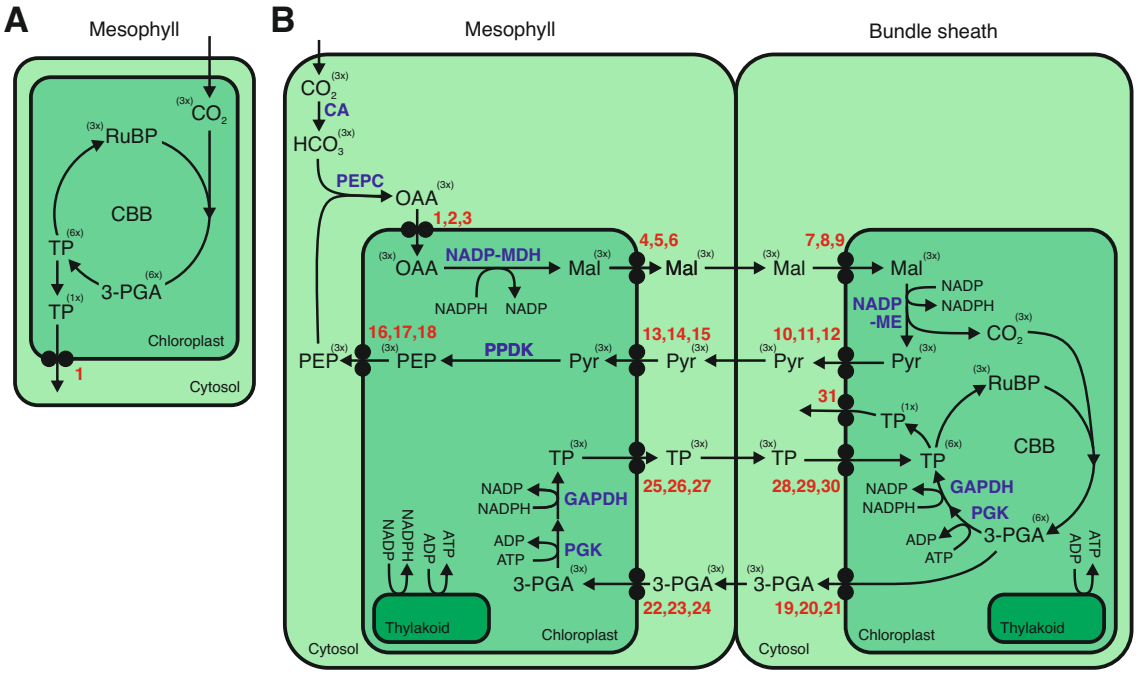


Figure 1.4: Intracellular metabolite transport in C_3 and C_4 photosynthesis. Diagrams showing minimal intracellular metabolite transport required for **A**) C_3 photosynthesis, and **B**) NADP-ME subtype C_4 photosynthesis. Metabolites are shown in black. Enzymes are shown in blue. Brackets show relative stoichiometries of each metabolite for the production and export of one molecule of TP. Counts of number of transport reaction required for the production and export of one molecule of TP are shown in red. For abbreviations see appendix A.1, table A.1.

Despite the importance of intracellular metabolite transport for C_4 photosynthesis, our knowledge of the transporters responsible for these transport steps lags significantly behind that of the soluble enzymes involved in the C_4 cycle. What follows is a review of our current knowledge of these transporters. No components of the C_4 cycle that have been identified so far have evolved *de novo*. Instead they have all been co-opted from existing enzymes and transporters used in plant metabolism. Therefore, much of the work on the transporter families that have been co-opted for use in C_4 photosynthesis has been done on C_3 plants (or even animals and fungi in the case of mitochondrial transporters), and this work will also be discussed here.

1.3 A review of metabolite transporters involved in C_4 photosynthesis

1.3.1 Chloroplastic dicarboxylate transport – the DCT family

Transport of dicarboxylates across the chloroplast envelope occurs primarily via transporters belonging to the divalent anion-sodium symporter (DASS) family. Although this family contains both symporters and antiporters, it is named after the Na^+:dicarboxylate symport mechanism exhibited by most of its bacterial and mammalian members[27]. However, members of the dicarboxylate transporter (DCT or DiT) subfamily from plants instead all function as chloroplast-localised dicarboxylate antiporters[28–30].

The first characterisation of dicarboxylate transport across the chloroplast envelope came with the observation of the release of pre-loaded radio-labelled dicarboxylates

(malate, succinate, or aspartate) from isolated spinach chloroplasts, in exchange for external dicarboxylates (malate, succinate, aspartate, glutamate, 2-oxoglutarate, fumarate, or OAA)[31]. Determination of the transport kinetics for these different dicarboxylates suggested the existence of two separate dicarboxylate exchangers of overlapping specificity: one that preferentially transports aspartate, and one that preferentially transports non-amino acid dicarboxylates[32]. One of the most important dicarboxylate fluxes across the chloroplast envelope in C_3 plants is the net exchange of 2-oxoglutarate (2-OG) for glutamate that is required during primary or photorespiratory nitrogen assimilation via the GS/GOGAT (glutamine synthetase/glutamine oxoglutarate aminotransferase) cycle in the chloroplast[33]. The function of two separate dicarboxylate transporters with differing specificities was further supported by the observation that NH_3 and 2-OG-dependent O_2 evolution and glutamate export (measures of the GOGAT reaction) from isolated spinach/pea chloroplasts are stimulated by malate[33–35]. Assuming the existence of only a single dicarboxylate transporter to facilitate 2-OG/glutamate exchange across the chloroplast envelope, malate would be expected to competitively inhibit this transporter and therefore NH_3 and 2-OG-dependent O_2/glutamate production. Instead, the observed stimulatory activity of malate can be explained by a two-transporter model in which 2-OG and glutamate are exchanged indirectly on separate carriers, facilitated by a cycling transport of malate (Figure 1.5 A)[33,34,36]. The purification and kinetic characterisation of a 2-OG/Malate transporter (OMT1, sometimes referred to as DiT1), that is unable to transport amino acids such as aspartate or glutamate[28,29,37,38], matches the first of these hypothesized transporters. Similarly, a general Dicarboxylate Transporter (DCT, sometimes referred to as DiT2)

capable of amino acid transport was identified through kinetic analysis of spinach chloroplasts pre-loaded with glutamate[39], and subsequent purification and kinetic characterisation of the transporter itself[28,29]. The role of DCT in nitrogen assimilation is further supported by the severely stunted or even non-viable phenotype (depending on the report) of *dct* mutant *Arabidopsis thaliana* plants grown under photorespiratory conditions (atmospheric O_2), but not under non-photorespiratory conditions (high CO_2)[28,29,40]. Similarly, RNAi-mediated OMT1 knock-down spinach plants and *omt1* mutant *A. thaliana* show photorespiratory phenotypes indicative of disrupted nitrogen assimilation[41,42]. The comparatively more severe phenotype of *dct* compared to *omt1* mutant *A. thaliana* is consistent with a two-transporter model for nitrogen assimilation in which the narrower substrate specificity of OMT1 means that OMT1 cannot compensate for the loss of DCT, but DCT can partially compensate for the loss of OMT1.

Nitrogen assimilation is not the only cellular process in which dicarboxylate exchange across the chloroplast envelope plays a role. Reducing equivalents generated by photosynthetic electron transport in the chloroplast thylakoid membranes must be exported to be used in the cytosol, for example in nitrate reduction, or may be imported into other subcellular compartments, for example into the peroxisome for photorespiration or into the mitochondria for mitochondrial electron transport. Additionally, under stress conditions, a decrease in the consumption of stromal reducing equivalents can lead to a backing up of photosynthetic electron transport and the generation of damaging reactive oxygen species (ROS) if excess reducing equivalents are not exported from the chloroplast[43]. One mechanism for the transport of reducing equivalents across intracellular membranes is the ‘malate valve’, in which an

OAA/malate exchanger acts in combination with malate dehydrogenases acting in opposite directions on either side of the membrane[44]. OMT1 is also capable of efficient OAA/malate exchange[29], but based on kinetic data suggesting that OAA uptake into the chloroplast by OMT1 would be strongly inhibited by much higher cytosolic 2-OG and malate concentrations, the existence of a separate, high-affinity OAA Transporter (OAT) was postulated[45]. However, a significantly increased K_m for OAA uptake into chloroplasts isolated from *omt1* mutant *A. thaliana* suggests that OMT1 is the only high-affinity OAA transporter expressed in *A. thaliana* chloroplasts[41]. It is still not understood how OAA import via OMT1 occurs however, given predicted inhibition by cytosolic 2-OG and malate. Increased indicators of photo-inhibition in *omt1* mutant *A. thaliana* plants exposed to high light further support a function for OMT1 as a malate valve that facilitates the indirect export of reducing equivalents from the chloroplast stroma to the cytosol[41] (Figure 1.5 B). The dual function of OMT1 in nitrogen assimilation as well as intracellular redox balancing may help to integrate these two pathways. During primary nitrogen assimilation, OMT1 may help to coordinate the supply of reducing power to cytosolic nitrate reductase (NR)[29]. Similarly, during photorespiratory nitrogen assimilation, OMT1 may help to coordinate the supply of reducing power to peroxisomal hydroxypyruvate reductase, although some of this reducing power could also be supplied through a mitochondrial malate valve coupling photorespiratory, peroxisomal hydroxypyruvate reductase reaction to photorespiratory, mitochondrial glycine decarboxylase[46,47].

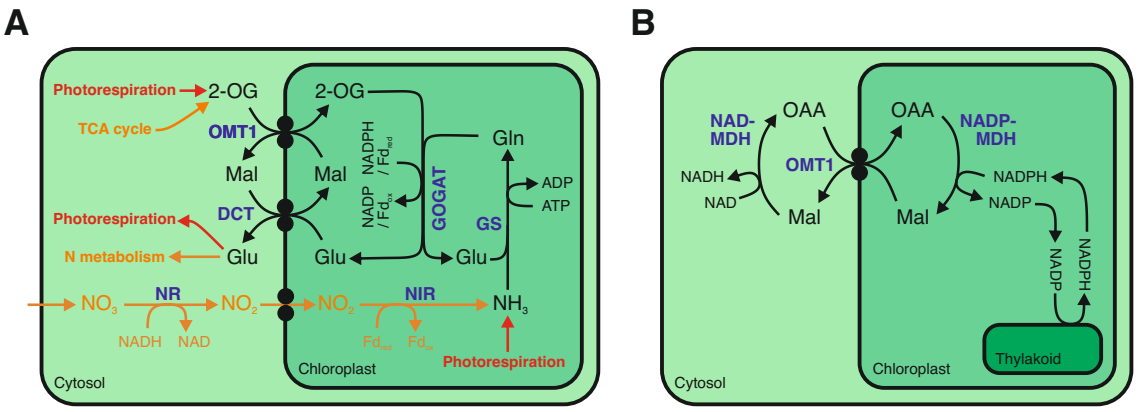


Figure 1.5: DCT transporters in C_3 plants. **A**) DCT function in nitrogen assimilation. Entry and exit points to primary N assimilation are shown in orange, and to secondary (photorespiratory) N assimilation are shown in red. **B**) OMT1 function in redox balancing. Metabolites are shown in black. Enzymes are shown in blue. Enzymes and transporters are shown in blue. For abbreviations see appendix A.1, table A.1.

In C_4 plants, DCT subfamily proteins have been co-opted from their ancestral roles to facilitate C_4 cycle dicarboxylate exchange across the chloroplast envelope[30]. As is typical for components of the C_4 cycle, *DCT* subfamily genes show higher expression in C_4 than C_3 plants, and this expression is asymmetrically distributed between mesophyll and bundle sheath cells. In maize there are three *DCT* subfamily genes – *OMT1*, *DCT1*, and *DCT2*. *ZmOMT1* is expressed strongly in the mesophyll only, *ZmDCT1* is expressed to a lower level, and preferentially in the mesophyll, and *ZmDCT2* is expressed strongly in the bundle sheath only[30,48] (Figure 1.6 A). This pattern is recapitulated in studies on chloroplast envelope proteomes of maize mesophyll and bundle sheath cells[24,49,50], which also show higher protein levels for all three proteins in C_4 maize than C_3 pea[51]. It is therefore proposed that ZmOMT1 localised in the mesophyll chloroplast envelope facilitates the exchange of cytosolic OAA (generated by PEPC) with chloroplastic malate

(generated by NADP-MDH), as part of the C_4 cycle[30] (Figure 1.6 B). In a similar manner, smaller quantities of ZmDCT1 localised in the mesophyll chloroplast envelope may facilitate the exchange of OAA with chloroplastic aspartate (generated by AspAT), as part of the auxiliary PCK cycle in maize (Figure 1.6 B). The generation and characterisation of *omt1* and *dct1* mutants in maize would be useful to confirm these hypotheses.

In contrast to the proposed role of OMT1 in the mesophyll cell chloroplast, ZmDCT2 is proposed to be involved in the import of malate into the bundle sheath chloroplast for the C_4 cycle (Figure 1.6 C). This is supported by the analysis of *dct2* mutant maize[52]. Bundle sheath chloroplasts isolated from *dct2* plants show significantly slower rates of pyruvate production and malate uptake compared to wild-type plants[52]. Additionally, radio-labelled CO_2 and malate feeding experiments (using whole plants and isolated bundle sheath cells respectively) indicate an increased production of aspartate in place of malate: consistent with a diversion in flux from the disrupted NADP-ME C_4 cycle to the auxiliary PCK cycle[52]. It should be noted that DCT2, being an obligate dicarboxylate exchanger, cannot by itself facilitate the net import of malate into the bundle sheath chloroplast that is required to run the C_4 cycle. Therefore, the model proposed by Weissmann et al. to explain the apparent role of DCT2 in bundle sheath chloroplast malate import involves proposing the existence of a separate, as-yet-unidentified bundle sheath chloroplast-localised aspartate uniporter (or cation symporter), which imports aspartate into the bundle sheath chloroplast so that DCT2 can subsequently carry out aspartate export in exchange for malate import (Figure 1.6 C). This would be consistent with previous observations that malate-dependent pyruvate formation from

wild-type maize bundle sheath cells or chloroplasts is stimulated by aspartate[53,54]. Alternatively, the most parsimonious explanation would simply be a missing bundle sheath chloroplast-localised malate uniporter (or cation symporter) (Figure 1.6 C).

Consistent with multiple independent origins of C_4 photosynthesis, different C_4 species have subfunctionalised different *DCT* subfamily members for different roles. For example, other NADP-ME subtype C_4 species *Sorghum bicolor* (sorghum), *Setaria italica,* and *Setaria viridis* show strong bundle sheath-specific expression of *DCT4*, a duplicate of *DCT1* that is not found in maize, in place of *DCT2*[55–57] (Figure 1.6 A). By contrast, species that do not utilise NADP-ME and instead carry out dicarboxylate decarboxylation in cellular compartments other than the chloroplast, for example the primarily NAD-ME species *Panicum virgatum* and *Cleome gynandra*, and primarily PCK species *Urochloa fusca*, have low expression of *DCT* genes[57–59] (Figure 1.6 A).

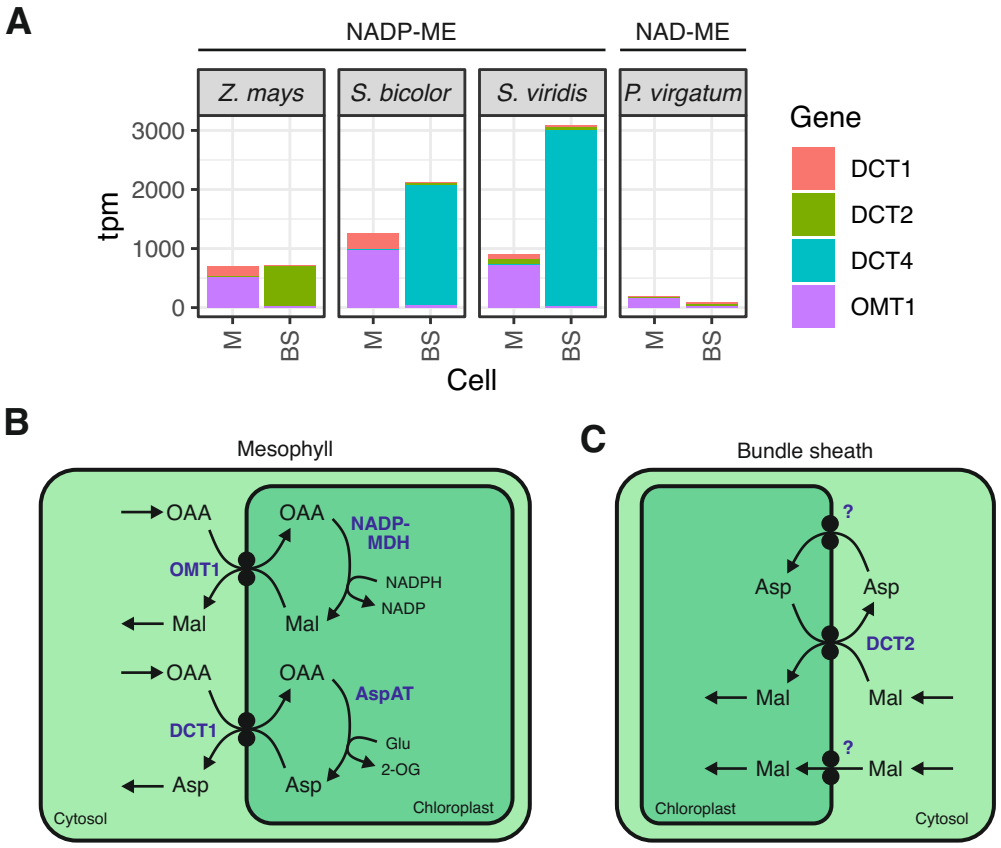


Figure 1.6: DCT transporters in C_4 plants. **A**) Cell-type specific transcript abundance of DCT family genes among NADP-ME and NAD-ME subtype C_4 monocots. *Z. mays* data from Chang et al. (2012)[48]. *S. bicolor* data from Emms et al. (2016)[55]. *S. viridis* data from John et al. (2014)[56]. *P. virgatum* data from Rao et al. (2016)[59]. M = mesophyll. BS = bundle sheath. tpm = transcripts per million. **B**) Function of OMT1 and DCT1 in mesophyll cell of C_4 plants. **C**) Hypothesised function of DCT2 in bundle sheath cell of C_4 plants. Metabolites are shown in black. Enzymes and transporters are shown in blue. For abbreviations see appendix A.1, table A.1.

1.3.2 Chloroplastic monocarboxylate/phosphate transport – the TPT family

The primary transport flux of monocarboxylates across the chloroplast inner envelope occurs via transporters belonging to the Triose Phosphate Translocator (TPT) family, members of the Drug/Metabolite Transporter (DMT) superfamily. TPT family proteins

are specific to eukaryotes[60], and catalyse the exchange of phosphorylated compounds with inorganic phosphate (P_i).

TPT

The first observation of TPT activity was the exchange of P_i with phosphorylated 3-carbon compounds across the envelopes of isolated spinach chloroplasts[31]. The rate of exchange was high with triose phosphates phosphorylated on the third carbon (e.g. 3-PGA, dihydroxyacetate phosphate (DHAP), glyceraldehyde 3-phosphate (GAP)), and lower for triose phosphates phosphorylated on the second carbon (e.g. PEP, 2-PGA), and lower still for sugar phosphates[31,61]. Given these transport properties, an obvious function for TPT is the export of 3C-phosphorylated triose phosphates, the primary products of photosynthesis, from the chloroplast to the cytosol to support sucrose biosynthesis. Interestingly, potato and spinach plants with RNAi-mediated knock-down of TPT and *tpt* mutant *A. thaliana* plants did not show severe impairment of growth or photosynthetic defects[62–64], as might be expected given disruption to the primary mechanism for chloroplastic export of fixed carbon. However, metabolic analysis demonstrates that this is due to compensatory diversion of carbon flux into chloroplastic starch generation and later break-down into glucose-1P for export from the chloroplast via a separate glucose translocator[62,64,65]. TPT itself plays a direct role in this coordination between sucrose and starch biosynthesis: reduced sucrose biosynthesis, and therefore reduced triose phosphate exchange with P_i through TPT, leads to reduced P_i and increased triose phosphate levels in the stroma, which leads to increased starch

biosynthesis via the allosteric activation of chloroplastic ADP-glucose pyrophosphorylase (Figure 1.7 A)[63].

Additional roles for TPT in C_3 plants have been proposed. Similarly to the chloroplast malate valve, TPT can act in concert with coupled chloroplastic and cytosolic enzymes to indirectly export reducing equivalents to the cytosol, although in this case TPT also exports ATP[43]. In the chloroplast, 3-PGA is converted by the ATP-consuming PGK and NADPH-consuming GAPDH into GAP, whilst in the cytosol GAP is converted by the NADH-producing GAPDH and ATP-producing PGK back to 3-PGA. In this way, both ATP and reducing power are indirectly "shuttled" from the chloroplast to the cytosol. The exchange of 3-PGA and GAP occurs indirectly through separate TPTs by cycling P_i (Figure 1.7 B).

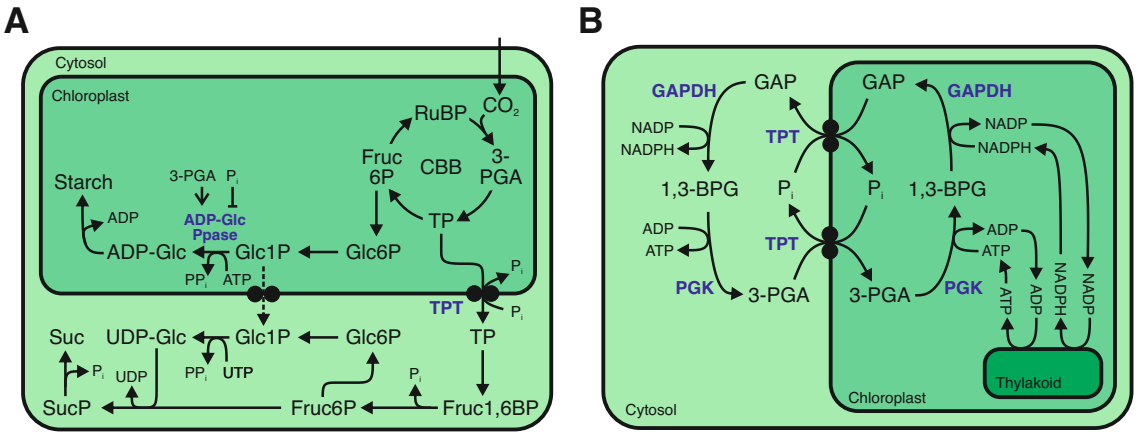


Figure 1.7: TPT in C_3 plants. **A**) TPT function in photosynthetic TP export and coordination of starch and sucrose biosynthesis. **B**) TPT function in exporting reducing equivalents and ATP from the chloroplast. Metabolites are shown in black. Enzymes and transporters are shown in blue. For abbreviations see appendix A.1, table A.1.

In C_4 plants, this redox shuttling across mesophyll and bundle sheath chloroplast envelopes, in concert with 3-PGA/TP exchange between mesophyll and bundle sheath cells, is used to transfer reducing equivalents produced by linear electron transport in mesophyll chloroplasts to bundle sheath chloroplasts (Figure 1.8B). As stated above (section 1.2), in NADP-ME subtype C_4 plants this supply of reducing equivalents from mesophyll to bundle sheath is proposed to compensate for the deficiency in reducing equivalents in the bundle sheath chloroplast due to down-regulated PSII[43]. Consistent with this role, NADP-ME subtype C_4 plants show higher leaf *TPT* RNA and TPT protein levels than C_3 plants[51,66], and high *TPT* RNA and TPT protein levels in both mesophyll and bundle sheath[24,48–50,55,56] (Figure 1.8A). However, it should be noted that *TPT* is also expressed to high levels in both mesophyll and bundle sheath of NAD-ME subtype C_4 plants such as *P. virgatum* and *C. gynandra*[59], despite the fact that NAD-ME plants do not show depleted bundle sheath chloroplast PSII activity[67] (Figure 1.8A). It is therefore likely that redox balancing is required between mesophyll and bundle sheath chloroplasts in all C_4 plants, not just in NADP-ME subtype C_4 plants where the down-regulation of PSII in the bundle sheath makes this need more apparent.

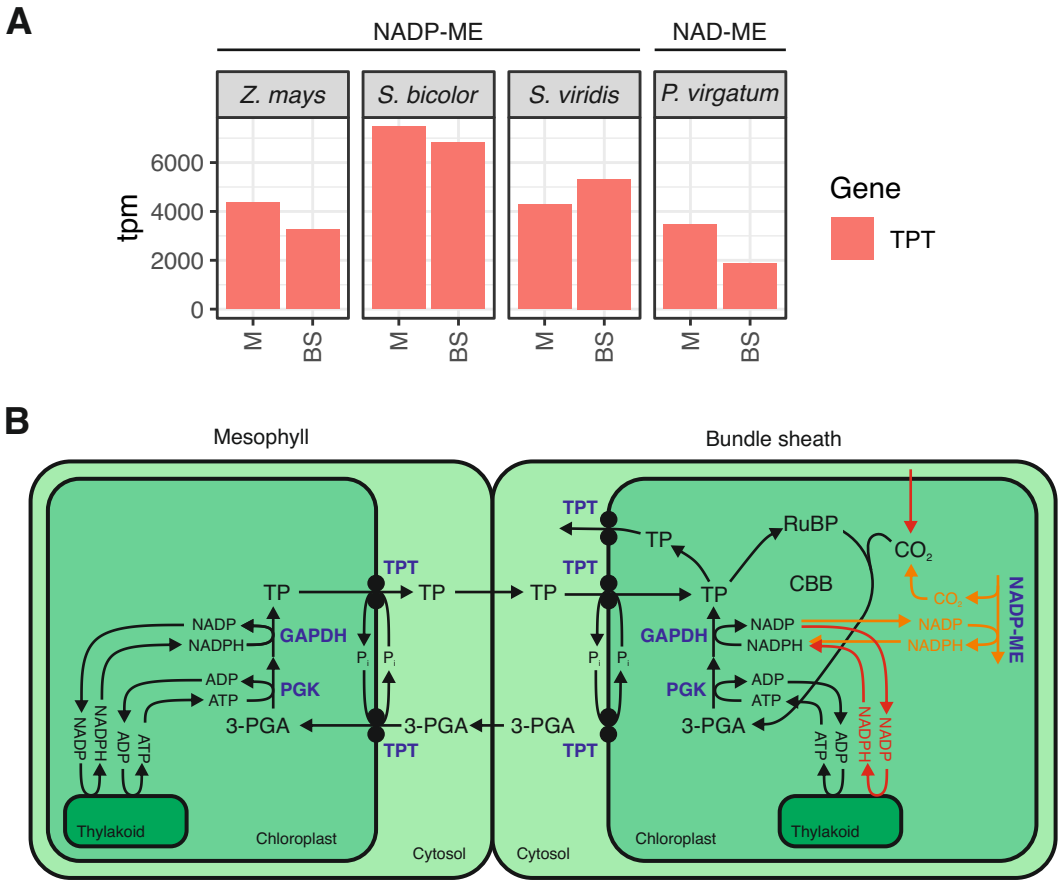


Figure 1.8: TPT in C_4 plants. **A**) Cell-type specific transcript abundance of TPT genes among NADP-ME and NAD-ME subtype C_4 monocots. *Z. mays* data from Chang et al. (2012)[48]. *S. bicolor* data from Emms et al. (2016)[55]. *S. viridis* data from John et al. (2014)[56]. *P. virgatum* data from Rao et al. (2016)[59]. M = mesophyll. BS = bundle sheath. tpm = transcripts per million. **B**) Functions of TPT in C_4 plants. Metabolites are shown in black. Enzymes and transporters are shown in blue. NADP-ME subtype-only parts of pathway are shown in orange. NAD-ME subtype-only parts of pathway are shown in red. For abbreviations see appendix A.1, table A.1.

GPT and XPT

The exchange of 3-PGA and GAP with P_i is not the only important transport reaction carried out by members of the TPT family. It was initially assumed that the exchange of

phosphate with triose phosphates and sugar phosphates was all carried out on the same broad-specificity TPT carrier. However, plastids from different plant tissues were found to exchange different phosphorylated compounds for P_i with differing efficiencies[68]. For example, plastids isolated from pea roots and maize endosperm can efficiently exchange P_i with glucose-6P, in order to compensate for their lack of the gluconeogenesis enzyme fructose 1,6-bisphosphatase[69,70]. This exchange reaction is facilitated by a separate TPT-family carrier, the glucose-6P/P_i translocator (GPT)[71] (Figure 1.9). In contrast to *TPT* itself, in C_3 plants *GPT* tends to be highly expressed in non-green tissues, where it is proposed to supply plastids with glucose-6P for starch biosynthesis and/or the oxidative pentose phosphate pathway (OPPP)[71]. Expression in photosynthetic tissues is generally low, although induction of *GPT* expression in photosynthetic tissues during growth under elevated CO_2[72], conditions causing sugar accumulation[73,74], and upon transition from low to high light[75,76] indicate a potential role under variable conditions[77]. More recently, another TPT-family member was identified based on sequence similarity to GPT: the xylulose-5P/P_i translocator (XPT). XPT, which is expressed throughout the plant, is an efficient transporter of xylulose-5P and is proposed to connect cytosolic and plastidic reactions within the OPPP[78] (Figure 1.9).

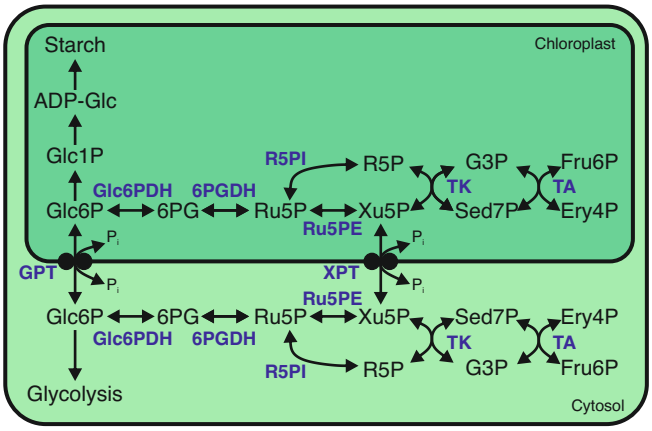


Figure 1.9: GPT and XPT in C_3 plants. For abbreviations see appendix A.1, table A.1.

Potential functions of GPT and XPT in C_4 plants have not yet been studied. Interestingly, *GPT* genes show high, cell-type specific expression in several C_4 plants. Maize and Sorghum (NADP-ME subtype grasses), as well as *C. gynandra* (NAD-ME subtype dicot) show bundle sheath-preferential expression[48,55], whilst *S. italica* shows mesophyll-preferential expression[56] (Figure 1.10). Clearly further work needs to be done to investigate a potential interaction between GPT transport function and C_4 photosynthesis.

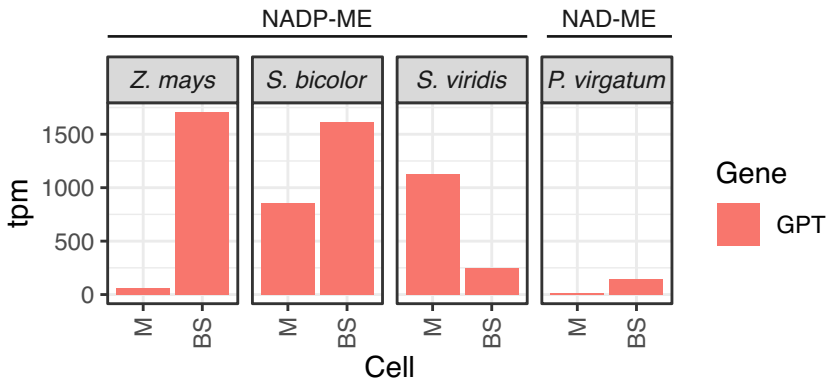


Figure 1.10: GPT in C_4 plants. Cell-type specific transcript abundance of GPT among NADP-ME and NAD-ME subtype C_4 monocots. *Z. mays* data from Chang et al. (2012)[48]. *S. bicolor* data from Emms et al. (2016)[55]. *S. viridis* data from John et al. (2014)[56]. *P. virgatum* data from Rao et al. (2016)[59]. M = mesophyll. BS = bundle sheath. tpm = transcripts per million.

PPT

A final member of the TPT family is the PEP/P_i translocator PPT, which is an efficient transporter of triose phosphates that have been phosphorylated on the second carbon, such as PEP and 2-PGA, although a low affinity for 2-PGA means that *in vivo* PEP is likely the main substrate[79]. In *A. thaliana* there are two *PPT* genes: *PPT1* and *PPT2. PPT2* is ubiquitously expressed throughout the leaf, and is hypothesized to carry out the housekeeping role of PEP import into the chloroplast, for entry into the shikimate pathway for the production of aromatic amino acids[80,81] (Figure 1.11 A). *PPT1*, on the other hand, is specifically expressed in the leaf vasculature (and not the mesophyll cells) and the roots[79,80]. PPT1 is also sometimes referred to as CAB (chlorophyll a/b binding protein) UnderExpressed 1 (CUE1), due to its initial identification in an *A. thaliana* mutant screen for regulators of light-responsive genes[82,83]. These plants show stunted

roots and a reticulate leaf phenotype caused by impaired mesophyll chloroplast development[82], which in turn is related to a deficit in the PEP-derived aromatic amino acid phenylalanine and downstream phenolic compounds (like flavonoids and anthocyanins)[81,83]. Therefore, it is proposed that PEP transport by *PPT1* is involved in the synthesis of an unknown plant secondary metabolite in the leaf vasculature that is not required in the vasculature itself, but diffuses into the mesophyll where it is required for mesophyll cell development[80,81] (Figure 1.11 B). This is consistent with work in tobacco showing that several phenolic compounds are involved in the regulation of cell division and expansion in leaf mesophyll cells[84], and that disruption in phenolic metabolism can lead to a reticulate leaf phenotype very similar to *A. thaliana cue1*[84,85].

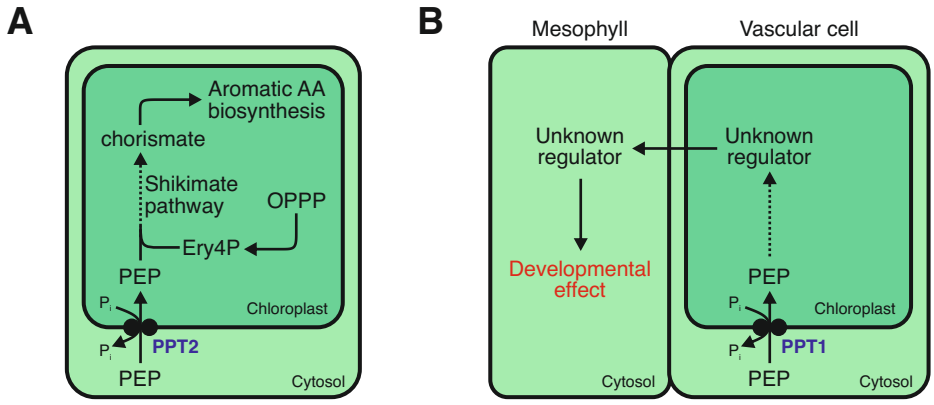


Figure 1.11: PPT in C_3 plants. **A**) PPT2 function in chloroplastic PEP import for the shikimate pathway. **B**) PPT1 developmental function. Metabolites and pathways are shown in black. Transporters are shown in blue. For abbreviations see appendix A.1, table A.1.

In C_4 plants of all subtypes, a novel chloroplast envelope PEP transport reaction is required. Rather than PEP import, PEP must be exported from the mesophyll chloroplast to provide the substrate for PEPC (Figure 1.12 B). Usually, an ortholog of *PPT1* has been co-opted for function in the C_4 cycle[86], and correspondingly shows higher RNA and protein levels in C_4 plants than in C_3 relatives[51,58,66,87], as well as preferential expression in the mesophyll[24,48–50,55,56,59] (Figure 1.12 A). Although transport activity and cell-type expression data are highly indicative, PPT function in the C_4 cycle has never been directly tested through analysis of *PPT* knock-down or knock-out C_4 plants. As well as confirming a role for PPT in the C_4 cycle, it would be interesting to investigate if the movement of *PPT1* expression from bundle sheath to mesophyll during the evolution of C_4 photosynthesis is related to the morphological adaptations present in C_4 plants, given the apparent developmental role played by PPT1 in C_3 plants.

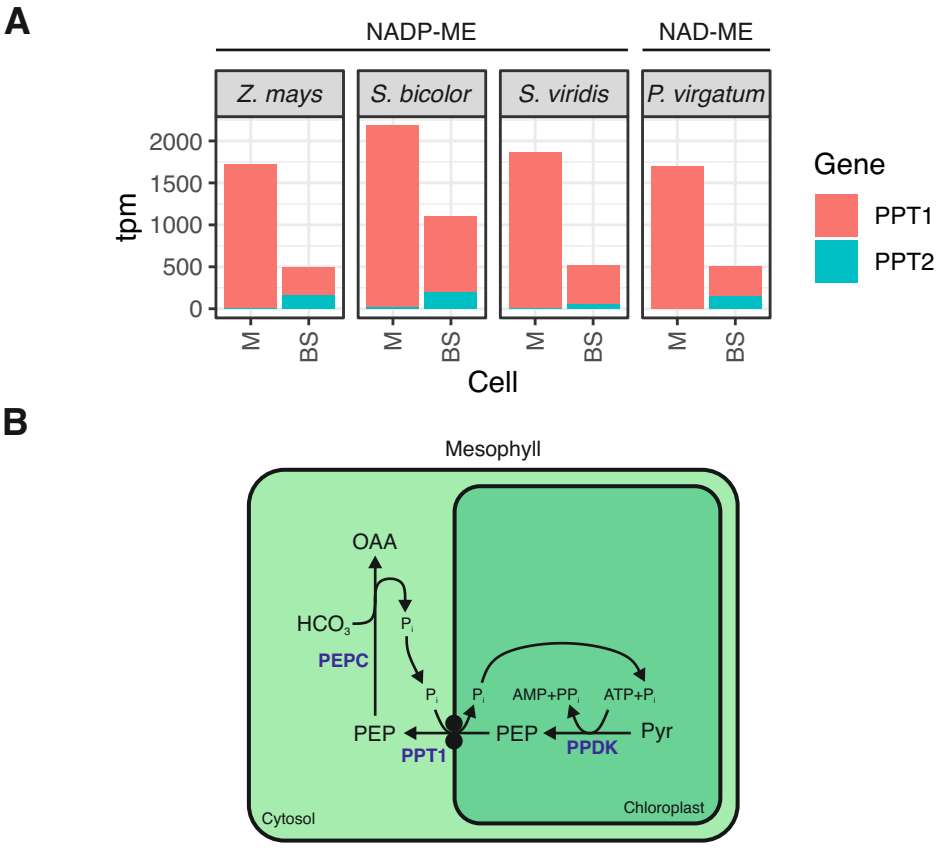


Figure 1.12: PPT in C_4 plants. **A**) Cell-type specific transcript abundance of PPT genes among NADP-ME and NAD-ME subtype C_4 monocots. *Z. mays* data from Chang et al. (2012)[48]. *S. bicolor* data from Emms et al. (2016)[55]. *S. viridis* data from John et al. (2014)[56]. *P. virgatum* data from Rao et al. (2016)[59]. M = mesophyll. BS = bundle sheath. tpm = transcripts per million. **B**) Function of PPT in C_4 plants. Metabolites are shown in black. Enzymes and transporters are shown in blue. For abbreviations see appendix A.1, table A.1.

1.3.3 Chloroplastic pyruvate transport – the BASS family

One family of transporter proteins has been linked with chloroplast pyruvate import so far. This is the Bile Acid/Sodium Symporter (BASS) family, of the Bile/Arsenite/Riboflavin Transport (BART) superfamily. These families are so-named due to their function in

animals of transporting bile salts[88]. However, BASS proteins are also found in prokaryotes and other eukaryotes, including plants. In plants, there are five BASS subfamilies, BASS1-5[89].

BASS2

Carrier-mediated chloroplast pyruvate uptake was first demonstrated using chloroplasts isolated from the mesophyll cells of the C_4 plant *Digitaria sanguinalis*[90]. It was later shown that this transport reaction in maize is light-dependent[91], and in *Panicum miliaceum* that pyruvate uptake is directly dependent on the proton gradient that is generated across the chloroplast envelope under illumination[92,93]. Subsequent sampling across a broader species range revealed that light-dependent chloroplast pyruvate uptake can be divided into two mechanisms: H^+- and Na^+-dependent[94]. H^+-dependent pyruvate uptake is limited to two sister lineages within the grasses that are both derived from the same origin of C_4 photosynthesis[95]: the Arundinelleae and Andropogoneae, the latter of which contains several important C_4 crop species such as maize and sorghum[94,96]. Na^+-dependent pyruvate uptake on the other hand is found in the remaining monocot as well as all dicot clades tested so far[94,97,98].

Na^+-dependent pyruvate import is now known to be facilitated by BASS2, in conjunction with Na^+/H^+ antiporter 1 (NHD1)[99] (Figure 1.12 A). Both *BASS2* and *NHD1* are expressed to higher levels in the leaves of Na^+-type C_4 species than in close C_3 relatives[58,87], and BASS2 protein was similarly found in the leaves of multiple Na^+-type C_4 species but not H^+-type C_4 species or C_3 species[99]. As expected, *BASS2* and *NHD1* are also preferentially expressed in the mesophyll cells of Na^+-type C_4 species *S. italica* and *P. virgatum*[56,59]

(Figure 1.13 A). The mechanism of BASS2 function was elucidated through heterologous expression in *Escherichia coli*. *E. coli* cells expressing BASS2 showed significantly increased rates of pyruvate uptake, which was further increased by the addition of Na^+: indicating that BASS2 acts as a pyruvate/Na^+ symporter[99]. The previously known Na^+/H^+ antiport activity of NHD1[100], co-expression in C_4 plants with BASS2, and the fact that heterologous co-expression with BASS2 in *E. coli* in the absence of added Na^+ led to pyruvate uptake rates correlating with NHD1 protein levels, led to a suggestion for the shared mechanism of BASS2 and NHD1[99]. It is proposed that NHD1 utilises the light-dependent proton gradient across the chloroplast inner envelope to couple the import of H^+ down the proton gradient to the export of Na^+ into the intermembrane space. Pyruvate is then imported by pyruvate/Na^+ symport through BASS2[99] (Figure 1.13 B).

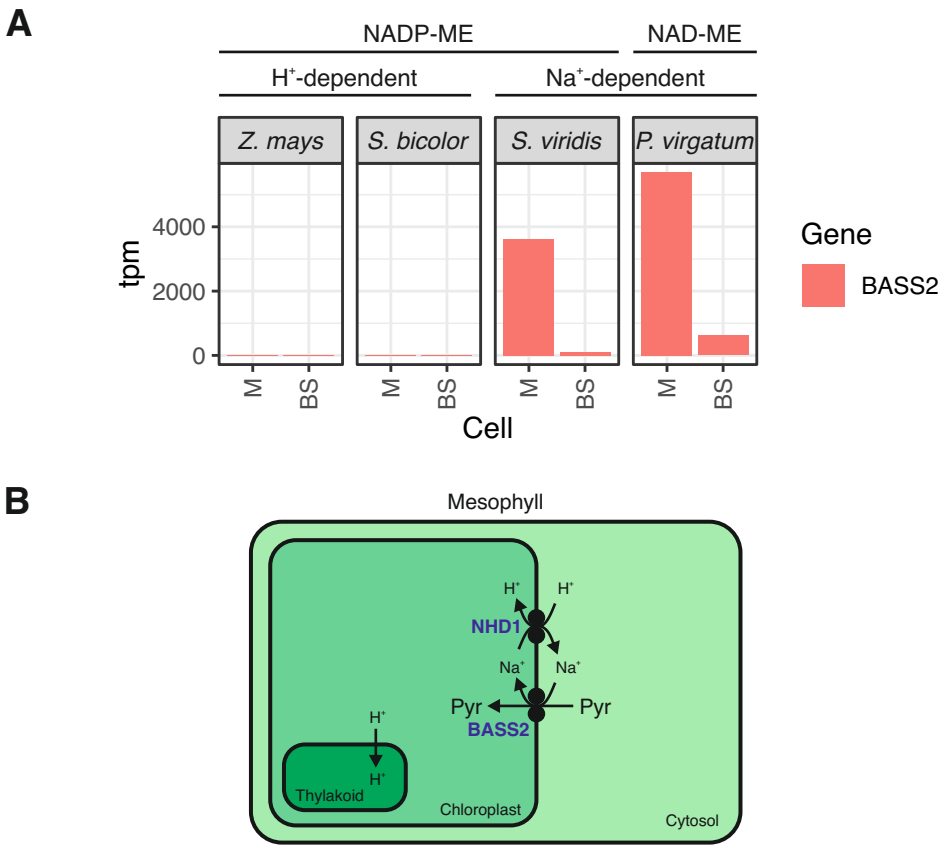


Figure 1.13: BASS2 in C_4 plants. **A**) Cell-type specific transcript abundance of BASS2 gene among NADP-ME and NAD-ME subtype C_4 monocots. H^+- or Na^+-dependence of mesophyll cell pyruvate uptake is also indicated. *Z. mays* data from Chang et al. (2012)[48]. *S. bicolor* data from Emms et al. (2016)[55]. *S. viridis* data from John et al. (2014)[56]. *P. virgatum* data from Rao et al. (2016)[59]. M = mesophyll. BS = bundle sheath. tpm = transcripts per million. **B**) Function of BASS2 in C_4 plants. Metabolites are shown in black. Enzymes and transporters are shown in blue. For abbreviations see appendix A.1, table A.1.

Whilst analysis of C_4 *bass2* mutant plants is yet to be performed, the function of BASS2 in C_3 plants has been investigated using *A. thaliana bass2* mutant plants. Chloroplasts isolated from *bass2* mutant *A. thaliana* plants show significantly reduced light- or Na^+-

dependent pyruvate uptake[99]. Under normal conditions the mutant plants appeared phenotypically normal, but growth reduction in the presence of mevastatin was more severe in *bass2* mutant plants compared to wild-type[99]. Mevastatin is an inhibitor of the rate-limiting enzyme in the mevalonic acid (MVA) pathway: the cytosolic terpenoid biosynthesis pathway. Terpenoids can also be synthesised by the chloroplastic methyl erythritol phosphate (MEP) pathway, one of the substrates for which is pyruvate. Therefore, it is proposed that in C_3 plants, BASS2 acts to import pyruvate into the chloroplast for entry into the MEP pathway for terpenoid biosynthesis, but that *bass2* mutants are only phenotypically affected when the alternative MVA pathway is also disrupted (Figure 1.14). In wheat (*Triticum aestivum*) BASS2 has been shown to be involved in salinity tolerance[101]. Abscisic acid (ABA), a terpenoid derived from the MEP pathway, is known to be important in the salinity stress response. Although strangely BASS2 seems to be involved not by modulated ABA levels but, through an unknown mechanism, by modulating the expression of a downstream transcription factor *ABA insensitive 4* (*ABI4*)[101].

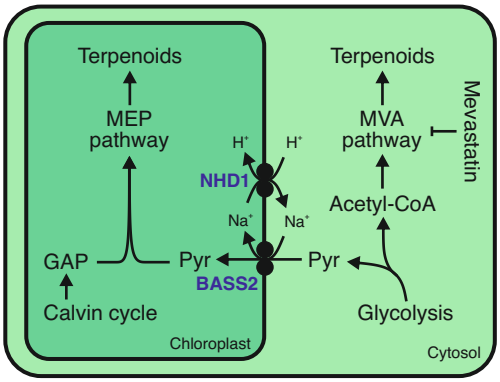


Figure 1.14: Function of BASS2 in C_3 plants. Metabolites and pathways are shown in black. Enzymes and transporters are shown in blue. For abbreviations see appendix A.1, table A.1.

Other BASS genes (BASS1, BASS4, BASS6)

Other BASS family members with different substrate specificities have also been characterised in plants. In a C_3 context, *A. thaliana* BASS1 was demonstrated to act as a chloroplast pantoate transporter involved in pantothenate biosynthesis[102], whilst *A. thaliana* BASS6 was shown to facilitate the export of photorespiratory glycolate from the chloroplast[103], alongside the previously characterised plastidic glycolate/glycerate transporter (PLGG1)[104]. In a C_4 context, BASS4 was identified alongside BASS2 as a potential C_4 cycle transporter due to high expression in some C_4 dicots compared to C_3 relatives (from the *Flaveria* genus)[99], although it is not highly expressed in the C_4 dicot *Cleome gynandra*[99] or any C_4 monocots tested so far[48,55,56,59]. *Flaveria bidentis* BASS4 reconstituted in liposomes shows Malate/Na^+ symport activity, and *FbBASS4* knock-down lines show significant growth defects, reduced photosynthetic rates, and metabolic alterations consistent with a function of bundle sheath chloroplast malate

import as part of the C_4 cycle[105]. Thus, whilst DCT and TPT family proteins seem to have been ubiquitously recruited for function in the C_4 cycle among independent origins of C_4 photosynthesis, BASS family proteins, perhaps owing to their comparatively smaller role and correspondingly lower expression in C_3 plants, have only been recruited for function in the C_4 cycle in a subset of C_4 origins.

1.3.4 Mitochondrial pyruvate transport – the MPC family

Transport of pyruvate across the mitochondrial membrane is facilitated by members of the Mitochondrial Pyruvate Carrier (MPC) family, which are found across the eukaryotes, and catalyse a pyruvate/H^+ symport reaction[106,107].

The import of pyruvate from cytosolic glycolysis into the mitochondria for entry into the tricarboxylic acid (TCA) cycle during aerobic respiration is a crucial transport step in most eukaryotes. Mitochondrial pyruvate import is also required to support the biosynthesis of branched-chain amino acids leucine and valine (Figure 1.15). The first biochemical evidence for a specific pyruvate carrier was obtained in the 1970s[106,108], but it was not until 2012 that two groups independently uncovered the molecular identity of the MPC, primarily through experiments in yeast[107,109]. These experiments showed that in yeast a complex of MPC1 with either MPC2 or MPC3 facilitates mitochondrial pyruvate import to support both the TCA cycle and valine/leucine biosynthesis[107,109]. This function is also conserved in MPC proteins from animals (humans, mice, and *Drosophila melanogaster*)[107,109].

MPC proteins have subsequently been characterised in the C_3 plant *A. thaliana*. The MPC family comprises four members in *A. thaliana*: MPC1, MPC2, MPC3, and an MPC2-like protein called NRGA1. Mutations in some of these genes lead to only minor phenotypic changes. *MPC1* and *NRGA1* (Negative Regulator of Guard Cell ABA signalling 1) are highly expressed in *A. thaliana* guard cells, and mutant lines for both show increased ABA sensitivity and drought tolerance[110,111]. This is proposed to be related to an ABA-induced increase in cytosolic pyruvate concentration, which is more pronounced in these lines that are defective in mitochondrial pyruvate uptake, and which interferes with signalling processes controlling stomatal closure[111]. *mpc1* mutant *A. thaliana* are also more susceptible to cadmium toxicity, with metabolic analysis of these plants showing increased pyruvate levels and decreased levels of TCA cycle intermediates compared to wild-type, consistent with reduced mitochondrial pyruvate uptake[112]. Susceptibility to Cadmium stress is proposed to be related to the re-direction of glutamate to support the TCA cycle in *mpc1* plants, which is therefore not available for the biosynthesis of glutathione: an important metabolite for cadmium tolerance[112]. Mitochondria isolated from *mpc1* mutant *A. thaliana* show significantly reduced pyruvate-dependent O_2 consumption (respiration) and pyruvate uptake rates, compared to wild-type[113]. The apparent discrepancy in *mpc1* mutant *A. thaliana* between relatively minor phenotypes (altered drought and cadmium tolerance), and a significant disruption to the crucial supply of pyruvate to the TCA cycle, is explained by the presence of two alternative pathways to indirectly supply mitochondrial pyruvate: namely the import of malate (e.g. via DIC, see below) and oxidation to pyruvate by NAD-ME, or the import of alanine via an unknown transporter and conversion to pyruvate by AlaAT[113] (Figure 1.15).

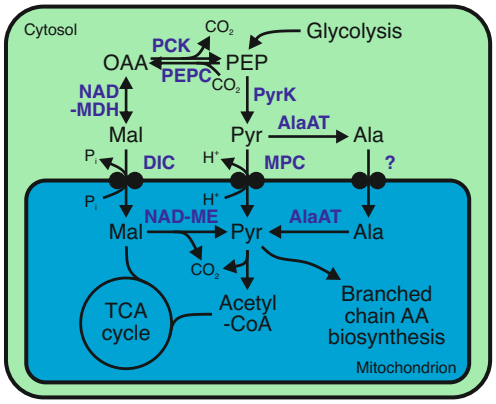


Figure 1.15: Function of MPC in C_3 plants. Metabolites and pathways are shown in black. Enzymes and transporters are shown in blue. For abbreviations see appendix A.1, table A.1.

NAD-ME subtype C_4 photosynthesis requires pyruvate transport across the bundle sheath mitochondrial membrane, although here the export of NAD-ME-derived pyruvate is required, in contrast to the characterised pyruvate import activity of MPC. The NAD-ME subtype dicot *C. gynandra* shows significantly upregulated *MPC1* and *MPC3* compared to close C_3 relative *C. spinosa*, although not to the extent of other C_4 cycle genes[58]. *MPC* genes in *C. gynandra* are also not differentially expressed between mesophyll and bundle sheath[114], as would be expected given the hypothesised bundle sheath-specific C_4 function. The NAD-ME subtype monocot *P. virgatum* does show preferential expression in the bundle sheath for *MPC1* and *MPC2*[59]. This is in contrast to NADP-ME subtype monocots, which show only very low bundle sheath expression of *MPC* genes, as would be expected given the lack of mitochondrial malate decarboxylation in this subtype[48,55,56] (Figure 1.16). Despite this somewhat promising expression data, there is no experimental evidence to support the hypothesis that MPC

can facilitate mitochondrial pyruvate export. In fact, mitochondria isolated from *mpc1* knock-out *A. thaliana* are still capable of exporting pyruvate[113], indicating the presence of another, as-yet-unidentified pyruvate exporter in plants that would also be a candidate for co-option into the C_4 cycle.

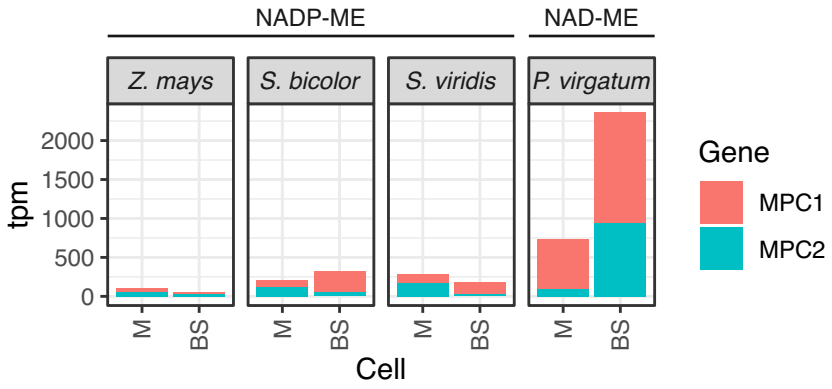


Figure 1.16: MPC in C_4 plants. Cell-type specific transcript abundance of MPC genes among NADP-ME and NAD-ME subtype C_4 monocots. *Z. mays* data from Chang et al. (2012)[48]. *S. bicolor* data from Emms et al. (2016)[55]. *S. viridis* data from John et al. (2014)[56]. *P. virgatum* data from Rao et al. (2016)[59]. M = mesophyll. BS = bundle sheath. tpm = transcripts per million.

1.3.5 Mitochondrial dicarboxylate transport – the MC family

Dicarboxylate transport across the mitochondrial membrane is facilitated by several transporters from the Mitochondrial Carrier (MC) family. This is a large family of transporters, found across eukaryotes, with a common structure of six transmembrane domains, but highly varied substrate specificity[115]. In relation to C_4 photosynthesis, three

MC subfamilies have been implicated: the Dicarboxylate Carrier (DIC), Uncoupling Protein (UCP), and Dicarboxylate and Tricarboxylate Carrier (DTC) subfamilies.

DIC

DIC proteins carry out the exchange of dicarboxylates (including malate, OAA, and succinate), phosphate, sulphate and thiosulphate[116]. In plants, this transport activity was first observed in isolated bean mitochondria[117]. *A. thaliana* DIC1-3 were identified based on homology to the previously identified human DIC, and transport assays using AtDIC1-3 reconstituted in liposomes demonstrated that these proteins account for this transport activity[116]. Several potential roles for DIC transporters in C_3 plants have been proposed (Figure 1.17). First, DICs may facilitate the anaplerotic provision of dicarboxylates for the TCA cycle[116], which is consistent with the fact that malate and succinate oxidation in plant mitochondria is inhibited by known inhibitors of DIC[117,118]. Second, DICs may exchange malate and OAA, and act in concert with cytosolic and mitochondrial MDHs to transfer reducing equivalents between cytosol and mitochondria, in an analogous way to OMT1 in the chloroplast[116]. Third, high expression of *DIC1* and *DIC2* in cotyledons may indicate a role in the exchange of OAA and malate with succinate across the mitochondrial membrane that is required during the mobilisation of storage lipids via the glyoxylate cycle and gluconeogenesis[116].

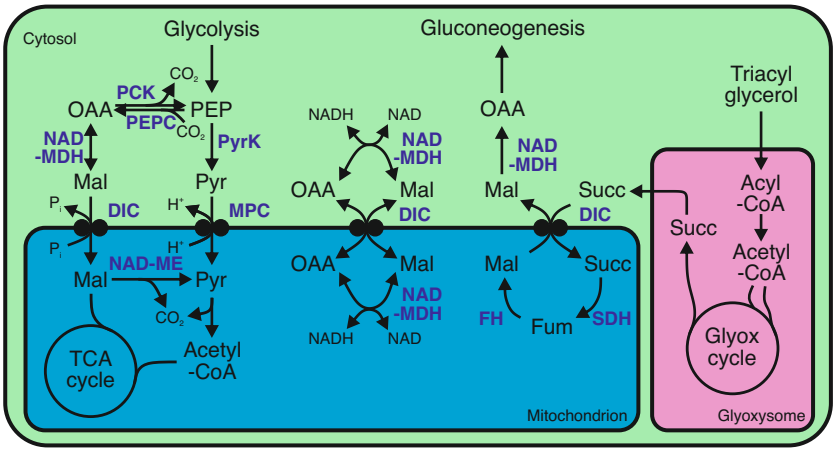


Figure 1.17: Possible functions of DIC in C_3 plants. Left: Anaplerotic provision of malate to TCA cycle. Centre: Exchange of reducing equivalents between cytosol and mitochondria. Right: Mitochondrial malate/succinate exchange in conjunction with glyoxylate cycle and gluconeogenesis. For abbreviations see appendix A.1, table A.1.

It was initially assumed that the NAD-ME subtype C_4 cycle required the mitochondrial import of only aspartate (the 4C transfer acid from the mesophyll) and 2-OG (required for the trans-amination reaction catalysed by AspAT to convert Aspartate into OAA). However, mitochondria isolated from the NAD-ME C_4 plant *Atriplex spongiosa* were found to require not only aspartate and 2-OG, but also malate and P_i, to sustain maximal rates of pyruvate and O_2 formation[119,120]. This process is also inhibited by a known inhibitor of DIC[117,119]. Thus, the exchange of malate and P_i by DIC seems to be required in NAD-ME subtype bundle sheath mitochondria. Transcriptomics data has subsequently supported a role for DIC in the NAD-ME cycle. Among C_4 monocots, two paralogs of *A. thaliana DIC3* are highly expressed, in a strongly bundle sheath-preferential manner in the NAD-ME subtype species *P. virgatum*[59], but not in NADP-ME subtype species[48,55,56] (Figure 1.18 A). In dicots, *DIC1* shows significantly bundle sheath-preferential expression

in the NAD-ME subtype C_4 species *C. gynandra*[114], and this gene is significantly upregulated compared to its close C_3 relative *C. spinoza*[58]. In contrast, no *DIC* genes show significant upregulation when comparing NADP-ME subtype C_4 species to close C_3 relatives within the *Flaveria* genus[66]. The expression patterns above for *DIC* correlate with the expression of genes encoding the mitochondrial Phosphate Carrier (PIC): a P_i/H^+ symporter[121]. This suggests a mechanism by which malate import (and P_i export) through DIC is coupled to the proton gradient across the mitochondrial membrane via H^+ and P_I symport through PIC (Figure 1.18 B).

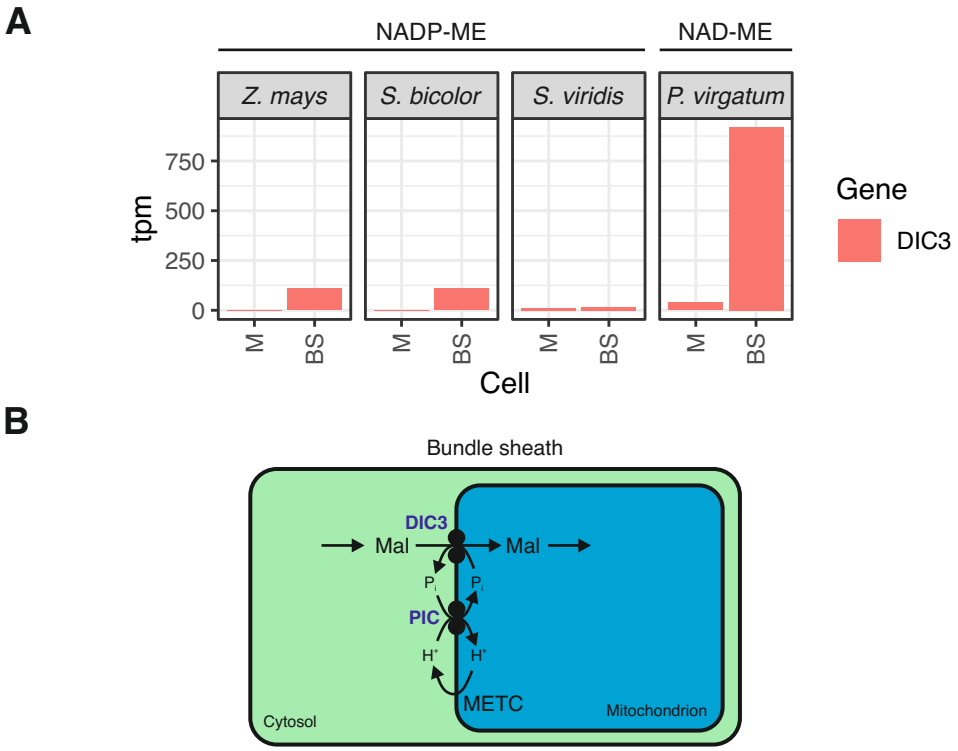


Figure 1.18: DIC in C_4 plants. **A**) Cell-type specific transcript abundance of DIC genes among NADP-ME and NAD-ME subtype C_4 monocots. *Z. mays* data from Chang et al. (2012)[48]. *S. bicolor* data from Emms et al. (2016)[55]. *S. viridis* data from John et al. (2014)[56]. *P. virgatum* data from Rao et al. (2016)[59]. M = mesophyll. BS = bundle sheath. tpm = transcripts per million. **B**) Function of DIC in C_4 plants. Metabolites are shown in black. Enzymes and transporters are shown in blue. For abbreviations see appendix A.1, table A.1.

DTC

DTC proteins catalyse the strict counter-exchange of dicarboxylates (including 2-OG, malate, OAA, succinate) and tricarboxylates (including citrate, isocitrate)[122,123]. DTC proteins are found only in plants and protozoa, although they are closely related to 2-OG/malate carriers found in animals, and were first identified in *P. miliaceum* based on homology to a previously characterised mammalian 2-OG/malate carrier[123]. A role in

NAD-ME subtype C_4 photosynthesis was immediately hypothesised, based on light-dependent, leaf bundle sheath-specific expression, and developmental regulation similar to other C_4 cycle genes[124]. This has subsequently been corroborated with transcriptomics data, with *DTC* genes showing high, bundle sheath-specific expression in NAD-ME[59], but not in NADP-ME subtype C_4 monocot species[48,55,56] (Figure 1.19). The NAD-ME subtype C_4 dicot species *C. gynandra* also shows significantly bundle sheath-preferential expression of *DTC*[114], although the gene is not upregulated in the NAD-ME subtype C_4 dicot species *C. gynandra* compared to its close C_3 relative *C. spinoza*[58].

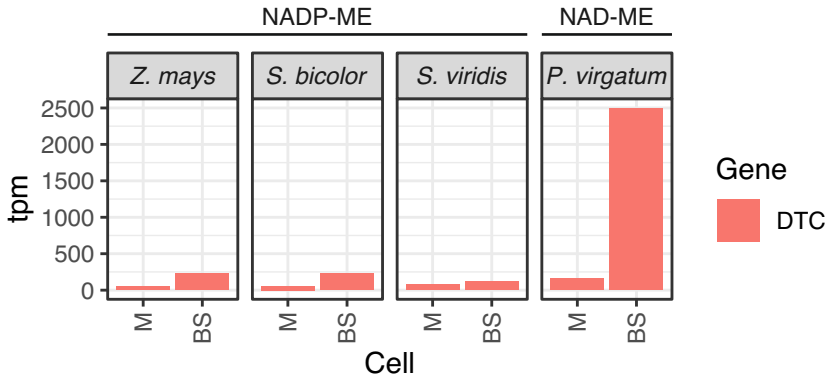


Figure 1.19: DTC in C_4 plants. Cell-type specific transcript abundance of DTC genes among NADP-ME and NAD-ME subtype C_4 monocots. *Z. mays* data from Chang et al. (2012)[48]. *S. bicolor* data from Emms et al. (2016)[55]. *S. viridis* data from John et al. (2014)[56]. *P. virgatum* data from Rao et al. (2016)[59]. M = mesophyll. BS = bundle sheath. tpm = transcripts per million

The role for DTC transporters in C_3 plants has not been well characterised, although multiple hypotheses have been proposed (Figure 1.20). First, the capacity for malate/OAA exchange means a redox valve function is possible[122]. Second, the capacity

for citrate exchange with OAA and/or malate could play a role in fatty acid elongation and isoprenoid synthesis[122]. The only hypothesis for which there is currently any experimental evidence is a potential role in nitrogen assimilation. Nitrogen assimilation via the GS/GOGAT cycle in the chloroplast requires the supply of 2-OG carbon skeletons. DTC, through its capacity to exchange various TCA cycle intermediates, may be involved in the direct export of 2-OG produced by mitochondrial isocitrate dehydrogenase to the cytosol, or the export of citrate that can then be converted to 2-OG by cytosolic aconitase and isocitrate dehydrogenase[125]. This is supported by the observation that *DTC* transcript levels in nitrogen-starved *Nicotiana tabacum* plants increased upon the addition of nitrate[122], in a similar pattern to that previously observed for isocitrate dehydrogenase transcripts[126,127].

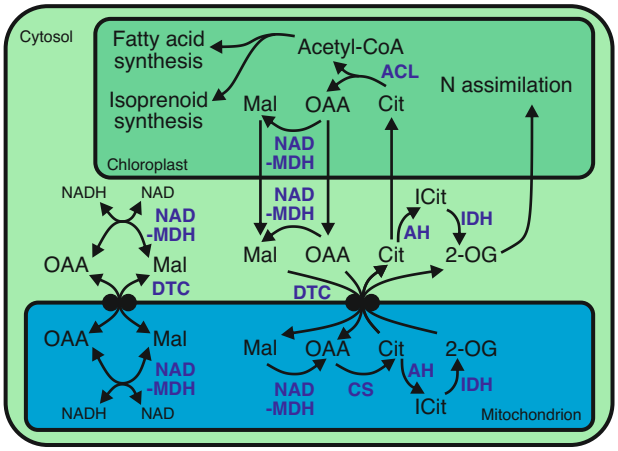


Figure 1.20: Possible functions of DTC in C_3 plants. Left: Exchange of reducing equivalents between cytosol and mitochondria. Right: Exchange of malate and/or oxaloacetate for citrate and 2-OG for chloroplastic fatty acid/isoprenoid synthesis and nitrogen assimilation respectively. For abbreviations see appendix A.1, table A.1.

UCP

UCPs are so-named because they were first identified based on their ability to transport H^+ in order to dissipate the proton gradient across the mitochondrial membrane and thus uncouple mitochondrial electron transport from ATP synthesis[128]. Whilst six family members were originally identified in *A. thaliana* (*UCP1-6*)[129], UCP4-6 were subsequently found to transport dicarboxylates, and to cluster with DICs from animals and yeast, and were therefore re-classified as DIC1-3 (discussed earlier)[116]. This leaves three UCP members in plants, UCP1-3, which cluster with UCPs from animals and yeast[116].

In plants, UCPs, sometimes called Plant Uncoupling Mitochondrial Proteins (PUMPs), were first identified in mitochondria isolated from potato[130]. This potato UCP, when reconstituted into liposomes, was found to transport H^+ [130]. Inhibiting its activity *in planta* led to a significant increase in mitochondrial H_2O_2 generation, suggesting a role in uncoupling mitochondrial electron transport in order to protect the mitochondria against oxidative stress[131]. A UCP was subsequently found in *A. thaliana*, where its expression was found to be cold-induced, indicating a role in uncoupling mitochondrial electron transport for the purpose of heat generation[132], similar to the thermogenic role played by UCPs in animals[133]. However, cold conditions can also induce oxidative stress[134], so induction of *UCP* expression by cold may actually be further evidence for a role in oxidative stress protection rather than thermogenesis[129]. Indeed, ubiquitous expression of *UCPs* in the arum lily, *Helicodiceros muscivorus*, and not just in the highly thermogenic floral organs, argues against a role for UCPs in thermogenesis[135]. Analysis

of *ucp1* mutant *A. thaliana* supported a role for UCPs in preventing oxidative stress, although not significantly under stress conditions such as low temperatures, and also identified a new role during photosynthesis[136]. Specifically, *ucp1* mutant plants showed significantly lower photosynthetic rates under photorespiratory conditions, and significantly reduced photorespiratory flux, as evidenced by a significant decrease in the rate of conversion of glycine to serine[136]. A proposed hypothesis for AtUCP1 function is therefore the uncoupling of mitochondrial electron transport from ATP synthesis under photorespiratory conditions. During photorespiration, some of the NADH produced by photorespiratory, mitochondrial GDC is shuttled to the peroxisome for use by photorespiratory hydroxypyruvate reductase. However, some is likely retained in the mitochondria and supplied to the mitochondrial electron transport chain. This is because, under high light conditions, a significant proportion of the reducing equivalents supplied to the peroxisome are thought to come from photosynthetic electron transport in the chloroplasts[47]. Supply of reducing equivalents to the mitochondrial electron transport chain will lead to the development of a proton gradient across the mitochondrial membrane. However, under these same conditions of high light, most of the cell's ATP requirement will be generated by photosynthetic electron transport in the chloroplasts, and thus the proton gradient across the mitochondrial membrane cannot be dissipated through ATPase alone. Instead, in order to prevent the backing-up of the mitochondrial electron transport chain, which could lead to oxidative stress, as well as inhibiting photorespiration by reduced consumption of NADH produced by photorespiratory GDC, the proton gradient can be dissipated through UCP1[136].

Interestingly, it was recently shown that AtUCP1 and 2, when reconstituted into liposomes, are able to exchange dicarboxylates (such as malate, OAA, 2-OG, succinate), amino acids (such as aspartate and glutamate), as well as other small compounds (such as sulfate, thiosulfate, and P_i)[137]. This prompted a re-investigation of *A. thaliana ucp1* mutant plants, as well as further investigation of *ucp2* single and *ucp1/ucp2* double mutant plants. GC-MS-based metabolic profiling showed that mutant plants showed reduced levels of the organic acids malate, fumarate, and citrate, compared to wild-type[137]. In light of this new biochemical and metabolic evidence, it is suggested that the photorespiratory role of AtUCP1/2 may be as part a redox valve to supply reducing equivalents from photorespiratory, mitochondrial GDC to photorespiratory, peroxisomal hydroxypyruvate reductase. It has already been suggested that reducing equivalents are transferred between mitochondria and peroxisome via a malate/aspartate shuttle, in which glutamate and malate are transferred from mitochondria to peroxisome, with aspartate and 2-OG moving in the opposite direction, and linked by MDH and AspAT in both cellular compartments[138]. Malate/2-OG exchange is likely facilitated by DTC, whilst aspartate/glutamate exchange could be carried out by UCP (Figure 1.21). Cleary further work must be done to fully elucidate the roles played by UCPs in C_3 plants, particularly with regards to photorespiration.

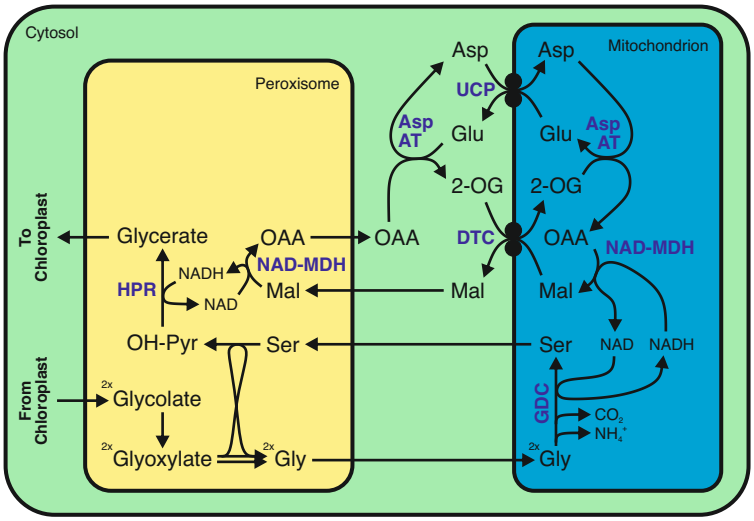


Figure 1.21: UCP in C_3 plants. Metabolites are shown in black. Enzymes and transporters are shown in blue. For abbreviations see appendix A.1, table A.1.

The discovery that UCPs can transport amino acids and dicarboxylates makes them clear candidates for co-option into the NAD-ME subtype C_4 cycle. Indeed, in dicots, *UCP1* shows bundle sheath-preferential expression in the NAD-ME subtype C_4 species *C. gynandra*[114], and this gene is significantly upregulated compared to its close C_3 relative *C. spinoza*[58]. In contrast, there is no consistent upregulation of *UCP* genes between C_3 and NADP-ME subtype C_4 *Flaveria* species[66]. In monocots, *UCP* genes show moderately high, bundle sheath-specific expression in the NAD-ME subtype C_4 species *P. virgatum*[59], and although *UCP* gene expression is also bundle sheath-preferential in NADP-ME subtype C_4 monocots, the expression level is much lower[48,55,56] (Figure 1.22).

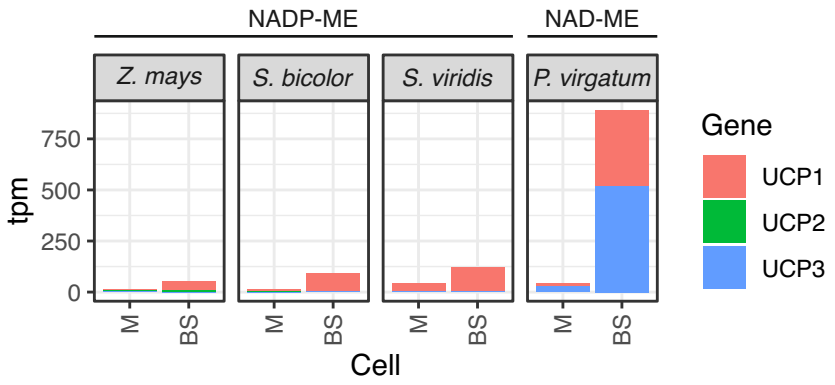


Figure 1.22: UCP in C_4 plants. Cell-type specific transcript abundance of UCP genes among NADP-ME and NAD-ME subtype C_4 monocots. *Z. mays* data from Chang et al. (2012)[48]. *S. bicolor* data from Emms et al. (2016)[55]. *S. viridis* data from John et al. (2014)[56]. *P. virgatum* data from Rao et al. (2016)[59]. M = mesophyll. BS = bundle sheath. tpm = transcripts per million.

Overall, the combination of MC family members currently identified in plants, their characterised transport activities, and their cell-specific expression levels across multiple C_4 species of different subtypes, suggests a mechanism for dicarboxylate transport across the mitochondrial membrane in NAD-ME subtype C_4 photosynthesis (Figure 1.23). This model was first proposed by Furbank et al., before any of the transporters involved had been identified, based on the observation that isolated *P. miliaceum* bundle sheath cells and chloroplasts carry out maximal O_2 evolution only in the presence of a combination of aspartate, 2-OG, malate and P_i[119]. The model was further supported by Taniguchi and Sugiyama after the initial characterisation of DTC in a C_4 plant[124]. Now the model can be fully described, including all transport steps. Briefly, UCP facilitates the import of aspartate in exchange for glutamate produced by mitochondrial AspAT-mediated transamination of the aspartate and 2-OG. 2-OG in turn

is imported by DTC, in exchange for malate export. Malate must be reimported through DIC in exchange for P_i export, driven by the P_i gradient generated by P_i:H^+ symport through PIC (Figure 1.23).

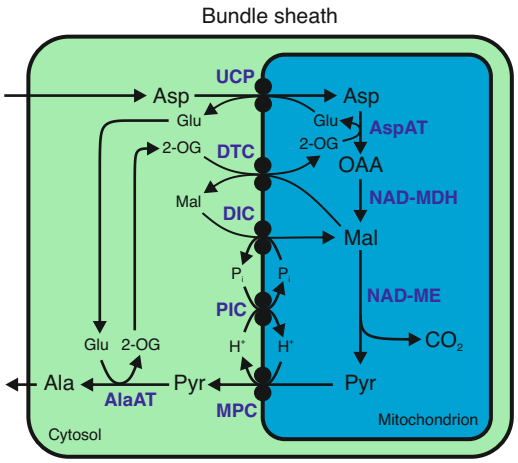


Figure 1.23: Hypothesis for MC (and MPC) family function in NAD-ME subtype C_4 plants. Metabolites are shown in black. Enzymes and transporters are shown in blue. For abbreviations see appendix A.1, table A.1.

1.4 Current gaps in our knowledge of NADP-ME subtype C_4 transporters

Despite over 50 years passing since the first identification of the transport of C_4 cycle metabolites across the chloroplast inner envelope[31], there are still several NADP-ME subtype C_4 cycle transport steps for which the corresponding metabolite transporter has not yet been identified and characterised (Figure 1.24).

1.4.1 H^+-dependent mesophyll chloroplast pyruvate import

As stated above, mesophyll chloroplast pyruvate import is known to occur by two mechanisms: an Na^+-dependent mechanism in the majority of C_4 plants, dependent on BASS2[99], and an H^+-dependent mechanism which has so far only been found in two tribes of grasses descended from the same C_4 origin[139], the Arundinelleae and Andropogoneae[94]. This suggests that all origins of NADP-ME subtype C_4 photosynthesis may have co-opted the Na^+:pyruvate symporter BASS2 to facilitate mesophyll chloroplast pyruvate import as part of the C_4 cycle, except for (at least) one origin of C_4 photosynthesis, which co-opted a separate, currently unidentified pyruvate uptake system, very likely an H^+:pyruvate symporter.

1.4.2 Bundle sheath chloroplast malate import and pyruvate export

As stated above, bundle sheath chloroplast malate import is proposed to be facilitated by members of the DCT subfamily, which is supported by experimental characterisation of maize *dct2* mutant plants[52]. However, the only efficient transport activity so far observed for DCT subfamily transporters is the exchange of four-carbon dicarboxylates (e.g. OAA, malate, succinate, fumarate, 2-OG) and four- and five-carbon amino acids (e.g. aspartate, glutamate)[28,30,140]. Significant dicarboxylate/amino acid unidirectional uptake was not been observed with DCTs from *A. thaliana* or maize[29,30], and DCT from C_3 spinach and C_4 *F. bidentis* were shown not to exchange malate and pyruvate[28]. Thus, whilst DCTs appear to be involved in the transport of dicarboxylates and amino acids across the bundle sheath chloroplast envelope during the C_4 cycle, they cannot facilitate

the net import of carbon that is required for the C_4 cycle to function as a carbon-concentrating mechanism. The existence of an additional aspartate transporter (likely a cation:aspartate symporter) was proposed to explain this[52], which is consistent with previous observations that malate-dependent pyruvate formation from wild-type maize bundle sheath cells or chloroplasts is stimulated by aspartate[53,54]. It is also plausible (and parsimonious) that there exists a separate transporter capable of either malate uptake (likely a cation:malate symporter) or simultaneous malate uptake and pyruvate efflux. Indeed, as stated above, it seems that cation:malate symport in the C_4 dicot *F. bidentis*, may be carried out by the Na^+:malate symporter BASS4 (Kuhnert et al., unpublished). However, this transporter is not highly expressed in the bundle sheath cells of NADP-ME subtype grasses[48,55,56], indicating a different transporter may have been co-opted for this role in monocots. In the absence of a transporter capable of carrying out malate import and pyruvate export, pyruvate export is presumably facilitated by a separate pyruvate exporter (likely a cation:pyruvate symporter).

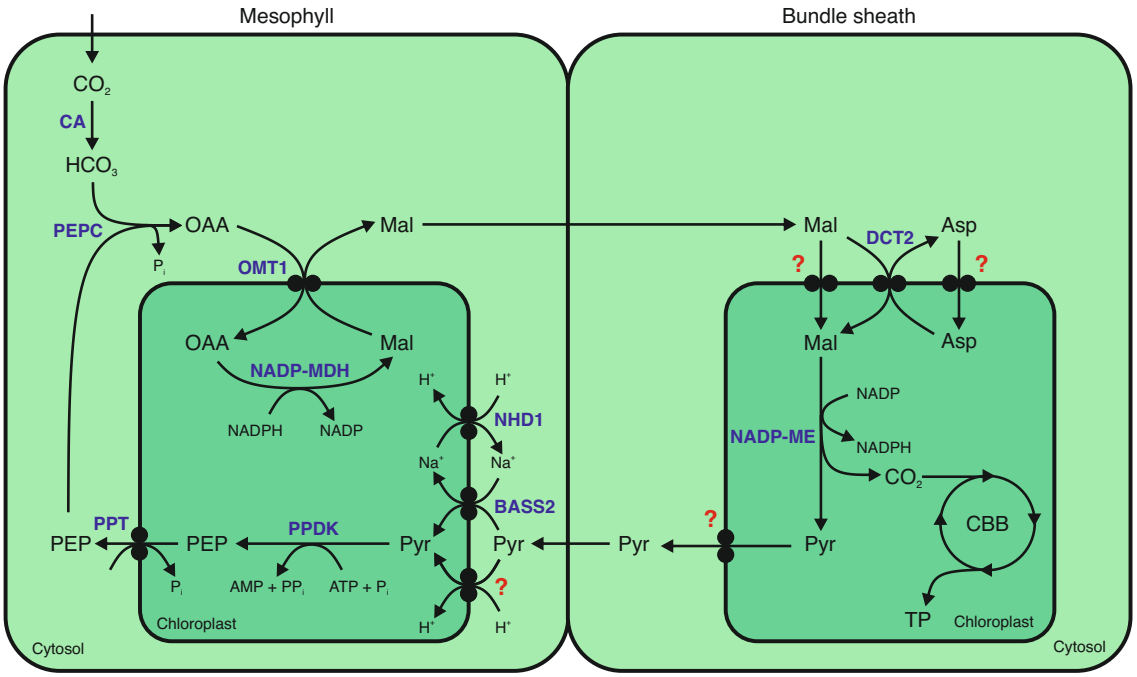


Figure 1.24: Transport steps missing from our understanding of the NADP-ME subtype C_4 cycle. Missing transporters are indicated with red question marks. These are 1) the import of dicarboxylates into the bundle sheath chloroplast (either directly via a single malate transporter, or an aspartate transporter working in concert with the dicarboxylate exchanger DCT2), 2) the export of pyruvate from the bundle sheath chloroplast, and 3) the H^+-dependent import of pyruvate into the mesophyll chloroplast. Metabolites and pathways are shown in black. Enzymes and transporters are shown in blue. For abbreviations see appendix A.1, table A.1.

Recent work by a previous PhD student in my lab identified a candidate pyruvate exporter, named HATCH1 (in homage to Hal Hatch, one of the co-discoverers of C_4 photosynthesis) in the C_4 model species *S. viridis*. Preliminary data also suggested a capacity for malate uptake by this transporter. The characterisation of this transporter is the primary goal of this thesis (chapters 2, 3, 5). However, the gene encoding this transporter is not highly expressed in the bundle sheath cells of maize or sorghum: C_4

plants from a separate origin of C_4 photosynthesis. This suggests that, as for pyruvate in the mesophyll chloroplast, a separate malate importer and pyruvate exporter exists in the bundle sheath chloroplasts of these species. An attempt to identify this transporter or transporters, as well as a transporter responsible for mesophyll chloroplast pyruvate import, is a secondary goal of this thesis (chapters 4, 5).

Chapter 2: Characterisation of HATCH1, a bundle sheath chloroplast malate importer and pyruvate exporter in *Setaria*

2.1 Chapter introduction and background

2.1.1 Missing transport steps of the C_4 cycle in *S. viridis*

C_4 photosynthesis is enabled by intracellular metabolite transport in both the mesophyll and bundle sheath cells. Many of the most widely grown C_4 crop species carry out NADP-ME subtype C_4 photosynthesis. However, a complete understanding of metabolite transport in NADP-ME subtype C_4 species is currently lacking. This chapter addresses this knowledge gap by characterising a novel chloroplast transporter responsible for malate import and pyruvate export in the model NADP-ME subtype C_4 grass species, *Setaria viridis.*

2.1.2 HATCH1 is a putative bundle sheath-localised C_4 cycle transporter in *S. viridis*

Previous bioinformatic work carried out by Steven Kelly and Basel Abu Jamous identified a putative chloroplast-localised transporter that was potentially involved in the C_4 cycle in *S. viridis*. Kelly and Abu Jamous used computational predictions for protein structural features (chloroplast transit peptides and hydrophobic transmembrane domains) as well as cell type-specific transcript abundance data (looking for bundle sheath-specific expression and elevated transcript abundance in the leaves of C_4 plants relative to C_3), to identify the uncharacterised gene Sevir.4G287300 as a potential candidate for being

involved in metabolite transport in the C_4 cycle. This gene was named *HATCH1* in homage to Hal Hatch one of the co-discoverers of C_4 photosynthesis. This gene belongs to an uncharacterised gene family which is defined by the presence of a UPF0114 PFAM domain (PF03350). A more in-depth analysis of this gene family is presented in chapter 5 of this thesis.

The amino acid sequence of HATCH1 is predicted to encode an N-terminal chloroplast transit peptide and four transmembrane domains (Figure 2.1 A). Whilst four transmembrane domains is less than other known chloroplast-localised C_4 cycle transporters (OMT1 and DCT2 = 13, PPT = 7, TPT = 8), HATCH1 could act as a homo-multimer. Consistent with a role in NADP-ME subtype C_4 photosynthesis, the relative abundance of *HATCH1* transcripts are significantly higher in leaves of NADP-ME subtype C_4 plants compared to C_3 relatives (Figure 2.1 B). In *S. viridis*, *HATCH1* transcripts also accumulate asymmetrically between mesophyll and bundle sheath cells, an expression pattern which is comparable to the other known C_4 cycle transporters in this species (Figure 2.1 C). Whilst all previously identified transporters involved in the *S. viridis* C_4 cycle are mesophyll-specific, *HATCH1* shows bundle sheath-specific expression, and is therefore a candidate malate importer and/or pyruvate exporter.

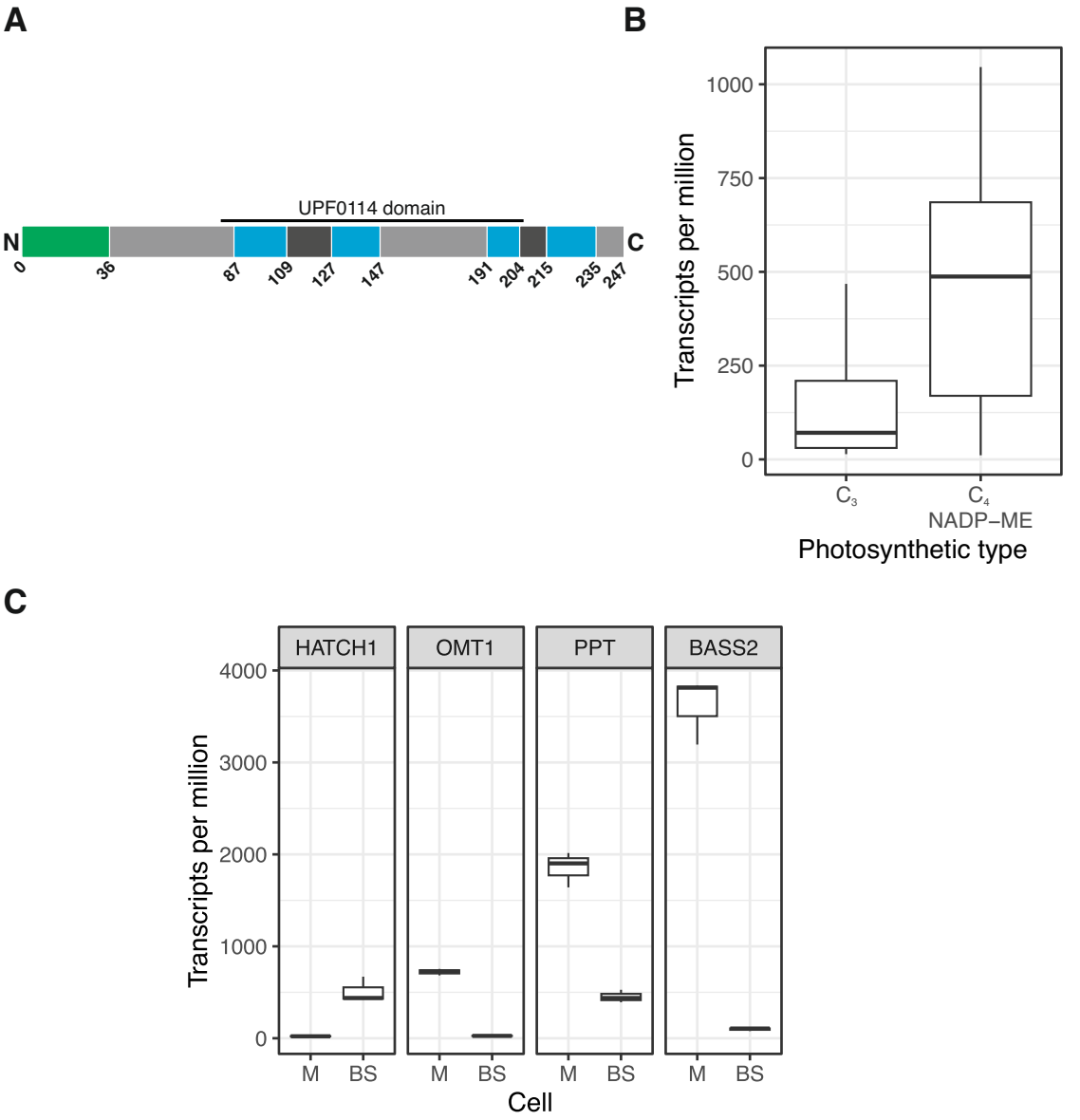


Figure 2.1: Bioinformatic evidence for putative involvement of HATCH1 in the C_4 cycle in *S. viridis*. **A**) Cartoon structure of HATCH1 protein. Colours represent different protein domains: Green = chloroplast transit peptide, Blue = transmembrane domain, Light grey = cytoplasmic domain, Dark grey = non-cytoplasmic domain. Conserved UPF0114 domain is indicated. Domain information obtained using Interpro[141], TargetP-2.0[142], and DeepTMHMM[143]. **B**) Transcript abundance of orthologs of *HATCH1* in NADP-ME subtype C_4 and closely related C_3 eudicot species. Average transcript abundance is higher in NADP-ME C_4 than C_3 species ($p < 0.01$, t-test). Data obtained from Kelly et al. (2017)[144]. **C**) Mesophyll (M) and bundle sheath (BS) cell-type specific transcript abundance for

genes encoding HATCH1 and other known C_4 cycle transporters (OMT1, BASS2, PPT) in *S. viridis*. Data obtained from John et al. (2014)[56].

2.1.3 A bacterial system to characterise chloroplast-localised malate and pyruvate transporters

Characterisation of transporter activity is typically carried out by expression and reconstitution of putative transporter proteins in liposomes, followed by observations of the uptake or efflux of radio-labelled metabolites. However, this process is relatively unsuited for the characterisation of transport activities involving pyruvate due to the ability of this metabolite to permeate lipid bilayers[99]. An alternative system is to exogenously express proteins in the inner envelope of bacteria, and observe any resulting changes in growth patterns, as well as metabolite exchange between the inside and outside of the cells. A number of K^+ transporters from various plants have been characterised in this manner[145–149], as well as a chloroplastic ATP/ADP transporter from *A. thaliana*[150]. There is also precedent for utilising this kind of bacterial expression system for the characterisation of plant proteins involved in the transport of monocarboxylates and dicarboxylates such as pyruvate and malate. For example, the plasma membrane-localised dicarboxylate transporter DCAT1 from Alder (*Alnus glutinosa*)[151] and the chloroplastic C_4 cycle pyruvate transporter BASS2[99] were characterised through expression in *E. coli* followed by measurement of metabolite uptake and, in the former case, by measurement of the effect on bacterial growth.

Malate and pyruvate are key metabolites involved in primary metabolism through the tricarboxylic acid (TCA) cycle in aerobic respiration, as well as serving as precursors for

the biosynthesis of other essential molecules such as amino and fatty acids (Figure 2.2). These metabolites are centrally positioned in metabolism in all domains of life, and are imported, exported and metabolised by a wide range of cells and organisms. Thus, a bacterial system such as *Escherichia coli,* which operates a functional TCA cycle and can be grown on malate[152] and/or pyruvate[153] as sole carbon sources is ideally suited for studying the activity of novel putative malate and/or pyruvate transporters from plants. Thus, this chapter describes experiments and results obtained from characterising the HATCH1 transporter expressed in *E. coli.*

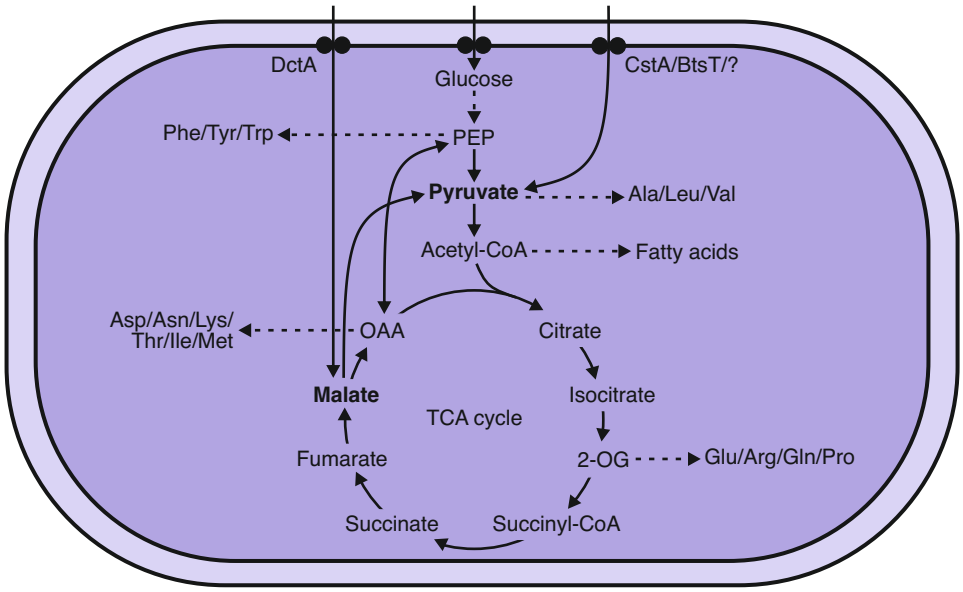


Figure 2.2: Diagram of *E. coli* central metabolism, indicating the main flux pathways through malate and pyruvate in normal aerobic growth.

2.1.4 HATCH1 has putative malate import and pyruvate export activity

Work by a previous member of the Kelly lab, Michael Niklaus, comprised an initial characterisation of HATCH1 through inducible, exogenous expression of the ortholog of *HATCH1* from the extremely closely related *Setaria italica* (Seita.4G275500) in *E. coli*. The amino acid sequence of the *HATCH1* gene is 100% identical in both species. When *HATCH1* expression was induced in *E. coli* strain BW25113 grown in minimal media with glucose as a sole carbon source, this led to an immediate cessation of growth, indicating that the presence of HATCH1 protein was deleterious for the growth of the cells. Analysis of the supernatant from these growth-arrested cell cultures showed that large amounts of pyruvate were exported from the cells. Metabolite mass spectrometry revealed that pyruvate export was specific and that only small amounts of some 4-carbon dicarboxylates (succinate, malate, fumarate) and the 3-carbon monocarboxylate 3-PGA, could be found in the supernatant of induced cells when compared to non-induced controls (Figure 2.3 A). Thus, HATCH1 facilitates export of pyruvate, but also potentially other monocarboxylates and dicarboxylates to a lesser extent.

Given that HATCH1 appeared to be somewhat permissive to the export of 4-carbon dicarboxylates it was also tested whether HATCH1 could facilitate malate import. To do this *HATCH1* was expressed in the *ΔdctA E. coli* strain. DctA is the sole transporter system in *E. coli* for the uptake of malate under aerobic conditions[154], and *ΔdctA* cells are therefore unable to grow in minimal media containing malate as a sole carbon source[155]. When *HATCH1* expression was induced in *ΔdctA E. coli* grown in minimal media with malate as a sole carbon source cells did not grow. However, pyruvate accumulated in

the culture medium to similar levels to cells that had glucose as a sole carbon source (Figure 2.3 B). It was calculated that the total mass of pyruvate exported from the *ΔdctA* cells expressing *HATCH1* when grown on malate was ~1/3rd of the total dry weight of the cells present: a value which is too high to be accounted for by a pre-existing pyruvate pool, or by other intracellular metabolites which could be converted into pyruvate for export. Thus, the most parsimonious explanation for these observations is that HATCH1 also facilitates malate import. Specifically, imported malate is likely converted to pyruvate via malic enzyme or via PCK, and is exported again from the cell by HATCH1.

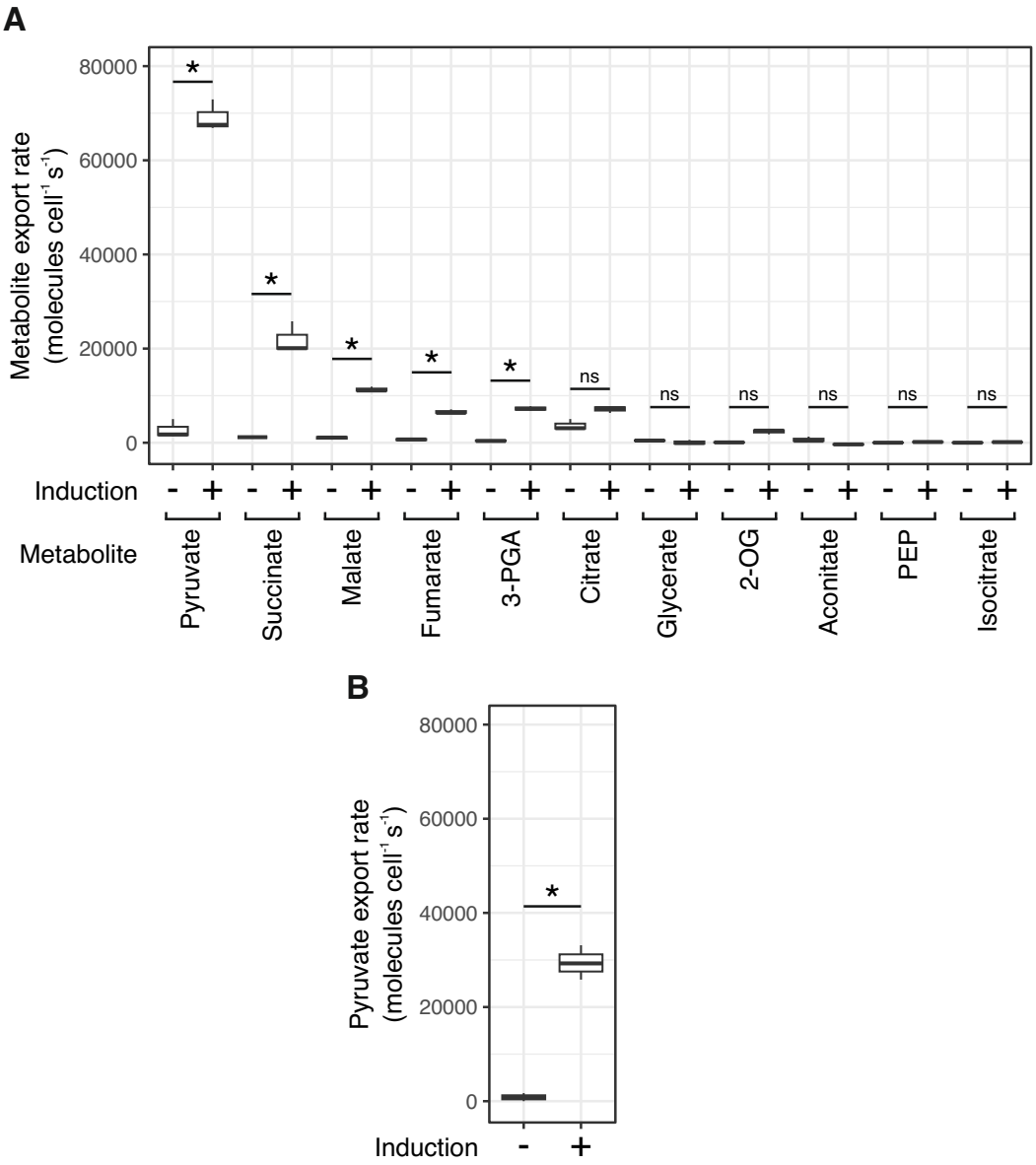


Figure 2.3: Transport properties of *S. italica* HATCH1 upon exogenous expression in *E. coli*. **A**) Metabolite profiling of cell culture supernatant of BW25113 *E. coli* after induction (+) or not (-) of *HATCH1* expression, on minimal media containing glucose as a sole carbon source. **B**) Measurement of pyruvate concentration in cell culture supernatant of *ΔdctA E. coli* after induction (+) or not (-) of *HATCH1* expression, on minimal media containing malate as a sole carbon source. Asterisks indicate significant differences ($p < 0.05$) between induced and non-induced samples using t-tests with Benjamini-Hochberg correction for multiple comparisons as appropriate. n.s. indicates not significant. Data collected by Michael Niklaus.

2.1.5 RNAi-mediated *HATCH1* knock-down *S. viridis* show photosynthetic defects

In order to test whether HATCH1 facilitates malate/pyruvate transport as part of a role in C_4 photosynthesis, Michael Niklaus also generated RNAi-mediated *HATCH1* knock-down *S. viridis* lines. Despite showing ~90% reduction in *HATCH1* transcript levels (Figure 2.4 A), these plants did not show any measurable growth defects under normal growth conditions. Although the plants did not show a growth defect, they showed reduced maximal photosynthetic rates in both CO_2 assimilation (Figure 2.4 B) and light response (Figure 2.4 C) curves. However, this phenotype proved highly variable, and it was hypothesised that this variability was due to variation in the efficacy of knock-down between plants and between leaves on individual plants.

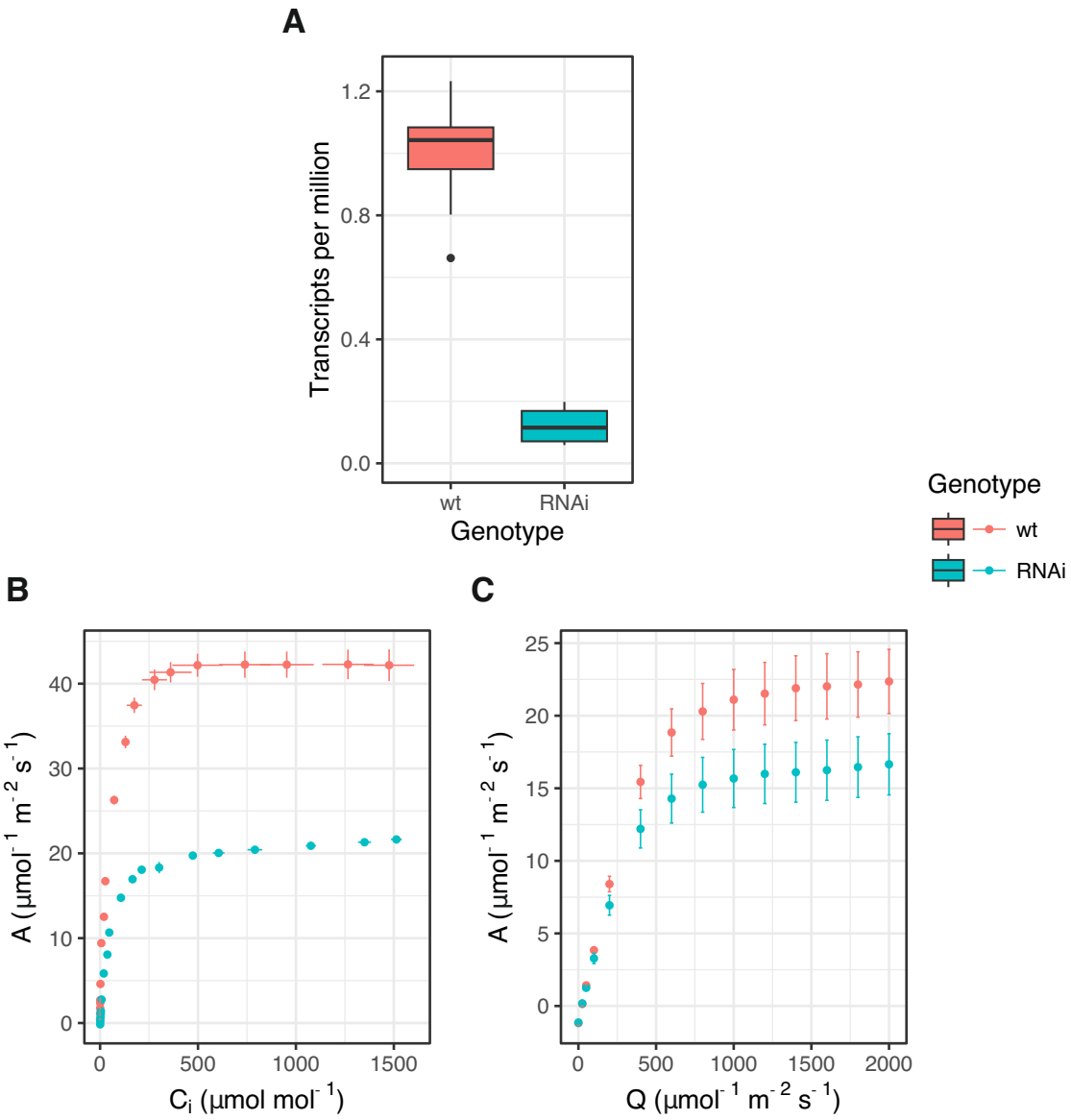


Figure 2.4: Photosynthetic gas exchange defects of *HATCH1* knock-down *S. viridis* plants. **A**) Relative *HATCH1* transcript abundance between *HATCH1* knock-down (RNAi) and segregating wild-type (wt) lines measured by RT-qPCR. **B**) CO_2 response and **C**) Light response curves of *HATCH1* knock-down (RNAi) and segregating wild-type (wt) *S. viridis* lines. Error bars indicate standard error of the mean. Data collected by Michael Niklaus.

2.1.6 Chapter aims and outline

Building on the previous work described above, all subsequent work described below was conducted by me. The first aim of the work in this chapter was to further characterise the transport mechanism of HATCH1 using *E. coli*. Specifically, does HATCH1 carry out a strict counter-exchange of malate and pyruvate, or does it merely facilitate the import and export of these metabolites, acting more as a selective pore or channel? The second aim of the work in this chapter was to further characterise the RNAi-mediated *HATCH1* knock-down *S. viridis* lines, and to confirm that their photosynthetic defects were associated with a reduction in *in planta* malate/pyruvate transport activity.

2.2 Results and discussion

2.2.1 HATCH1 is localised to the chloroplast inner envelope

HATCH1 is predicted to have a chloroplast transit peptide, and the proteins encoded by the *A. thaliana* and *Z. mays* orthologs of *HATCH1* were previously identified in their respective chloroplast envelope proteomes[49,156]. In order to confirm the chloroplast envelope localisation of HATCH1, Michael Niklaus previously expressed a *C*-terminal green fluorescent protein (GFP) fusion of *S. italica* and *A. thaliana HATCH1* transiently in mesophyll protoplasts isolated from *S. viridis*. However, the results from this experiment were inconclusive, partially owing to the low efficiency of protoplast isolation and transformation. Therefore, I decided to try an alternative system to determine protein localisation: transient expression of *C*-terminal GFP fusion proteins in *Nicotiana*

benthamiana leaves by agroinfiltration, followed by direct confocal fluorescence microscopy on leaf segments.

Surprisingly, the *S. italica* HATCH1-GFP fusion did not localise to the chloroplast, as expected, but instead showed a reticulate fluorescence pattern indicative of mis-localisation to the endoplasmic reticulum (ER) (Figure 2.5 A). A similar pattern of mis-localisation was observed for two positive controls of transiently expressed *C*-terminal GFP fusions of *S. italica* – *the* Triose Phosphate/P_i Translocator (TPT) (Figure 2.5 B) and Bile Acid:Sodium Symporter family protein 2 (BASS2) (Figure 2.5 C). These two proteins were chosen as positive controls as they are both experimentally validated chloroplast envelope-localised transporters[99].

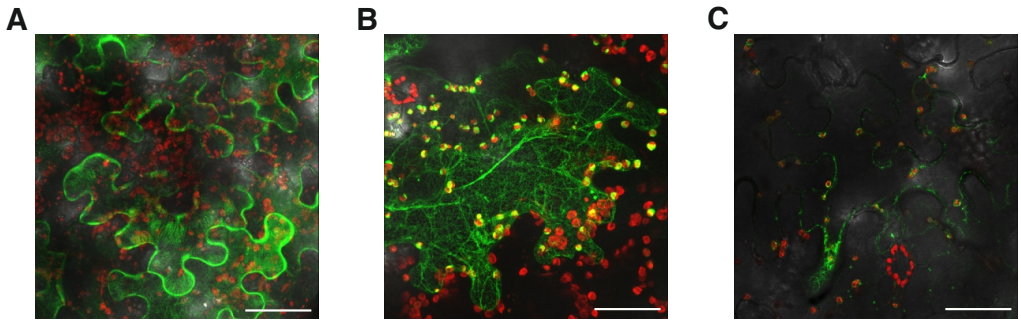


Figure 2.5: Confocal microscopy images showing the subcellular localisation of GFP-tagged *S. italica* proteins overexpressed in *N. benthamiana* leaves. Representative wide-field, maximum intensity projection images are shown for leaves overexpressing GFP-tagged versions of **A**) SiHATCH1, and two positive controls **B**) SiTPT and **C**) SiBASS2. GFP fluorescence = green, Chlorophyll autofluorescence = red. Scale bars represent 50 μm.

There are multiple possible explanations for mis-localisation of GFP-tagged proteins in this transient expression system. First, the fusion of GFP, a 26.9 kDa protein, to the *C*-terminus of these proteins could interfere with protein folding, multimerisation and/or localisation. This could be particularly problematic for a comparatively small protein such as HATCH1, which, at 27.05 kDa, is only marginally bigger than GFP itself. Second, whilst having broadly similar physical properties, chloroplast transit peptides from monocots and dicots have significantly different amino acid compositions[157]. It is therefore possible that transit peptides from the monocot *S. italica* do not function properly in the eudicot *N. benthamiana*, resulting in the mis-localisation of these proteins in the context of the transient expression system described above. Given that the two positive controls of experimentally verified chloroplast transporters also failed to localise correctly to the chloroplast inner envelope it is most parsimonious that the later scenario was the cause of the mis-localisation.

In order to confirm the localisation of HATCH1 using the *N. benthamiana* system, it was therefore decided to express a *C*-terminal GFP fusion of *A. thaliana HATCH1*. *A. thaliana* and *N. benthamiana* are both eudicots, so *A. thaliana* proteins were predicted to have a greater chance of localising correctly when expressed in *N. benthamiana*. Moreover, experiments conducted by Michael Niklaus above demonstrated that *A. thaliana* HATCH1 has identical transport properties to *S. viridis* HATCH1 when expressed in *E. coli*. As expected, the *A. thaliana* HATCH1-GFP localised to the chloroplast, showing a punctate localisation around the periphery of the chloroplast (Figure 2.6 A). A similar pattern was observed for the positive control *C*-terminal GFP fusion of *A. thaliana* TPT (Figure 2.6 B), but not in the negative control of cytosolic GFP (Figure 2.6 C). This is

consistent with previous observations that fluorescently-tagged chloroplast inner envelope proteins form distinct, punctate or crescent moon-like structures around the edges of the chloroplast, as opposed to outer envelope proteins, which show more continuous fluorescence around the entire circumference of the chloroplast[158,159].

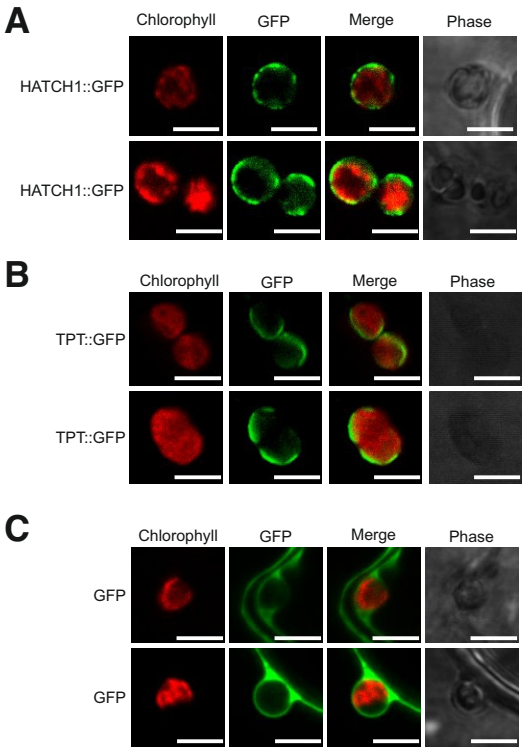


Figure 2.6: Confocal microscopy images showing the subcellular localisation of GFP-tagged *A. thaliana* proteins overexpressed in *N. benthamiana* leaves. Representative images of two chloroplasts each are shown for leaves overexpressing GFP-tagged versions of **A**) AtHATCH1, **B**) AtTPT (positive control), **C**) cytosolic GFP (negative control). Scale bars represent 5 μm.

Overall, despite mis-localisation of *S. italica* HATCH1 (and two positive controls) in the *N. benthamiana* expression system, the clear localisation of *A. thaliana* HATCH1 to the chloroplast inner envelope, in line with bioinformatic prediction and proteomic evidence in Arabidopsis and maize, strongly indicates that HATCH1 is localised to the chloroplast inner envelope.

2.2.2 HATCH1 facilitates selective dicarboxylate import

As it was already shown that HATCH1 is a selective pyruvate exporter, and capable of facilitating malate import, it was necessary to determine whether HATCH1 import was similarly selective for malate, or whether a wider range of dicarboxylates are accepted as import substrates. We sought to answer this question through further application of the previously described *E. coli* exogenous expression system. We first tried to demonstrate metabolite import through the uptake of radio-labelled dicarboxylates into *E. coli* cells. However, it might be predicted that accumulation of significant radio-label within *E. coli* cells expressing HATCH1 would not be observed, due to the fact that any malate imported through HATCH1 would be immediately converted to CO_2, which would diffuse out of the cell, and pyruvate, which would be exported through HATCH1 itself. Indeed, this is what was observed when preliminary 14C labelled malate uptake assays were performed. We therefore sought to instead use pyruvate export rate as a proxy for the import of dicarboxylates through HATCH1.

E. coli is able to use a wide range of dicarboxylates, including malate, fumarate, succinate, and aspartate, as substrates for aerobic growth[154]. The primary uptake

system for these dicarboxylates under aerobic conditions and neutral pH is DctA[154,155]. Subsequently, these dicarboxylates either directly enter into the tricarboxylic acid (TCA) cycle, or are metabolised via the PEP/pyruvate bypass to produce acetyl-CoA so that the TCA cycle can be sustained (see figure 2.2)[154]. Thus, if HATCH1 is capable of importing a given dicarboxylate, it would be expected that *ΔdctA* mutant *E. coli* expressing *HATCH1* should take up the dicarboxylate through HATCH1 specifically, convert this to pyruvate through the PEP/pyruvate bypass, and subsequently export pyruvate back through HATCH1 at a higher rate than in the absence of the dicarboxylate.

To test this, *HATCH1* expression was first induced in cells grown on minimal media containing glucose as a sole carbon source, before being transferred to media containing either just glucose, or both glucose and one of the four aforementioned dicarboxylates (malate, fumarate, succinate, and aspartate), as sole carbon sources. Supernatant samples were subsequently taken and tested for the presence of pyruvate, from which pyruvate export rates could be calculated (Figure 2.7). As had been previously observed, pyruvate was exported in the presence of glucose as the sole carbon source (Ø: 7.7 x 10^4 molecules pyruvate $cell^{-1}$ s^{-1} ± 3.2 x 10^4 (SD, N = 18)). Also as expected, compared to glucose-only samples, the addition of malate led to a large increase in pyruvate export rate (Ø: 15 x 10^4 molecules pyruvate $cell^{-1}$ s^{-1} ± 3.6 x 10^4 (SD, N = 9)). This difference is significant according to post-hoc comparison using Dunnett's test ($p<0.001$), and is consistent with previous observations that HATCH1 facilitates malate import. The addition of succinate led to a smaller increase in pyruvate export rate (Ø: 11 x 10^4 molecules pyruvate $cell^{-1}$ s^{-1} ± 2112 (SD, N = 3)), which was not significant according to Dunnett's test ($p=0.312$). No significant increase in pyruvate export rate was observed

with the addition of either aspartate (Ø: 55187 molecules pyruvate cell^{-1} s^{-1} ± 9740 (SD, N = 3)) or fumarate (Ø: 48068 molecules pyruvate cell^{-1} s^{-1} ± 28897 (SD, N = 3)).

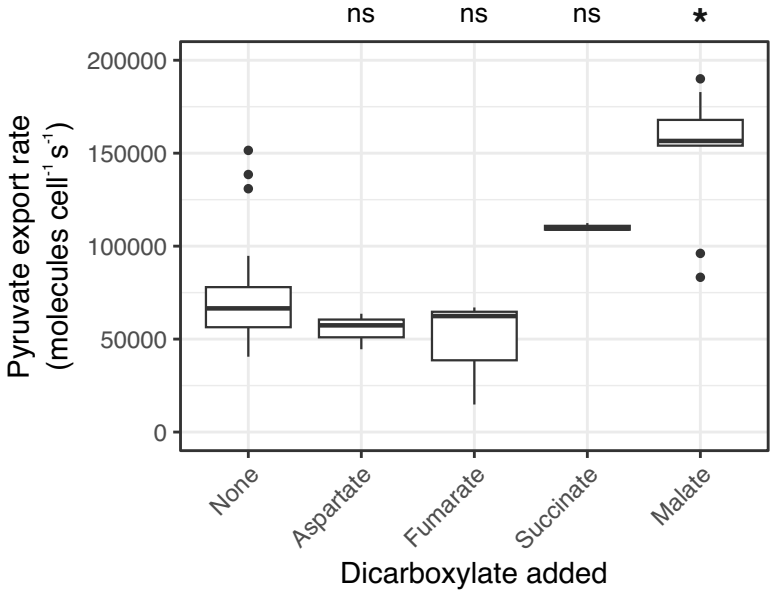


Figure 2.7: Pyruvate efflux experiment in *ΔdctA E. coli* to determine import specificity of exogenously expressed *S. italica* HATCH1 with regard to various 4-carbon dicarboxylates. Supernatant samples were taken 15 minutes after resuspension of induced cultures in minimal media containing either glucose or glucose plus the indicated dicarboxylate. Asterisk indicates significant difference ($p \leq 0.05$) in pyruvate export rate for cultures with dicarboxylate added compared to those without according to post-hoc Dunnett's test. ns = not significant.

It should be noted that increased pyruvate export in the presence of malate and succinate could be attributed to a couple of factors other than more efficient import of these dicarboxylates through HATCH1. First, although the major succinate uptake system under neutral pH (as used in this experiment) is DctA, succinate uptake can also

occur by other transporters DauA and SatP, although these transporters generally function at lower pHs[154,160,161]. This could explain why succinate caused a slight increase in pyruvate export rate, even though this difference was not significant. Second, some dicarboxylates may be more efficiently converted to pyruvate within the *E. coli* cell, although it would be expected that all these dicarboxylates that feed into the TCA cycle would be converted to pyruvate with approximately equal efficiency. However, the consistency between this experiment, and the previously generated metabolite export profiles of *E. coli* expressing HATCH1, specifically the comparatively high export rates of malate and succinate (see figure 2.3 A), lend support to the conclusion that HATCH1 is selective in which dicarboxylates it transports in either direction, with preference for malate and succinate.

2.2.3 HATCH1 facilitates pyruvate export in the absence of external malate

The C_4 cycle in *S. viridis* requires a 1:1 ratio of malate import to pyruvate export across the bundle sheath chloroplast envelope. Given the ability of HATCH1 to import malate and export pyruvate, we wanted to investigate whether HATCH1 is an antiporter that facilitates a strict malate/pyruvate counter-exchange reaction, or whether it functions more like a pore or a channel in facilitating the import of malate and export of pyruvate down their respective concentration gradients. Evidence against the antiporter hypothesis is given by the fact that export of cellular pyruvate is caused by induction of *HATCH1* expression in *E. coli* grown on minimal media containing glucose as a sole carbon source, i.e. in the absence of externally supplied malate (see figures 2.3 A and

2.7). However, we hypothesised that under these conditions, the bacteria could be converting glucose into dicarboxylates, and these could be leaking out into the periplasm through transporters other than the knocked-out DctA. Specifically, the Dcu family of transporters primarily facilitate the exchange of dicarboxylates such as fumarate and succinate during fumarate respiration under anaerobic conditions[154]. These antiporters may leak dicarboxylates into the periplasm, and DcuC is proposed to facilitate succinate efflux[162]. Even a slow rate of dicarboxylate efflux into the periplasm could be sufficient to stimulate pyruvate efflux via HATCH1, assuming an antiport mechanism for this transporter. In order to eliminate this hypothesis, an *E. coli* line containing a quadruple knock-out in *DctA*, *DcuA*, *DcuB*, and *DcuC* was obtained: IMW213[163], along with its corresponding wild-type strain: AN387[164].

To verify that the obtained *E. coli* IMW213 strain did indeed have mutations in all four genes, the whole genomes of IMW213 and AN387 were sequenced. The paired-ends reads for each strain were mapped against the wild-type *E. coli* K12 genome, and visualised (Figure 2.8 A). As expected, AN387 showed no disruption in any of the four genes. IMW213 showed clear disruption of *DcuA* and *DcuB*, with no reads from IMW213 mapping to whole or part of the gene respectively, indicating the loss of these gene regions and a potential insertion at this location. In contrast, IMW213 reads mapped to the whole length *DctA* and *DcuC*. However, they each contained one break, where mapped reads did not cross, suggesting the insertion of an extra DNA sequence at this location. Indeed, by blasting the sequences of the genes in question against the IMW213 genome assembled from the sequencing reads, it was confirmed that all four genes

sequences had been disrupted by the insertion of an antibiotic resistance gene and/or transposable element, and thus were true knock-outs.

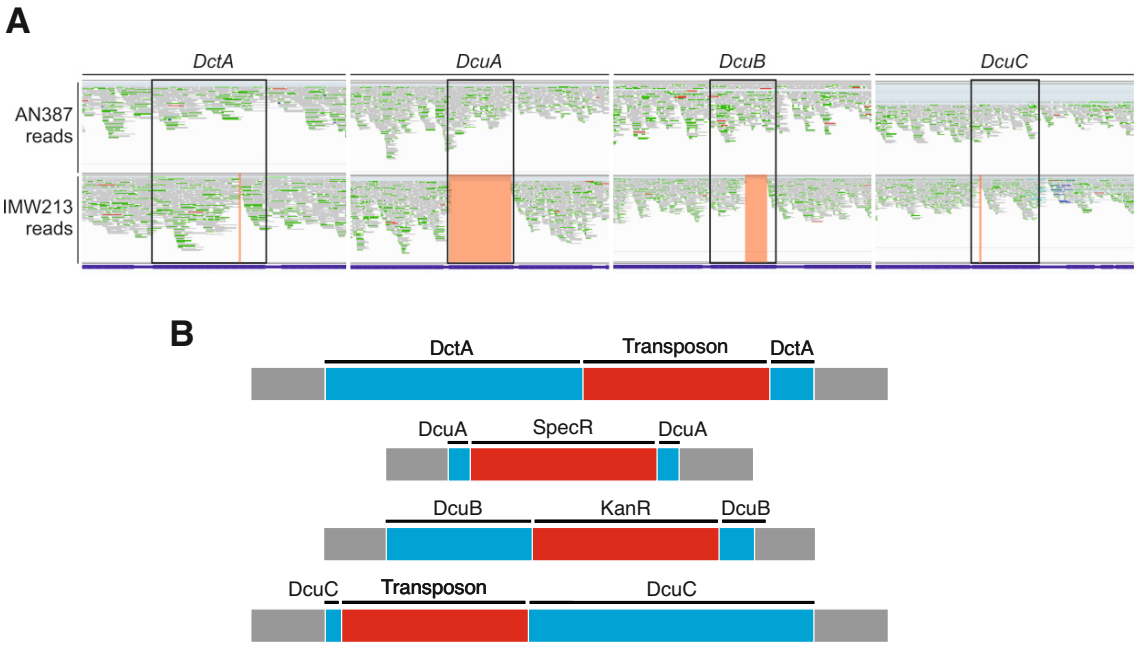


Figure 2.8: Whole genome sequencing of AN387 (wild-type) and IMW213 (*ΔdctAΔdcuAΔdcuBΔdcuC*) *E. coli* lines. **A**) The top and bottom panels show reads from AN387 and IMW213 mapped against the genome of another *E. coli* wild-type strain, K12. Orange boxes show locations where no reads map to the K12 genome, indicating a gene-disrupting insertion. **B**) Cartoon representations of the four genes in question showing inserted sequences. Blue = gene of interest. Red = disruptive element.

As the quadruple knock-out strain should not leak dicarboxylates into the cytoplasm, this system can be used to test the mechanism of HATCH1 activity. If IMW213 cells expressing *HATCH1*, grown on minimal media containing glucose as a sole carbon source, do not export pyruvate, this would indicate that HATCH1 is a strict malate/pyruvate antiporter, as the absence of leaked malate is preventing

malate/pyruvate counter-exchange through HATCH1. Alternatively, if these cells do export pyruvate, this would indicate that pyruvate export can occur in the absence of malate import, and therefore that HATCH1 does not operate by a strict antiport mechanism. Of these two alternate outcomes, the latter was found to be the case (Figure 2.9), supporting the conclusion that malate import and pyruvate export through HATCH1 are not strictly coupled, and thus that HATCH1 acts as a selective channel rather than a strict antiporter.

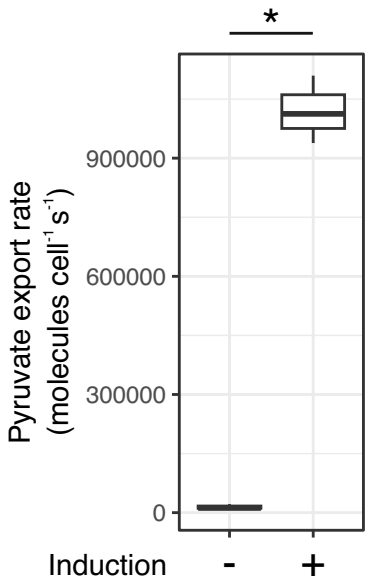


Figure 2.9: Pyruvate efflux experiment in *ΔdctAΔdcuAΔdcuBΔdcuC E. coli* to demonstrate pyruvate export in the absence of malate import. Supernatant samples were taken 3 hours after addition (+) or not (-) of inducer, on minimal media containing glucose as a sole carbon source. Asterisk indicates significant difference ($p \leq 0.01$) in pyruvate export rate between induced and non-induced cultures (t-test, p=0.00231).

2.2.4 Computational prediction of HATCH1 complex structure

Elucidating the structure of a transporter can also shed light on the transport mechanism. We have recently begun a collaboration with Dr Jani Bolla, who is attempting to over-express tagged versions of *HATCH1* and its *E. coli* ortholog, *YqhA*, in bacteria, for purification and subsequent experimental determination of transporter structure. Mass spectrometry will be used to determine the number of subunits in a potential multimeric HATCH1 complex, and subsequently, if possible, a protein structure will be obtained.

Whilst it still remains the case that the structure of proteins and protein complexes can only be fully known experimentally, in the past two years it has become possible to generate putative structural predictions using Google DeepMind's AlphaFold2[165]. Even though biological membranes are not explicitly represented in the model, the neural network underpinning AlphaFold2 can accurately predict the structure of many transmembrane proteins[166]. AlphaFold2 can also predict the structure of some multimeric protein complexes with reasonable accuracy[167]. Thus, structural predictions for both the monomeric and several multimeric forms of multiple members of the HATCH1 family were computed to determine if they could provide insight into the transport mechanism of HATCH1

S. italica HATCH1 (with the predicted chloroplast transit peptide removed) is predicted to form four alpha helical domains, arranged in an approximately parallel row, and each connected by an unstructured loop (Figure 2.10 A). In agreement with the prediction of four transmembrane domains, the central region of each alpha helix consists primarily

of hydrophobic (non-polar) residues, which are expected to interact with the hydrophobic interior of the chloroplast envelope. Towards the ends of the helices, and in the unstructured loops, there is a higher density of hydrophilic (polar uncharged, acidic, and basic) residues, which are likely exposed to the chloroplast stroma and intermembrane space (Figure 2.10 B). This predicted structure scores highly by the two confidence metrics calculated by AlphaFold2: predicted Local Distance Difference Test (pLDDT) (a measure of confidence in the local structure of the protein/complex) (Figure 2.10 C), and Predicted Aligned Error (PAE) (a measure of confidence in the global structure of the protein/complex) (Figure 2.10 D). This is despite relatively low depth of the multiple sequence alignment generated at the start of the AlphaFold2 algorithm run on HATCH1 (Figure 2.10 E).

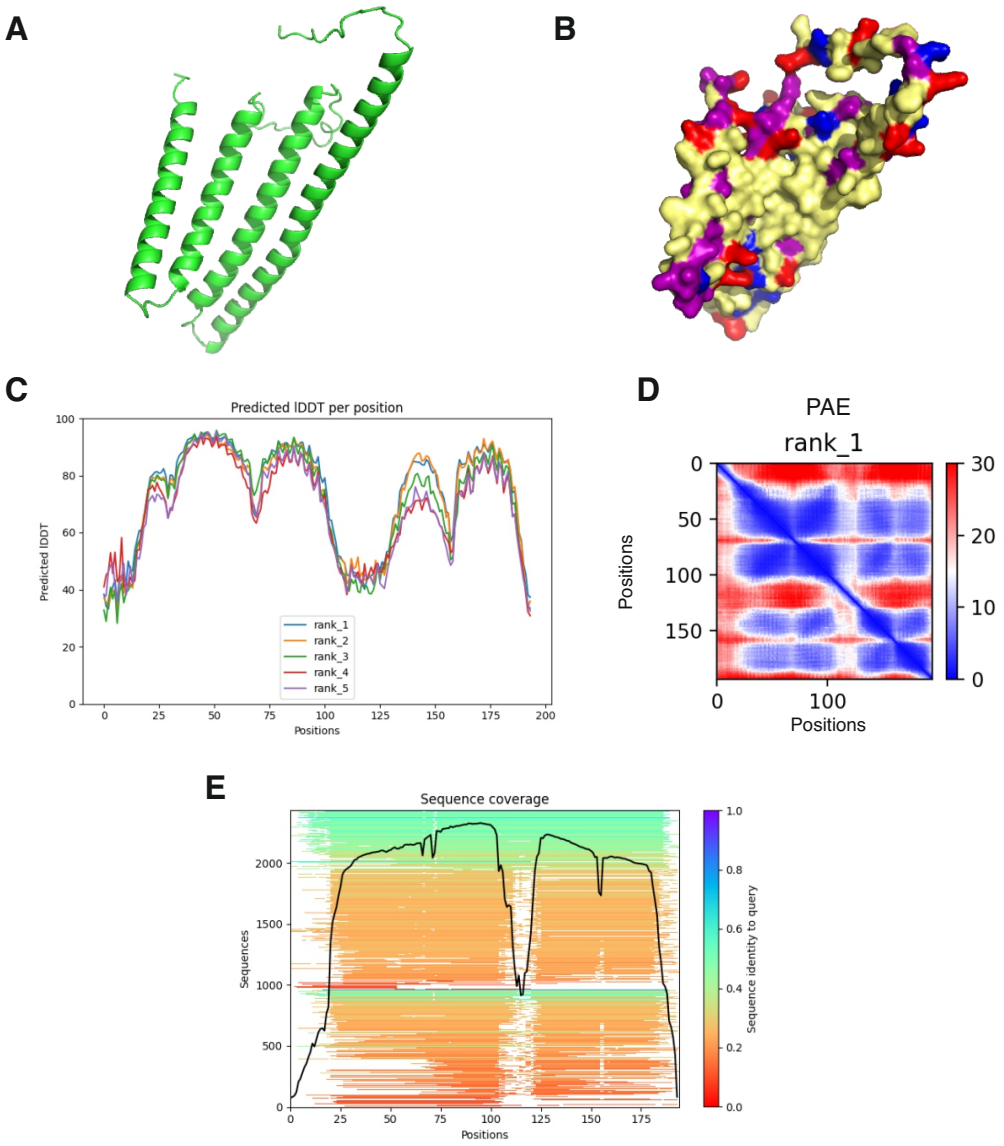


Figure 2.10: Structural prediction of the *S. italica* HATCH1 monomer using AlphaFold2. **A**) Cartoon representation, side view. **B**) Space-filling representation, side view. Colours indicate amino acid side-chain property: Yellow = hydrophobic, red = positively charged, purple = polar uncharged, blue = negatively charged. **C**) Graph showing predicted lDDT along the length of the protein. Rank_1 model is the relevant model. High plDDT indicates high confidence in the local structure of the prediction at that position. **D**) Graph showing PAE between all pairs of residues within the protein structure prediction. Low PAE (blue) indicates high confidence in relative position of the two residues. **E**) Graph showing alignment depth along the length of the protein during the initial MSA generation stage of the AlphaFold2 algorithm.

However, this structure does not appear suited to function as an intracellular metabolite transporter. Therefore, it is possible that HATCH1 functions as a multimer. Given that the protein expressed on its own can facilitate malate import and pyruvate export in *E. coli*, it was hypothesised that it acted as a homo-multimer as we hypothesised that it would be unlikely to form an interaction with a native *E. coli* membrane protein to provide this function. In the absence of experimental data to confirm the number of subunits in this hypothesised multimeric HATCH1 complex, we decided to generate predicted structures for increasing numbers of HATCH1 subunits. When multiple HATCH1 subunits are simultaneously subject to structural prediction they are predicted to interact along the length of the row of alpha helices. However, as each subunit is added, the confidence in the model declines, resulting in potentially erroneous predictions in which much of the subunit secondary structure is lost (Figure 2.11). This occurs due to the relatively low multiple sequence alignment depth, which limits the structural inferences that can be made about protein/complex structure.

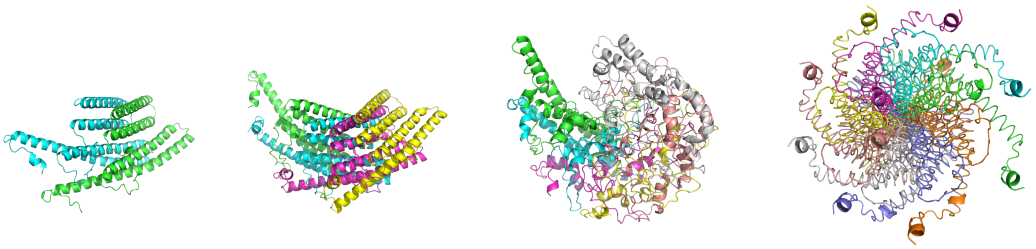


Figure 2.11: Structural prediction of *S. italica* HATCH1 multimers. Cartoon representation, top view. Increasing multimeric number from left to right (dimer, tetramer, hexamer, octamer).

In contrast, the coverage of the multiple sequence alignment generated when running AlphaFold2 on the *E. coli* ortholog of *HATCH1*, *YqhA* (see chapter 5), is much greater (Figure 2.12 E). This results in a very high confidence prediction (by both pLDDT and PAE metrics) of YqhA monomer structure (Figure 2.12 C, D). The predicted YqhA monomer structure is almost identical to that observed for HATCH1 (Figure 2.12 A, B).

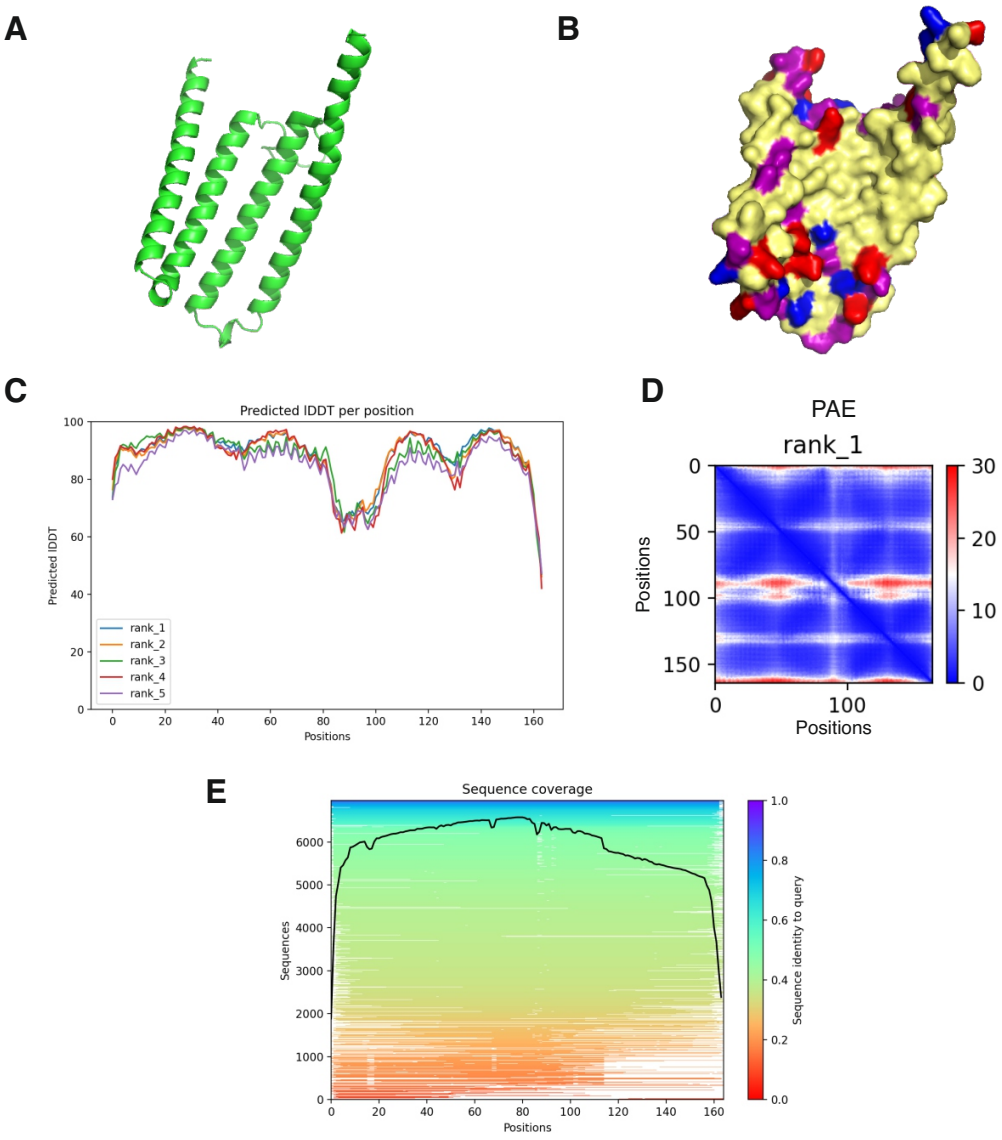


Figure 2.12: Structural prediction of the *E. coli* YqhA monomer using AlphaFold2. **A**) Cartoon representation, side view. **B**) Space-filling representation, side view. Colours indicate amino acid side-chain property: Yellow = hydrophobic, red = positively charged, purple = polar uncharged, blue = negatively charged. **C**) Graph showing predicted lDDT along the length of the protein. Rank_1 model is the relevant model. High plDDT indicates high confidence in the local structure of the prediction at that position. **D**) Graph showing PAE between all pairs of residues within the protein structure prediction. Low PAE (blue) indicates high confidence in relative position of the two residues. **E**) Graph showing alignment depth along the length of the protein during the initial MSA generation stage of the AlphaFold2 algorithm.

The increased prediction confidence allows for complex structure prediction with a larger number of YqhA subunits than was possible for HATCH1. As subunits are added, a curved complex structure is formed, and eventually, when eight subunits are reached, this curvature completes into a ring, to form a pore-like structure (Figure 2.13 A). Interestingly, the pore formed is approximately the right size to accommodate malate, succinate and pyruvate, with a positively charged ring of residues around one end that could attract these metabolites (Figure 2.13 B).

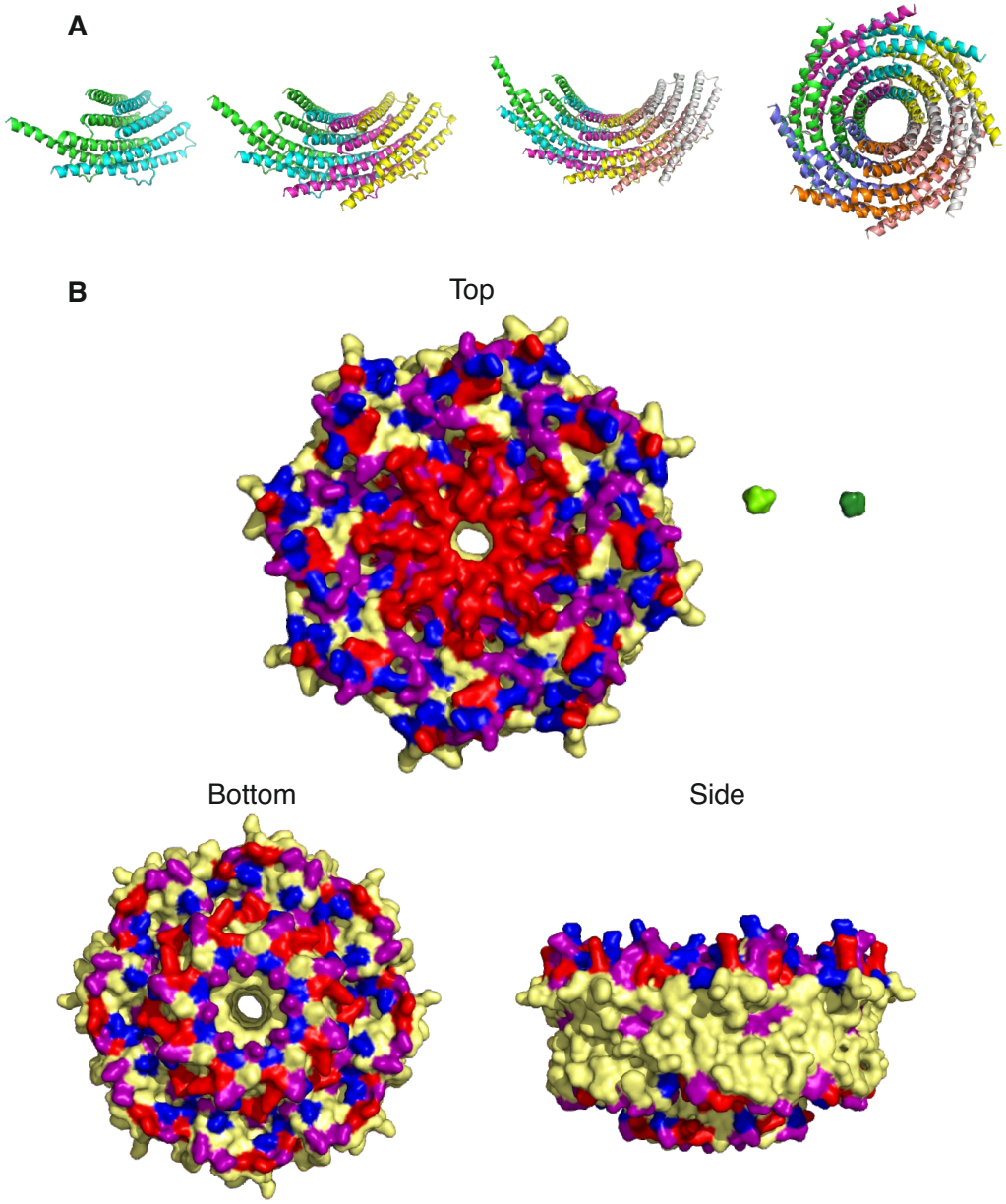


Figure 2.13: Structural prediction of *E. coli* YqhA multimers. **A**) Cartoon representation, top view. Increasing multimeric number from left to right (dimer, tetramer, hexamer, octamer). **B**) Space-filling representations, view angle indicated. Colours indicate amino acid side-chain property: Yellow = hydrophobic, red = positively charged, purple = polar uncharged, blue = negatively charged. For size comparison with the top view model, space-filling models of malate (light green) and pyruvate (dark green) are also shown.

Overall, the structural predictions presented above for HATCH1 and YqhA can only provide us with a hypothesis for the real HATCH1 complex structure, which awaits experimental validation. However, it is tempting to speculate that the predicted pore-like structure of the YqhA octomer, which is of approximately the right size to accommodate the metabolites that HATCH1 has been shown to transport, could explain the observed channel-like transport activity of the *Setaria* HATCH1 protein.

2.2.5 HATCH1 carries out malate/pyruvate transport *in planta*

The bioinformatic data (expression patterns and structural predictions) and experimental data (localisation studies, mechanistic characterisation in bacteria, and characterisation of knock-down plants) presented above support a role for HATCH1 in the C_4 cycle in *S. viridis*. However, we still wanted to confirm that the malate/pyruvate transport activity observed in bacteria also occurred *in planta*. To this end, we aimed to develop an assay to measure malate/pyruvate transport in isolated *S. viridis* bundle sheath strands or chloroplasts. These transport rates could then be compared between wild-type and *HATCH1* knock-down plants.

We initially attempted to isolate *S. viridis* bundle sheath chloroplasts using a modified version of a protocol developed by Jenkins and Boag for the isolation of maize bundle sheath chloroplasts[54]. Unfortunately, despite many repeated attempts we were unable to sufficiently optimise the protocol to isolate intact bundle sheath chloroplasts from *S. viridis* in sufficient quantity for our experiments. This was likely due to the suberised cell walls of the bundle sheath cells, a key component of C_4 anatomy, that renders bundle

sheath cell walls resistant to enzymatic digestion, and requires significant force for mechanical disruption. In all our attempts, disruption of the cell wall resulted in simultaneous disruption of the chloroplasts contained within the cells.

Given the difficulty in obtaining intact chloroplasts from bundle sheath cells, an alternative approach was also attempted. Here, we attempted to perform metabolite transport assays on isolated bundle sheath strands (Figure 2.14 A). Bundle sheath strands from a C_4 plant such as *S. viridis* are still a suitable material for the investigation of differences in chloroplast transport between plant lines, as their highly suberised cell walls (which prevent gentle cell disruption) ensure that the bundle sheath cells remain intact and that a large number of plasmodesmata connecting mesophyll and bundle sheath remain open after the removal of the mesophyll cells, thus facilitating metabolite exchange with the medium[168]. Precisely measuring chloroplast metabolite transport using isolated bundle sheath strands is not possible due to potential interference from other cellular compartments, such as the cytosol and mitochondria. However, the system should be sufficient for identifying differences in chloroplast transport between wild-type and RNAi lines. If the bundle sheath strands are fed with malate then they should produce pyruvate by return, and if transport at the chloroplast envelope is perturbed in the RNAi lines then the rate of pyruvate production should be reduced in strands isolated from RNAi compared to wild-type plants.

To investigate whether there was a malate/pyruvate transport defect in RNAi plants, it was necessary to provide the isolated bundle sheath strands with malate, to support the conversion of malate to pyruvate by NADP-ME, and then to measure the resulting

pyruvate concentration in the supernatant of the strand suspension. Initially the strands were supplied with malate, as well as 3-PGA, ADP, and light, which together support the generation of NADP within the chloroplast (Figure 2.14 C). Similar experimental set-ups were previously used for isolated maize chloroplast[54] and bundle sheath strands[20]. However, in experiments with *S. viridis* strands the addition of malate+PGA+ADP led to equal levels of pyruvate production to just supplying PGA+ADP, indicating that, under these conditions, 3-PGA was being converted directly to pyruvate rather than indirectly supporting the conversion of malate to pyruvate via NADP-ME in the chloroplast.

Therefore, rather than indirectly supplying the chloroplasts within the bundle sheath strands with NADP by facilitating its production from 3-PGA and ADP in the light, it was decided to instead supply NADP directly in the dark (Figure 2.14 D). This was shown to strongly stimulate pyruvate production, and was dependent on the presence of both malate and NADP (Figure 2.14 B). NAD was found not be able to substitute for NADP (Figure 2.14 B). No transporters capable of NADP uptake into the chloroplast have yet been identified, and the permeability of *S. viridis* bundle sheath chloroplasts to NADP has not been tested. However, there are no known routes by which malate and NADP specifically could support pyruvate production except for through chloroplastic NADP-ME. Therefore, we took this assay forward to test for differences in malate+NADP-dependent pyruvate production rate between bundle sheath strands isolated from wild-type and *HATCH1* knock-down plants.

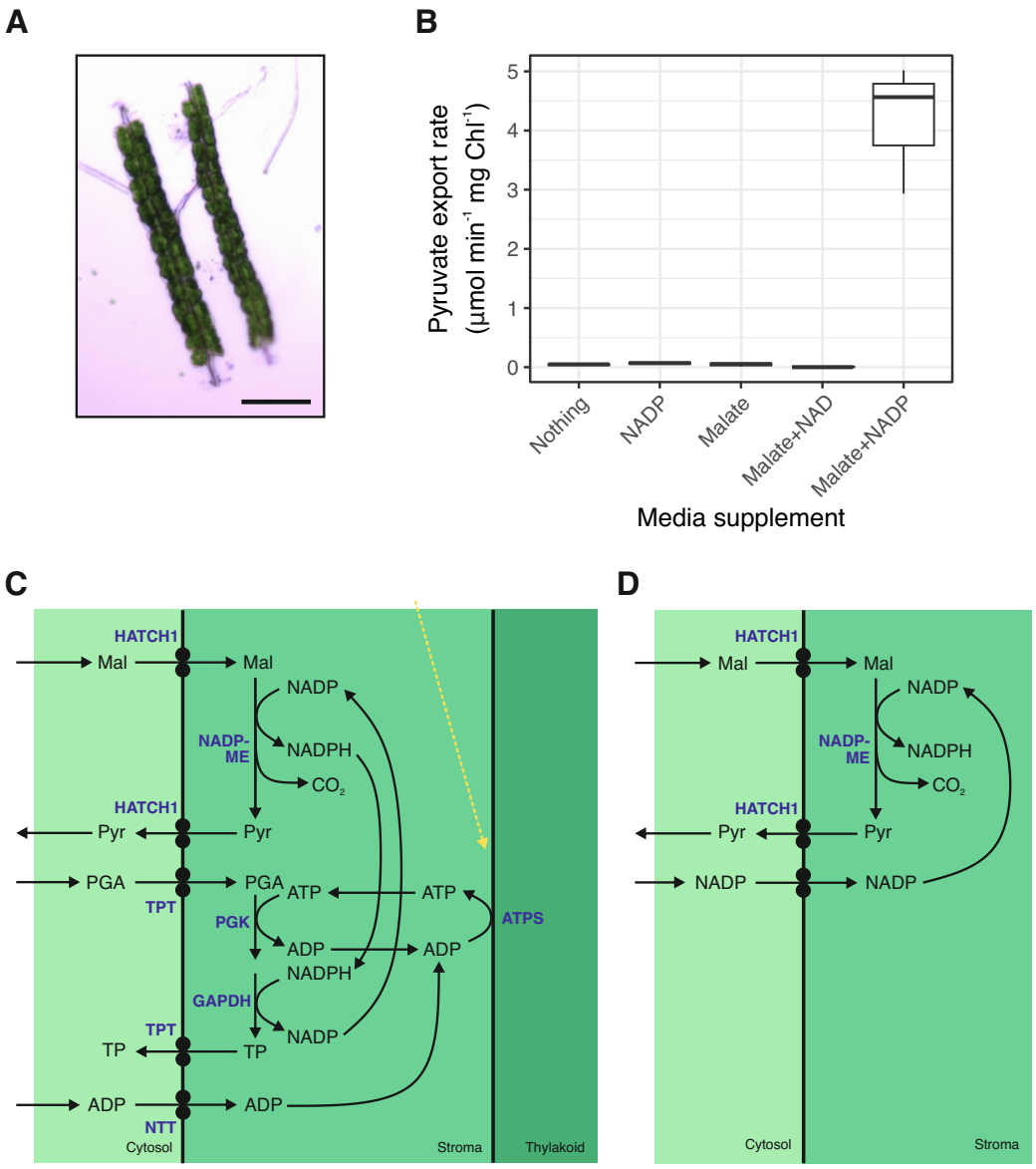


Figure 2.14: A bundle sheath strand assay for chloroplast malate/pyruvate transport. **A**) Representative bright-field stereomicroscope image of isolated bundle sheath strands. Scale bar = 100 µM. **B**) Bundle sheath strand transport assay validation. Isolated bundle sheath strands were added to a reaction buffer containing +/- NAD/NADP/malate, as indicated. Supernatant samples were taken after 30 minutes. Pyruvate export rate is standardised by chlorophyll content. **C**) Schematic representation of original method of supporting bundle sheath chloroplast NADP-ME reaction by supplying 3-PGA, ADP, and light. **D**) Schematic representation of new method of supporting bundle sheath chloroplast NADP-ME reaction by directly supplying NADP in the dark. Yellow arrow = light. For abbreviations see appendix A.1, table A.1.

As predicted, bundle sheath strands isolated from *HATCH1* knock-down *S. viridis* exported pyruvate at a significantly slower rate (Ø: 1.600 µmol min^{-1} mg Chl^{-1} ± 0.125 (SD, N = 13)) than segregating wild-type plants (Ø: 2.066 µmol min^{-1} mg Chl^{-1} ± 0.233 (SD, N = 15)) (Figure 2.15). Until NADP-ME activity assays are performed, it cannot be ruled out that *HATCH1* knock-down is associated with a decrease in NADP-ME activity, which would also explain the observed decrease in malate+NADP-dependent pyruvate production rate. However, the most parsimonious explanation for the observed result is that *HATCH1* knock-down leads to reduced malate/pyruvate transport across the chloroplast envelope, consistent with the transport activity observed in *E. coli*. Interestingly, the observed decrease in pyruvate production rate was only ~25%, compared to a ~90% reduction in *HATCH1* transcript level (see Figure 2.4 A). Transcript abundance is not always a good predictor of protein abundance, particularly in plants[169]. Thus, it is possible that a ~90% reduction in *HATCH1* transcript levels leads to a less substantial reduction in HATCH1 protein level and transport rate. A decrease in HATCH1 protein levels may also be compensated for by its paralogs, for example HATCH3 (discussed further in chapter 3). Either way, a relatively small reduction in transport activity was consistent with the lack of any growth defects in the *HATCH1* knock-down plants: the bundle sheath cells of *HATCH1* knockdown plants were still capable of exchanging malate for pyruvate, albeit at lower levels than wild-type.

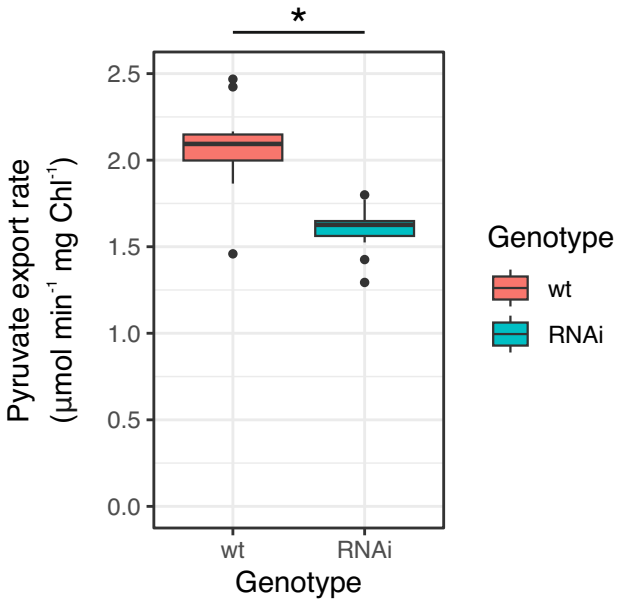


Figure 2.15: Bundle sheath strand transport assay comparing *HATCH1* knock-down (RNAi) and segregating wild-type (wt) *S. viridis*. Isolated bundle sheath strands were added to a reaction buffer containing malate+NADP, and supernatant samples were taken after 30 minutes. Pyruvate export rate is standardised by chlorophyll content. Error bars show standard error of the mean. Asterisk indicates significant difference ($p \leq 0.05$) in pyruvate export rate between RNAi and wild-type samples (t-test, $p=1.0461 \times 10^{-6}$).

2.2.6 Flux balance analysis modelling of HATCH1 knock-down plants recapitulates observed gas-exchange phenotype

Whilst it made intuitive sense that reduced malate/pyruvate transport across the bundle sheath chloroplast envelope (Figure 2.15) would be associated with the observed decrease in maximal photosynthetic rate in *HATCH1* knock-down plants (Figure 2.4 A), we wanted to support this connection using a modelling approach.

One way to answer this kind of question is by flux balance analysis (FBA) modelling. An FBA model utilises a matrix containing stoichiometric coefficients for all known reactions in a biological system. This stoichiometry provides constraints on flux through the system, and further constraints can be added in the form of maximal and minimal fluxes through specific reactions. A biological objective is then identified as some number of reactions from within the stoichiometric matrix that should be maximised. By solving the set of linear equations that are defined by the stoichiometric matrix, it is possible to identify the minimal pattern of flux through the metabolic network that maximises the biological objective[170].

Thus, given a metabolic model of two-cell, NADP-ME subtype C_4 photosynthesis, we proposed to constrain flux through the transport reactions predicted to be catalysed by HATCH1, as well as through the CO_2 input into the model, and observe the effect on CO_2 assimilation by rubisco in the model. This would allow for the generation of an *in silico* A-C_i curve, for comparison with the experimentally determined A-C_i curves (Figure 2.4 A).

I modified a two-cell C_4 FBA model that had previously been generated to investigate the evolution of C_4 photosynthesis[171]. This involved the addition of traits found only in fully evolved C_4 species, the modification of C_4 cycle reactions to more closely match *S. viridis* specifically, as well as the removal of several experimentally unsupported reactions that produced unphysiological metabolic cycles when the model was run (see Methods for more detail). CO_2 uptake into the mesophyll could be varied, and this was converted to an approximate value for C_i (intercellular CO_2 concentration) using curve-

fitting based on experimentally collected physiological data for *S. viridis* leaf gas exchange (see Methods for more detail). When the FBA model was run at these varying C_i levels, flux values for the rubisco carboxylation reaction, assumed to be equivalent to A (net CO_2 assimilation rate) could be extracted, thus generating *in silico* A-C_i curves. Flux through the transporter in the model responsible for bundle sheath chloroplast malate import and pyruvate export could then be restricted, to represent *HATCH1* knock-down plants, or left unconstrained, to represent wild-type plants.

The simulated A-C_i curves predicted that restriction in flux through HATCH1 would lead to a reduction in maximal CO_2 assimilation rate (Figure 2.16). This is in agreement with the experimentally determined A-C_i curves for *HATCH1* knock-down plants (Figure 2.4 A). A HATCH1 flux restriction to 53.2% was required to generate simulated A-C_i curves with maximal assimilation rates equivalent to the experimentally determined curves for RNAi plants. This predicted flux is surprisingly low, and would appear to be in disagreement with the higher ~75% rate of malate+NADP-stimulated pyruvate production rate in the bundle sheath strand assay on RNAi plants (Figure 2.15). Predictions of the precise reduction in transporter flux that are required to cause a specific reduction in CO_2 assimilation are likely beyond the scope of a relatively simple FBA model. Nevertheless, FBA modelling does support the broader conclusion that the experimentally observed reduction in maximal CO_2 assimilation rate occurs as a direct consequence of restriction in flux through HATCH1 in knock-down plants.

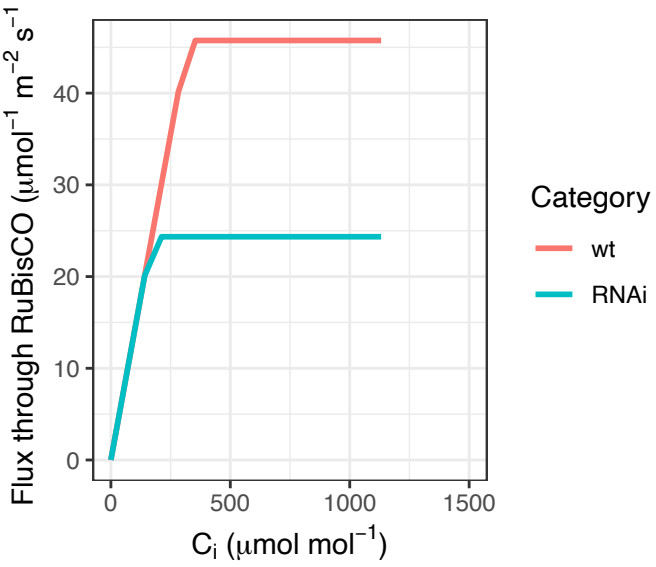


Figure 2.16: Simulated A-C_i curves using flux balance analysis modelling. Flux through rubisco is plotted against estimated intercellular CO_2 concentration (C_i), for both modelled wild-type plants (with unrestricted flux through the reaction corresponding to HATCH1) (red) and *HATCH1* knock-down plants (with flux through the reaction corresponding to HATCH restricted to 53%) (blue).

2.3 Chapter conclusion

In this chapter, I extended the characterisation of a putative chloroplast-localised metabolite transporter involved in the C_4 cycle in *S. viridis*, that had previously been identified and undergone preliminary characterisation by previous lab members.

I showed through a subcellular localisation experiment in *N. benthamiana* that *A. thaliana* HATCH1 protein localises to the chloroplast inner envelope, in agreement with its bioinformatically predicted chloroplast transit peptide and multiple transmembrane domains, and proteomic evidence from *A. thaliana* and maize chloroplast envelopes.

I functionally characterised HATCH1 by exogenous expression in an *E. coli* system. This provided evidence that HATCH1, which had already been demonstrated to have specific pyruvate export activity, has import activity specific to malate, and to a lesser extent succinate. Through *HATCH1* over-expression in a quadruple knock-out *E. coli* strain defective in dicarboxylate transport, evidence was provided to support the hypothesis that HATCH1 does not act as a strict malate/pyruvate antiporter, but rather as a selective channel, which facilitates the independent import of malate and export of pyruvate. It is possible that HATCH1 does have an antiport mechanism, but that it is also able to facilitate significant pyruvate efflux in the absence of external malate. Alternatively, HATCH1 may really act as a true channel, facilitating the import of malate and export of pyruvate in an entirely unlinked manner. This hypothesised mode of transport would be unique to HATCH1 among known C_4 cycle transporters, and would thus represent a very interesting avenue for further inquiry. Distinguishing between the two hypotheses of a leaky antiporter and a true channel has proven to be beyond the capability of the *E. coli* characterisation system used in this project. Work is currently under way in collaboration with Dr. Jani Bolla to reconstitute HATCH1 in liposomes, which would allow for more in-depth characterisation of transporter mechanism and kinetics in a cleaner system that is not complicated by interaction with *E. coli* metabolism.

I also carried out an *in silico* investigation into HATCH1 complex structure using AlphaFold2. Whilst it was not possible to predict the structure of a multimeric HATCH1 complex, it was possible to predict a structure for the protein encoded by the *E. coli* ortholog of *HATCH1*, YqhA. Whilst it should be stressed that the result of this analysis is only a prediction, a YqhA octamer was found to form a channel-like structure of

approximately the right size to facilitate malate and pyruvate transport. The aforementioned effort underway to over-express and reconstitute HATCH1 into liposomes should facilitate the determination of HATCH1 oligomeric number through mass spectrometry, and hopefully the eventual obtaining of an exact structure by crystallography or cryogenic electron microscopy.

Photosynthetic gas-exchange experiments performed by Michael Niklaus demonstrated that *HATCH1* knock-down *S. viridis* show a significant photosynthetic defect. We wanted to confirm that this was related to a restriction in the malate/pyruvate transport mechanism that is observed upon *HATCH1* expression in *E. coli*. I developed a protocol for the isolation of *S. viridis* bundle sheath strands, and an assay for subsequently testing malate+NADP-dependent pyruvate production from these strands. The reduction in malate+NADP-dependent pyruvate production in strands isolated from *HATCH1* knock-down plants compared to wild-type was consistent with down-regulation of a bundle sheath chloroplast envelope malate/pyruvate transport system in these plants. Flux balance analysis modelling was subsequently used to further support the connection between reduced malate/pyruvate transport across the bundle sheath chloroplast envelope, and the reduction in maximal photosynthetic rate that was observed in the *HATCH1* knock-down plants.

This work allows us to propose, for the first time, a complete map of the core C_4 cycle in an NADP-ME subtype C_4 plant, *Setaria* (Figure 2.18). As well as shedding light on a fascinating biological trait, this knowledge will also be useful in efforts to engineer C_4 photosynthesis into C_3 plants for increased yield. For this reason, in conjunction with the

possibility of utilising a malate/pyruvate transporter for biotechnological processes in industry, a patent application has been filed for this transporter (WO2021028876).

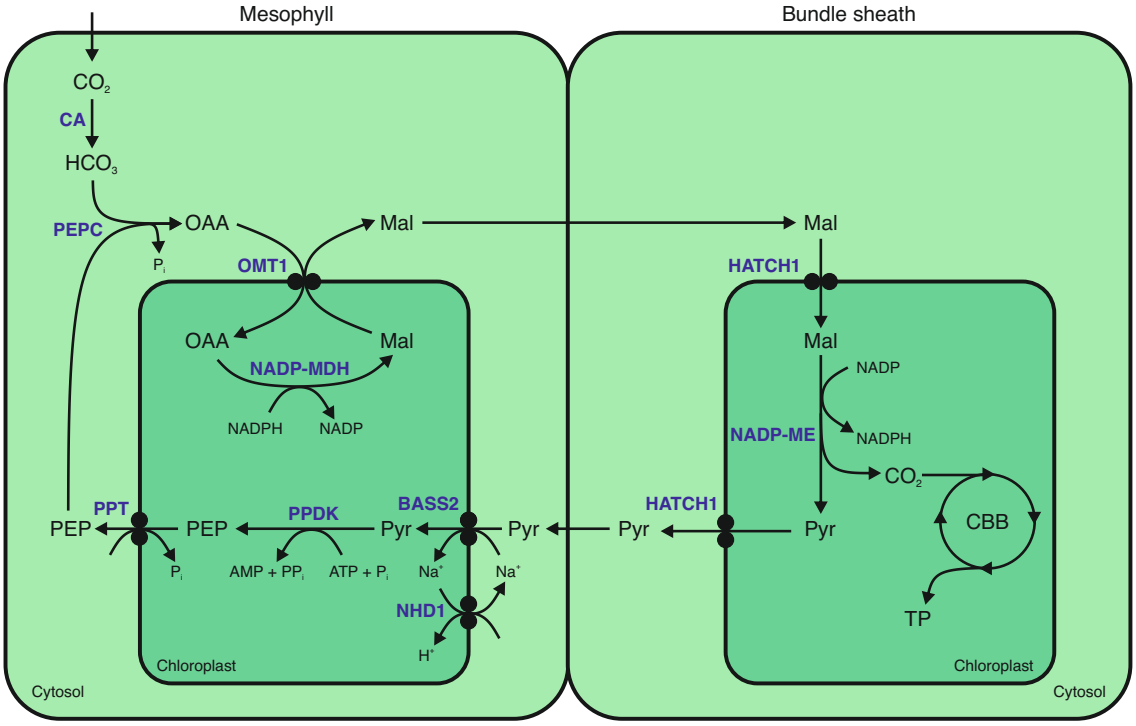


Figure 2.18: Schematic representation of our current understanding of the NADP-ME subtype C_4 cycle in *S. viridis*, including HATCH1. Metabolites are shown in black. Enzymes/transporters are shown in blue. For abbreviations see appendix A.1, table A.1.

2.4 Materials and methods

2.4.1 Generation of GFP fusion constructs for expression in *N. benthamiana*

The coding sequences for *S. italica TPT* (Seita.5G002900) and *BASS2* (Seita.5G242900), and *A. thaliana TPT* (AT5G46110) were acquired from Phytozome V13[172]. These sequences were modified by the addition of Golden Gate SC1 module adaptors,

domestication to remove BsaI recognition sites, and stop codon removal (Appendix A.5.1-3). For *SiBASS2*, further codon optimisation for ease of synthesis was also required. The sequences were synthesised by Twist Bioscience. Coding sequences for *S. italica HATCH1* (Seita.4G275500) and *A. thaliana HATCH1* (AT4G19390) had already been acquired and modified in the same manner by Michael Niklaus (Appendix A.5.4-5).

These sequences were assembled into binary vectors for transient expression of *C*-terminally GFP-tagged proteins in *N. benthamiana* (see Appendix A.5.8 Tables A.55-60) by Golden Gate reactions using the restriction enzyme BsaI-HFv2 (NEB), and the recipe and thermocycler settings at Appendix A.3.1, table A.11 and table A.12 respectively. After *E. coli* transformation and plasmid isolation, correct assembly was confirmed by Sanger sequencing (Source Bioscience) using the primers in Appendix A.5.8, table A.63.

2.4.2 Transient protein over-expression in *N. benthamiana* leaves

Freeze-thaw competent *A. tumefaciens* strain GV3101 pMP90 was transformed with the binary expression plasmids (section 2.4.1). Single colonies were inoculated in 5 mL LB (containing 50 μg/mL Rifampicin, 10 μg/mL Gentamycin, 50 μg/mL Kanamycin) and incubated at 28°C, 200 rpm overnight. The cells were collected by centrifugation (2500 xg, 10 mins) and discarding the supernatant, before being resuspended in infiltration buffer (10 mM $MgCl_2$, 10 mM MES pH 5, 0.15 mM acetosyringone) to an OD_{600} of 0.5. This suspension was infiltrated into 4-week-old fully-expanded *N. benthamiana* leaves.

2.4.3 Confocal microscopy analysis of transiently transformed *N. benthamiana* leaves

Infiltrated plants were used for imaging 3-5 dpi. Leaf samples were cut from infiltrated leaves, placed upside-down on a glass slide, and treated with perfluorodecalin. Leaf

samples were examined with a Zeiss Axio Imager Z.2 confocal microscope using the C-Apochromat 40x/1.2 W objective lens. The laser wavelength was 488 nm, and the channels for detection of GFP fluorescence and chlorophyll autofluorescence were set to 495-551 nm and 650-750 nm respectively. The images were processed using the Fiji distribution of ImageJ.

2.4.4 Generation of *S. italica HATCH1* construct for expression in *E. coli*

This construct had already been generated by Michael Niklaus. Briefly, the coding sequence for *S. italica HATCH1* (Seita.4G275500) was acquired from Phytozome V12[172], before being codon optimised for expression in *E. coli* using OPTIMIZER[173], modified by the addition of Golden Gate SC module adaptors, and domesticated to remove BsaI recognition sites (Appendix A.4.1). The sequences were synthesised by Gene Universal, and assembled into the lv1 vector (Appendix A.4.28, table A.27).

2.4.5 Experimental evaluation of dicarboxylate uptake specificity of *HATCH1* in *ΔdctA E. coli*

ΔdctA E. coli was transformed with the *SiHATCH1* inducible expression vector using heat-shock, and plated on LB agar (containing 10 μg/mL Gentamycin – as in all media from here on). One colony was picked and inoculated into 5 mL LB, and incubated at 37°C, 200 rpm overnight. The next day, 1 mL of this culture was inoculated into 50 mL M9 glucose (see Appendix A.2.1) in a 250 mL Erlenmeyer flask, and incubated at 37°C, 200 rpm for 24 hours. The next day, the cells were inoculated to an OD_{600} of 0.2 in M9 glucose, and incubated at 37°C, 200 rpm until reaching an OD_{600} of 0.4. To induce the expression of *HATCH1*, 50 μM inducer (2,4-Diacetylphloroglucinol, or DAPG) was added,

and the cultures were incubated as before for 2 hours. Next, the OD was measured, before the cells were collected by centrifugation (5000 xg, 5 mins), the supernatant was discarded, and the pellet was resuspended in sufficient M9 media containing carbon source(s) of interest and 50 μM DAPG to give an OD_{600} of 2. The M9 media were: M9 glucose, M9 glucose + malate, M9 glucose + fumarate, M9 glucose + succinate, M9 glucose + aspartate (see Appendix A.2.1). After a further 15-minute incubation as before, supernatant samples were taken, sterilised through a 0.22 μm filter, and flash frozen in liquid nitrogen. Supernatant pyruvate content was measured as in section 2.4.7.

2.4.6 Expression of *HATCH1* in IMW213 *E. coli* on minimal media containing glucose as a sole carbon source

IMW213 *E. coli* was transformed with the *SiHATCH1* inducible expression vector using heat-shock, and plated on LB agar (containing 50 μg/mL Kanamycin and 10 μg/mL Gentamycin – as in all media from here on). One colony was picked and inoculated into 5 mL LB, and incubated at 37°C, 200 rpm overnight. The next day, 1 mL of this culture was inoculated into 50 mL M9 glucose (see Appendix A.2.1) in a 250 mL Erlenmeyer flask, and incubated at 37°C, 200 rpm for 24 hours. The next day, the culture was transferred to a 50 mL falcon tube and cells were collected by centrifugation (3200 xg, 10 mins). The cells were washed twice in M9 - (see Appendix A.2.1) by decanting the supernatant, resuspending in M9 -, centrifugation, and repeating. After the second wash, the cells were resuspended again in 5 mL M9 -, and OD was measured. Subsequently, the cells were inoculated to an OD_{600} of 0.1 in 6 mL M9 glucose, and incubated at 37°C, 200 rpm,

for 3 hours. The OD was measured again, and 1 mL T_0 supernatant samples were taken, sterilised through a 0.22 μm filter, and flash frozen in liquid nitrogen. To induce expression of *HATCH1* 50 μM DAPG was added to three of the cultures, whilst the others remained un-induced. The cultures were incubated for 3 hours, as before. The OD was measured again, and 1 mL T_1 supernatant samples were taken and processed as before.

2.4.7 Pyruvate measurement in cell supernatant

The concentration of pyruvate in the supernatant was assessed using a colorimetric assay (abcam ab65342), according to the instructions provided. Briefly, 50 μl pyruvate standard or supernatant were mixed with 50 μl reaction mix and aliquoted in duplicate into a 96-well plate. After incubation for 30 minutes in the dark at room temperature, the absorbance at 570 nm was measured using a plate reader (FLUOstar Omega, BMG Labtech).

2.4.8 Whole genome sequencing and read mapping of AN387 and IMW213 *E. coli*

Whole genome sequencing to 30x coverage of *E. coli* strains was carried out using the commercial sequencing service provided by MicrobesNG (https://microbesng.com). The genome sequence for *E. coli* K12 was obtained from NCBI and the genome assembly file (accession GCA_000981485) was obtained from Ensembl Bacteria. In order to generate a genome index for *E. coli* K12 and map the trimmed reads against this genome index, I developed a pipeline in Snakemake[174], which can be accessed on GitHub (https://github.com/Olliemattinson/RNA_seq_read_mapping_pipeline). Briefly, the K12 genome index was generated and trimmed reads were mapped to it using hisat2 (v

2.1.0)[175], before Samtools (v 1.3.1)[176] was used to sort and index the mapped reads. Mapped reads were visualised using IGV[177].

2.4.9 *YqhA* and *HATCH1* structure prediction using AlphaFold2

The amino acid sequence for *E. coli YqhA* was obtained from EcoCyc[178]. Protein and multimer complex prediction was performed using ColabFold[179]: an online server for running AlphaFold2[165,167] combined with MMseqs2[180] for faster homology search. The amino acid sequence for *S. italica HATCH1* (Seita.4G275500) with predicted cTP removed was taken from the previously designed expression construct (see Appendix A.4.1). Protein and multimer complex prediction was performed using AlphaFold2[165,167] on the university's Advanced Research Computing clusters[181].

2.4.10 *S. viridis* plant growth conditions

Seeds were incubated for three weeks in moist Sphagnum moss at 4°C to break dormancy, before being dried at 37°C for two days. Seeds were scattered on soil (Sinclair Modular Seed Growing Medium), covered with a further thin layer of soil, drenched with water, and covered with a transparent, plastic cover to maintain humidity. Pots were transferred to a greenhouse facility at 28/20°C (day/night) with long photoperiod (16 hour light, 8 hour dark). As plants began to touch the lid, the lid was removed. Plants were watered every other day, with enough water to soak the soil without leaving any standing water.

2.4.11 *S. viridis* bundle sheath strand isolation

RNAi and segregating wild-type *S. viridis* were placed in the dark for two days prior to bundle sheath strand isolation in order to deplete chloroplastic starch reserves which could damage the chloroplasts during isolation. All subsequent steps, unless otherwise indicated, were performed on ice and protected from light. ~40 fully expanded leaves were taken from each line and leaf ends and midribs were removed using a razor, and weighed to ensure approximately equal tissue for each line. Leaf halves were then sliced into ~1 mm slices and transferred to chilled resuspension buffer (Appendix A.2.2, table A.4), and stored at 4°C overnight. The next morning, leaf slices were collected on a strainer and transferred to 50 mL enzyme solution (Appendix A.2.2, table A.5) in a conical vacuum flask. Vacuum was applied for 15 minutes to allow the enzyme solution to infiltrate into the leaf slices. The flask was then incubated at 30°C, 65 rpm for 1 hour to allow digestion to occur. Subsequently, the digested leaf material was collected on miracloth and transferred to a glass beaker containing 200 mL resuspension buffer. The leaf slices were agitated with a pipette tip for 2 minutes to release bundle sheath strands, and the solution was poured through a metal strainer into a second glass beaker, which now contained separated strains in resuspension buffer. The leaf material collected on the strainer was returned to the first glass beaker with 40 mL resuspension buffer, and agitation followed by filtration was repeated a further three times. Next, the remaining undigested leaf material was returned to the conical vacuum flask containing 50 mL fresh enzyme solution, vacuum was applied again for 5 minutes, and the flask was incubated as before. Leaf slice agitation and filtration was performed as before, two times. The contents of the beaker containing isolated strands and other debris were

filtered twice through 70 μm nylon filters (Spectrum), and further rinsed with resuspension buffer to remove ruptured cell debris. Clean strands collected on the filter were then transferred to ~10 mL resuspension buffer in a petri plate, and stored in the dark at 4°C until use.

2.4.12 *S. viridis* bundle sheath chloroplast isolation

Chloroplast isolation from isolated BS strands was attempted through several methods, including collection in miracloth and gently crushing using a pre-chilled pestle and mortar, or further strand digestion using enzyme solution (Appendix A.2.2, table A.5) for 1 hour at 30°C, 65 rpm, followed by collection and agitation on a strainer. The broken strand mixture was then filtered successively through 70 μm and 40 μm nylon filters (Spectrum). The filtrate was layered on top of either 10 mL 40% Percoll, or 10 mL 40% Percoll on top of 80% Percoll in a 50 mL tube. The tubes were centrifuged (3200 xg, 4°C, 15 mins) and respectively the pellet or band at 40%-80% Percoll boundary (containing intact chloroplasts) were collected, and added to 5 mL resuspension buffer (Appendix A.2.2, table A.4) in a fresh 50 mL tubes. The tubes were centrifuged again (1700 xg, 4°C, 5 mins), the supernatant was removed, and the pellet was carefully resuspended in 1 mL resuspension buffer. Chloroplast intactness was assessed visually using Zeiss Axio Imager Z.2 confocal microscope using the C-Apochromat 40x/1.2 W objective lens, using the fluorescent dye CFDA[182].

2.4.13 *S. viridis* bundle sheath strand transport assay

0.5 mL bundle sheath strands in resuspension buffer were added to 24-well plate wells (in the case of 3-PGA+ADP experiments) or 2 mL tubes (in the case of NADP

experiments). 0.5 mL resuspension buffer was also added to wells/tubes as a negative control. 0.5 mL of either 3-PGA+ADP assay buffer (Appendix A.2.2, table A.6) or NADP assay buffer (Appendix A.2.2, table A.7) were then added as appropriate. The samples were incubated for 30 mins (30°C, 150 rpm) either in the light (for 3-PGA+ADP experiments) or in the dark (for NADP experiments). Next, the strands were allowed to settle, and 0.5 mL supernatant was taken, filtered through a 0.22 μm filter, flash frozen in liquid N, and stored at -20°C until later measurement of pyruvate content, as in section 2.4.7. A further 0.25 mL supernatant was removed from the strand solution and discarded. The remaining 0.25 mL strand suspension was taken forward for chlorophyll quantification, as in section 2.4.14. Pyruvate export rate over the incubation period for each sample was then standardised by chlorophyll content of that sample.

2.4.14 Chlorophyll quantification

Chlorophyll extraction and quantification was carried out based on Lichtenthaler and Buschmann (2001)[183]. 1 mL 100% acetone was added to each 0.25 mL strand suspension sample, and vigorously mixed. Chlorophyll extraction was allowed to occur at 4°C for 24 hours. The next day, the extracted chlorophyll solution was filtered through a 0.22 μm filter and aliquoted into a clear 96-well plate. Absorbance was measured using a plate reader (FLUOstar Omega, BMG Labtech) at wavelengths of 647 nm, 663 nm, and 750 nm. Absorbance values at each wavelength were blank corrected by subtracting mean absorbance value of blank wells (containing just 80% acetone and strand resuspension buffer). Blank-corrected A750 values were then subtracted from blank-corrected A647

and A663 values for each well to remove background. Chlorophyll concentration was carried out according to the following equation:

$$mg\ Chl\ L^{-1} = 7.15\ A663 + 18.71\ A647$$

2.4.15 Flux balance analysis modelling

The two-cell C_4 metabolic model was adapted from Blätke and Bräutigam (2019)[171], with the following alterations. Fluxes through NAD-ME and PCK were set to zero, as *S. viridis* carries out NADP-ME subtype C_4 photosynthesis with negligible expression of either PCK or NAD-ME[56]. Flux through a chloroplast pyruvate:H^+ symporter was eliminated, as *Setaria* imports pyruvate into the chloroplast using a pyruvate:Na^+ symporter[94]: BASS2[56,99]. Flux through BASS2 was restricted to the mesophyll cell, and in the forward direction (import of pyruvate into the chloroplast), as neither bundle sheath expression[56] nor reverse direction transport[99] have been demonstrated. As the model was originally generated to investigate the evolution of C_4 photosynthesis from a C_3 ancestor, it did not include many of the subsequent C_4 adaptations present in *S. viridis*, which therefore had to be added into the model. Specifically, photosystem II was restricted to the mesophyll cell[56], rubisco was restricted to the bundle sheath cell[56], and oxygen uptake into the bundle sheath cell was also removed, due to the diffusion barrier created by the suberized bundle sheath cell wall[184]. To prevent the emergence of non-physiological metabolic cycles in the model, fluxes through a chloroplast alanine uniporter, glutamate uniporter, and pyrroline-5-carboxylate uniporter were eliminated, as none of these transport activities have been experimentally described in the literature. The maximum values of flux through the carboxylation reactions of rubisco

and PEPC were set to 45.7 µmol m^{-2} s^{-1} and 128.1 µmol m^{-2} s^{-1} respectively, as was previously estimated by Michael Niklaus directly from the measured A-C_i curves on segregating azygous lines (Figure 2.4 B) using mathematical models of C_4 photosynthesis[185]. The flux balance analysis model was optimized for minimal total flux through the metabolic network, minimal oxygenation rate by rubisco, and maximal sucrose output. The original model contained a chloroplast malate importer and pyruvate exporter, which was assumed to be encoded by *DiT2*[171]. Pyruvate export activity has not been demonstrated for this transporter[30,52]. However, in this work we show that *HATCH1* encodes the transporter with this activity in the bundle sheath cell and so this component of the model is retained, but restricted to the bundle sheath cell. To simulate an A-C_i curve, CO_2 uptake rate into the mesophyll cytosol was varied between 0 µmol m^{-2} s^{-1} and 160 µmol m^{-2} s^{-1}. The CO_2 uptake rate into the mesophyll cytosol in the model was converted to an approximate value for the difference in leaf intercellular airspace CO_2 partial pressure (C_i) and mesophyll CO_2 partial pressure (C_m), assuming, in the absence of measured values for *S. viridis*, a plasma membrane CO_2 conductance of 0.35 cm s^{-1} [186], cell wall CO_2 resistance of 79.5 m_{chlor}^2 s bar mol^{-1} [186], and surface area of chloroplasts exposed to intercellular airspace per unit leaf area of 14.67 m^2 m^{-2} [187]. Steven Kelly then converted this to an approximate value for C_i by fitting a linear model to the measured relationship between C_i-C_m and C_i (Appendix A.6, Figure A.1) using data obtained directly from measurements in *S. viridis*[188]. To simulate wild-type plants, flux through HATCH1 was unconstrained. To simulate HATCH1 RNAi plants, a flux value of 53.2% was obtained such that the assimilation rate of the model matched the assimilation rate of the HATCH1 RNAi plants. Flux through the rubisco carboxylation

reaction is equivalent to net CO_2 assimilation rate (A). The original model, and analysis workflow including alterations made to the two-cell model can be accessed on GitHub (https://github.com/Olliemattinson/C4_CBM_knock-down_plus_A-Ci).

Chapter 3: Investigation into the role of HATCH1 in C_4 photosynthesis in *Setaria viridis*

3.1 Chapter introduction and background

3.1.1 Why did knockdown of a transporter of the C_4 cycle not produce a growth effect?

In chapter 2 I presented evidence that HATCH1 facilitates the transport of malate and pyruvate across the bundle sheath chloroplast envelope as part of the NADP-ME subtype C_4 photosynthetic pathway in *Setaria viridis*. However, the question arises as to why *S. viridis* plants with a 90% reduction in *HATCH1* gene expression do not exhibit a growth defect. Three possible explanations present themselves. 1) The remaining ~10% *HATCH1* expression may be enough to facilitate sufficient bundle sheath chloroplast malate/pyruvate transport to support the NADP-ME C_4 cycle. 2) Significant loss of HATCH1 may be compensated for by its ortholog *HATCH3*, which is expressed to reasonably high levels in both the mesophyll and bundle sheath cells. 3) *S. viridis* is able to compensate for reduced NADP-ME C_4 cycle activity through other, ancillary C_4 cycles, that can also facilitate the provision of CO_2 to bundle sheath chloroplast-localised rubisco. The rationale for the proposition of hypotheses 2 and 3 above are each described in turn below.

3.1.2 Hypothesis 2: *HATCH1* RNAi may be compensated for by the expression of a *HATCH1* paralog in *S. viridis*

The HATCH family (characterised by the presence of a UPF0114 domain) contains two members in *S. viridis*. *HATCH1* shows high, bundle sheath-preferential expression, whilst *HATCH3* shows intermediate expression in both mesophyll and bundle sheath cells (Figure 3.1 A). *HATCH2* has been lost in *S. viridis*. The lower, but still appreciable, bundle sheath-specific expression of *HATCH3* compared to *HATCH1* suggests that HATCH3 may be able to compensate for reduced HATCH1 protein levels. Like HATCH1, HATCH3 is also predicted to have an N-terminal chloroplast transit peptide and four transmembrane domains (Figure 3.1 B). The 3D structure of the HATCH3 monomer, as predicted by AlphaFold2, is also very similar to that of HATCH1 (Figure 3.1 C). Thus, it is possible that 90% knockdown of *HATCH1* in the presence of HATCH3 enables sufficient flux through the NADP-ME pathway to support normal plant growth.

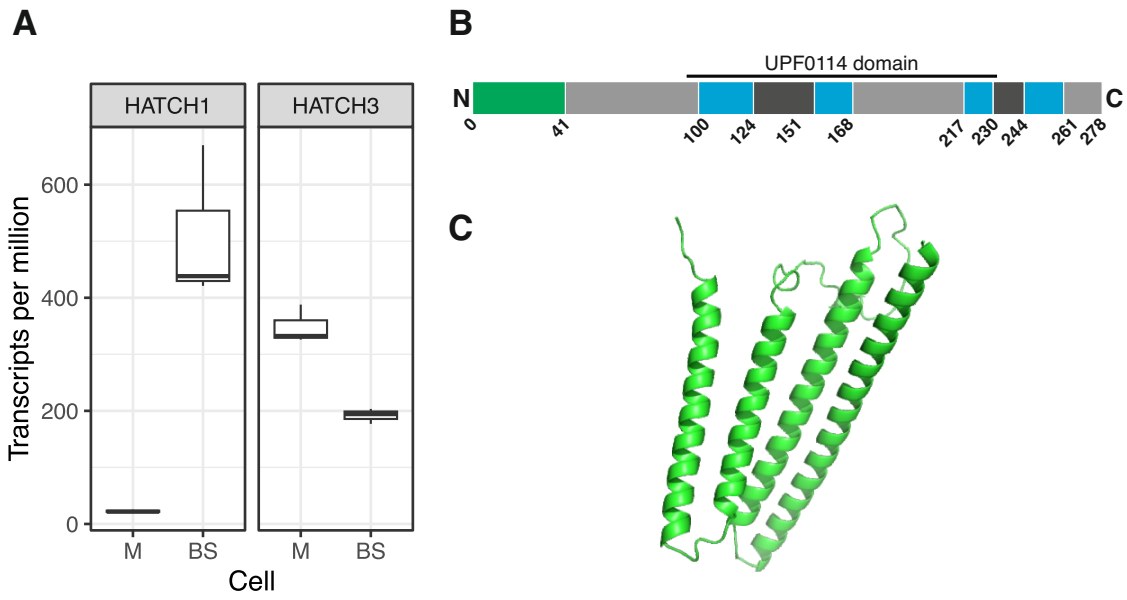


Figure 3.1: Bioinformatic characterisation of *HATCH* genes in *S. viridis*. **A**) Mesophyll (M) and bundle sheath (BS) cell-type specific transcript abundance for *HATCH* family genes in *S. viridis*. Data obtained from John et al. (2014)[56]. **B**) Cartoon structure of HATCH3 protein. Colours represent different protein domains: Green = chloroplast transit peptide, Blue = transmembrane domain, Light grey = cytoplasmic domain, Dark grey = non-cytoplasmic domain. Conserved UPF0114 domain is indicated. Domain information obtained using Interpro[141], TargetP-2.0[142], and DeepTMHMM[143]. **C**) Structural prediction of the HATCH3 monomer using AlphaFold2[165].

3.1.3 Hypothesis 3: *HATCH1* RNAi may be compensated for by the operation of auxiliary C4 cycles in *S. viridis*

In order to evaluate the three hypotheses outlined above it is necessary to consider potential complexities in our understanding of the C_4 cycle in *S. viridis*. As outlined in the introduction to this thesis, C_4 plants were initially categorised into three subtypes based on the decarboxylation enzyme with highest activity: NADP-ME (NADP malic enzyme), NAD-ME (NAD malic enzyme), and PCK (phospho*enol*pyruvate carboxykinase)[17].

However, subsequent studies across the angiosperm phylogenetic tree have revealed that the diversity of C_4 cycles is more complex than originally thought. Both the monocot maize (*Zea mays*) and the dicot *Flaveria bidentis*, which primarily carry out NADP-ME subtype C_4 photosynthesis, were shown to have significant flux through the primary transfer acid for NAD-ME/PCK subtype C_4 photosynthesis, aspartate, as well as the primary NADP-ME subtype transfer acid, malate[189,190]. With the discovery that PCK is also present at appreciable activities in the bundle sheath cells of maize and other 'NADP-ME subtype' C_4 plants[191], and that this enzyme is capable of supporting high rates of aspartate-dependent photosynthesis in isolated maize bundle sheath cells[20], it became clear that the NADP-ME cycle frequently operates in combination with an auxiliary PCK cycle. Similarly, the NAD-ME cycle has been shown frequently to operate alongside an auxiliary PCK cycle[192,193]. In fact, a C_4 species that relies exclusively on PCK as a decarboxylase has not yet been identified, with species traditionally designated as 'PCK subtype' generally being found to also utilise substantial NAD-ME[194]. It has therefore been proposed that a more accurate paradigm for C_4 photosynthesis involves mutually exclusive NADP-ME and NAD-ME subtypes, and varying degrees of PCK activity providing an auxiliary C_4 cycle in many species[19].

Whilst utilising a mixed C_4 pathway is expected to increase leakiness of decarboxylated CO_2 from the bundle sheath and therefore decrease the efficiency of the C_4 cycle[195], a number of advantages to mixed C_4 cycles have been proposed. Utilising multiple C_4 cycles with different transfer acids would reduce the concentration gradient requirements of each individual metabolite to facilitate diffusion through the plasmodesmata between mesophyll and bundle sheath cells[196]. Furthermore, modelling

analysis has predicted that mixed C_4 cycles would help to modulate NADPH/ATP balance between mesophyll and bundle sheath cells, and that flexibility in C_4 cycle usage could help to maintain high photosynthetic efficiency under varying environmental conditions[19,197]. Indeed, the primarily NADP-ME subtype species maize and sugarcane (*Saccharum spp.*) show flexibility in the relative levels of NADP-ME and PCK activity in response to changing conditions including shading, water- and salinity-stress[198–200].

S. viridis (and its close relative *S. italica*) is generally thought to be a "pure" NADP-ME subtype C_4 plant[201]. However, in different experiments by different research groups evidence has been presented for the existence of both auxiliary PCK and NAD-ME cycles[202–204]. Specifically, whilst samples isolated from *S. italica* bundle sheath cells show only low transcript levels for both NAD-ME and PCK[56], enzyme activity assays using separate leaf samples showed low levels of NAD-ME activity but high levels of PCK activity, comparable with NADP-ME[204]. In contrast, one multi-omics study could not detect significant levels of PCK transcript or protein in samples isolated from mature *S. italica* tissue, but did observe significant levels of NAD-ME transcript and protein (alongside much higher levels of NADP-ME)[203]. Another study also could not detect PCK protein in samples isolated from *S. viridis* mesophyll and bundle sheath cells, but did find significant, bundle sheath-preferential accumulation of NAD-ME protein[202]. The regulatory properties of both isolated *S. viridis* NAD-MEs are consistent both with their ancestral role in respiration as well as with a potential role in a supplementary C_4 pathway[202]. Thus, depending which study one believes, all three C_4 cycles apparently exist side-by-side in *S. viridis*. The reason for the discrepancy between different studies regarding the contribution of different decarboxylases in *Setaria* likely lies in the

metabolic flexibility that multiple co-existing C_4 cycles are hypothesised to provide[19,197]. Differing growth conditions and sampling methodologies between experiments may have also led to different balances in the relative transcript, protein, and enzyme activity levels of NADP-ME, NAD-ME, and PCK in *S. viridis*. Although further work clearly remains to be done to validate and elucidate the roles of auxiliary C_4 cycles in *S. viridis*, it seems plausible that disruption to the NADP-ME cycle in *S. viridis* by knocking down *HATCH1* may lead to at least some compensation by NAD-ME and/or PCK cycle activity, and thus may explain why no growth defect was observed in the *HATCH1* RNAi plants. Thus, to address all three of the hypotheses outlined at the start of this chapter it was decided to knockout both *HATCH1* and *HATCH3* using CRISPR-Cas mutagenesis.

3.1.4 The use of CRISPR/Cas9 to generate gene knock-outs in plants

Genomic mutations can be introduced into eukaryotes by generating double-stranded breaks (DSBs) in DNA. In plants, these DSBs are rescued through the DNA-repair process of non-homologous end-joining (NHEJ). However, NHEJ is error prone and often leads to insertions or deletions at the site of DNA damage[205], which may subsequently disrupt the function of the target DNA sequence, for example by causing a frameshift in the reading frame of a protein-coding gene. The CRISPR/Cas ('Clustered Regularly Interspaced Short Palindromic Repeats'/'CRISPR-Associated') system was discovered only ten years ago as an immune response system in prokaryotes, and was immediately recognised as a potential tool for genetic engineering[206]. Whilst ZFNs (Zinc-Finger Nucleases) and TALENs (Transcription Activator-Like Effector Nucleases) had previously

been used to generate targeted DNA DSBs, and thus induce targeted mutations[207,208], these systems relied on protein-DNA interactions for nuclease targeting, making construct design laborious and challenging[209]. CRISPR/Cas circumvents this issue by targeting the Cas nuclease (commonly Cas9 from *Streptococcus pyogenes*) to the DNA sequence of interest using a complementary single guide RNA (sgRNA) (a fusion of two naturally occurring short CRISPR RNAs found in prokaryotes)[206]. As CRISPR/Cas is based on the much simpler 1:1 interaction between DNA and RNA (compared to complex DNA-protein interactions), it is easily programmed to target an extremely large range of sites. The one requirement in targeting sites is the presence of a protospacer-adjacent motif (PAM): a specific short DNA sequence immediately downstream of the DNA sequence targeted by the sgRNA, which is required for DNA cleavage by Cas (5'-NGG-3' in the case of SpCas9)[210].

The simplicity of sgRNA design and construct cloning allows for the inclusion of multiple guides within a single construct, and thus facilitates the simultaneous targeting of multiple sites within the genome. This can increase the efficiency of mutagenesis (through redundant targeting of multiple sites within a gene) and can also facilitate the parallel mutagenesis of multiple separate genes (thus significantly expediting the knocking-out of multiple genes, which would otherwise have to be performed in series over several generations). One method for multiplexing sgRNAs in a single construct, without expressing them all under individual promoters and thus risking a large, repetitive construct that is prone to sequence loss by recombination, is to express the sgRNAs under a single promoter separated by tRNAs. These tRNAs are then cleaved out

using highly conserved cellular tRNA processing machinery, thus releasing the individual sgRNAs[211].

Generating targeted, heritable gene knock-outs in plant species amenable to genetic transformation is now routine using the CRISPR/Cas system[212]. *S. viridis* can be efficiently transformed by *Agrobacterium*-mediated transformation of embryogenic callus, followed by plant regeneration[213]. This method has recently been used to transform *S. viridis* with a CRISPR/Cas9 construct and generate stable, targeted, genomic mutations[214].

3.1.5 Chapter aims and outline

Whilst the work in chapter 2 supported a role for HATCH1 in C_4 photosynthesis in *S. viridis*, in the absence of a complete gene knock-out, it was not possible to tell if HATCH1 is *necessary* for the C_4 cycle to function. Thus, the primary aim of the work in this chapter was to address this question through the generation and phenotypic analysis of CRISPR/Cas9-mediated *hatch1* knock-out lines. As a secondary objective, we wanted to investigate whether HATCH3 partially compensated for *HATCH1* knock-down/out. Thus, I also aimed to generate and phenotypically characterise *hatch3* single- and *hatch1,hatch3* double-knock-out lines.

3.2 Results and discussion

3.2.1 Generation of *hatch1* single- and *hatch1,hatch3* double-knock-out lines in *S. viridis* using CRISPR/Cas9

In order to investigate whether HATCH1 is necessary for C_4 photosynthesis in the NADP-ME subtype C_4 species, *S. viridis*, I attempted to generate *hatch1* knock-out lines. Phenotypic analysis of these lines would allow us to determine whether the *HATCH1* knock-down lines showed limited phenotypic disruption due to the remaining ~10% *HATCH1* expression still leading to the accumulation of enough HATCH1 protein to facilitate sufficient transport to support significant NADP-ME subtype C_4 cycle activity. Additionally, I decided to attempt to generate *hatch3* knock-out lines, as well as *hatch1,hatch3* double knock-out lines. By comparing the *hatch1* and *hatch1,hatch3* knock-out lines, this would allow us to determine whether loss of HATCH1 activity can be compensated at least partially by HATCH3, which could explain the limited phenotypic disruption observed in the *HATCH1* knock-down lines.

Four guides each were designed to target the first half of *HATCH1* and *HATCH3*. DNA encoding these sgRNAs were assembled into units encoding two polycistronic tRNA-sgRNA arrays, for targeting *HATCH1* and *HATCH3* respectively, which were finally assembled into three binary vectors: pOM078 (for the targeting of *HATCH1* only), pOM079 (for the targeting of *HATCH3* only), and pOM080 (for the targeting of *HATCH1* and *HATCH3* together).

These constructs were transformed into *S. viridis* calli, which were transferred to selective media and regenerated to form T_0 plants. T_0 plants were genotyped for the

presence of the integrated construct, with construct-positive plants subsequently being sequenced to check for mutations in the genes of interest. Unfortunately, many of the T_0 plants showed mixed Sanger sequencing traces, as well as variable sequencing results between different leaf samples from the same plant. This indicates genetic chimerism caused by action of the CRISPR/Cas9 system later in embryonic development, leading to mutations that are present only in parts of the plant derived from the initial cell in which the mutation first occurred. It was therefore not possible to confirm specific mutations in the T_0 plants, but promising lines, in which mutations seemed to have occurred with high frequency, were taken forward to the next generation.

The T_1 plants were similarly genotyped for the presence of the construct, although in this generation I looked for plants in which the construct had also been segregated out. These plants were again sequenced to check for mutations in the genes of interest. Homozygous mutations were sought, in order to allow phenotypic characterisation to begin in the next generation. This process resulted in the identification of four independent, transgene-free, homozygous *hatch1* mutant lines (three having been transformed with pOM068, one having been transformed with pOM070, but with *HATCH3* editing failing to occur), and one transgene-free homozygous *hatch1,hatch3* double mutant line. There was often variability in mutations amongst transgene-free mutant T_1 plants from the same T_0 parent, as would be expected given chimerism in the T_0 generation. Unfortunately, no T_1 plants were found to be transgene free and have homozygous edits in *HATCH3* alone. Seed was collected from transgene free, heterozygous *hatch3* single mutant plants. However, as obtaining *hatch3* single mutant lines was deemed the lowest priority for this study, and given that an extra generation

would be required to obtain homozygous *hatch3* mutants and therefore that insufficient time was left in my DPhil to obtain and characterise these *hatch3* lines, it was decided not to progress further here with the *hatch3* single mutant lines.

One plant per transformation event each for three *hatch1* single mutant lines (68-7-3, 68-16-4, and 70-19-15) and the one *hatch1,hatch3* double mutant line (70-9-49), were taken forward for seed collection for further analysis in the T_2 generation. These T_1 plants all contained homozygous mutations at the predicted Cas9 cut-site for at least one of the four guides targeting *HATCH1* (for 68-7-3, 68-16-4, and 70-19-15), and *HATCH1* and *HATCH3* (for 70-9-49). The mutations were all either single base pair insertions or deletions, which are predicted to cause frameshifts and premature stop codons (Figure 3.2).

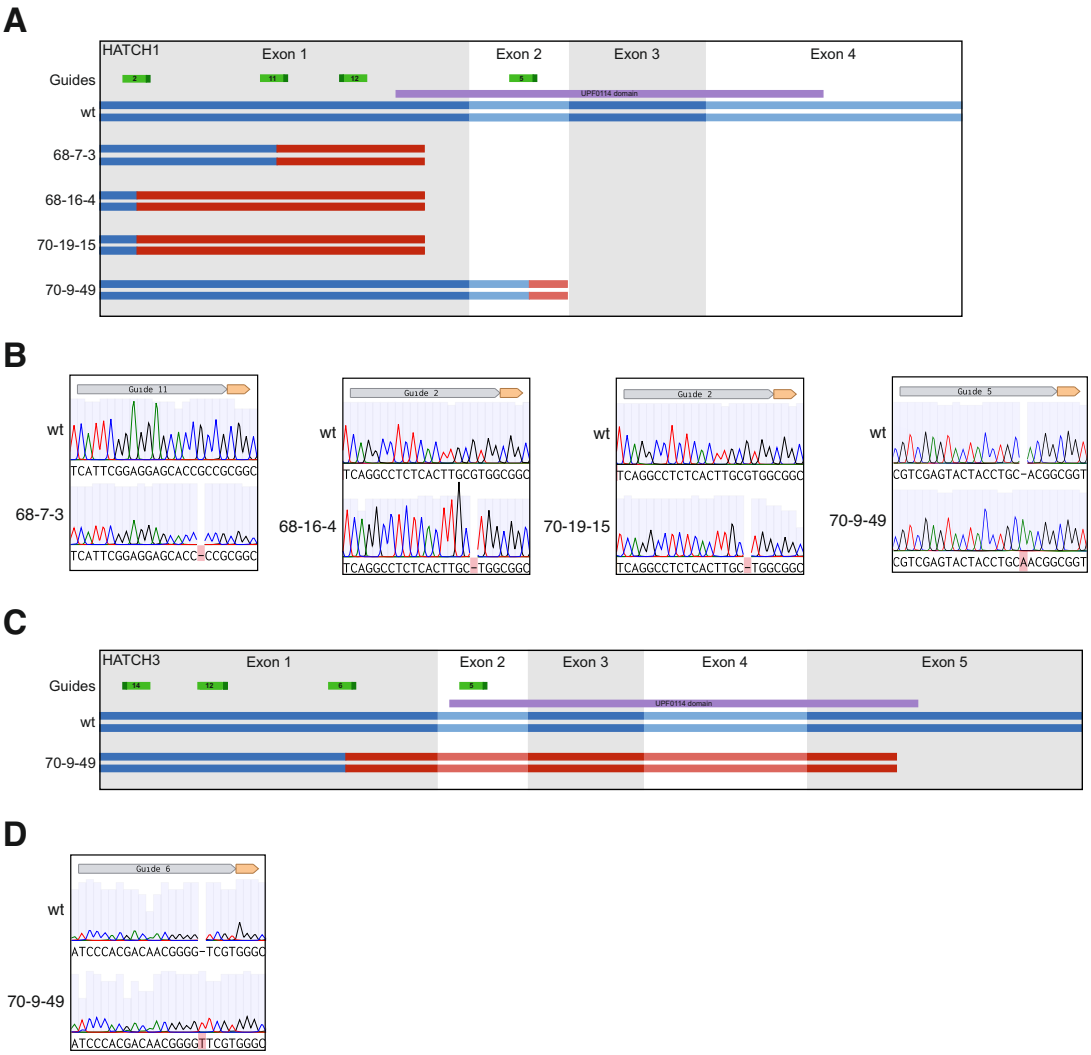


Figure 3.2: Mutations in *HATCH1* and *HATCH3* observed in T_1 plants taken forward for further study. **A**) Cartoon representations of *HATCH1* coding sequences (CDS) in wild-type and selected mutant lines (68-7-3, 68-16-4, 70-9-15, 70-9-49). **B**) Sanger sequencing traces comparing *HATCH1* sequences of wild-type vs mutant lines at the predicted binding site for the CRISPR sgRNA that has induced the mutation. **C**) Cartoon representations of *HATCH3* CDS in wild-type and selected mutant line (70-9-49). **D**) Sanger sequencing traces comparing *HATCH3* sequences of wild-type vs mutant line at the predicted binding site for the CRISPR sgRNA that has induced the mutation. For the cartoon representations (**A, C**): Blue indicates regions encoding wild-type amino acid sequence. Red indicates regions encoding frame-shifted amino acid sequence, stopping where premature stop codon occurs. Green indicates predicted sgRNA binding sites. Purple indicates UPF0114 domain.

3.2.2 *hatch1* single- and *hatch1,hatch3* double-knock-out lines do not show significantly different dry matter carbon isotope discrimination or carbon/nitrogen ratios

Phenotypic characterisation of *hatch1* and *hatch1,hatch3* lines, compared to wild-type plants, was carried out in the T_2 generation. First, to confirm whether the knock-out plants were still performing C_4 photosynthesis, or if the absence of HATCH1 (and HATCH3) led to a reversion to C_3 photosynthesis, dried leaf material from mature plants was subjected to carbon isotope analysis. C_3 plants, which fix CO_2 directly using rubisco, have a ratio of ^{13}C to ^{12}C ($\delta^{13}C$) of between -20 and -37 ‰[215]. In contrast, C_4 plants have $\delta^{13}C$ values between -10 and -16 ‰[216]. This is because they carry out primary CO_2 fixation using PEPC, which discriminates less strongly against ^{13}C than rubisco, resulting in a higher ratio of ^{13}C to ^{12}C (although still lower than atmospheric)[217]. Plants belonging to all four mutant lines had carbon isotope signatures very similar to wild-type, and well within the expected range for a C_4 plant (Figure 3.3 A). Thus, knocking out *HATCH1* and *HATCH3* did not lead to a reversion to C_3 photosynthesis. Furthermore, the ratio of carbon to nitrogen of the dried leaf samples was also not significantly different between any of the mutant lines and wild-type (Figure 3.3 B). This indicates a lack of large-scale disruption to cellular carbon or nitrogen metabolism in these leaves.

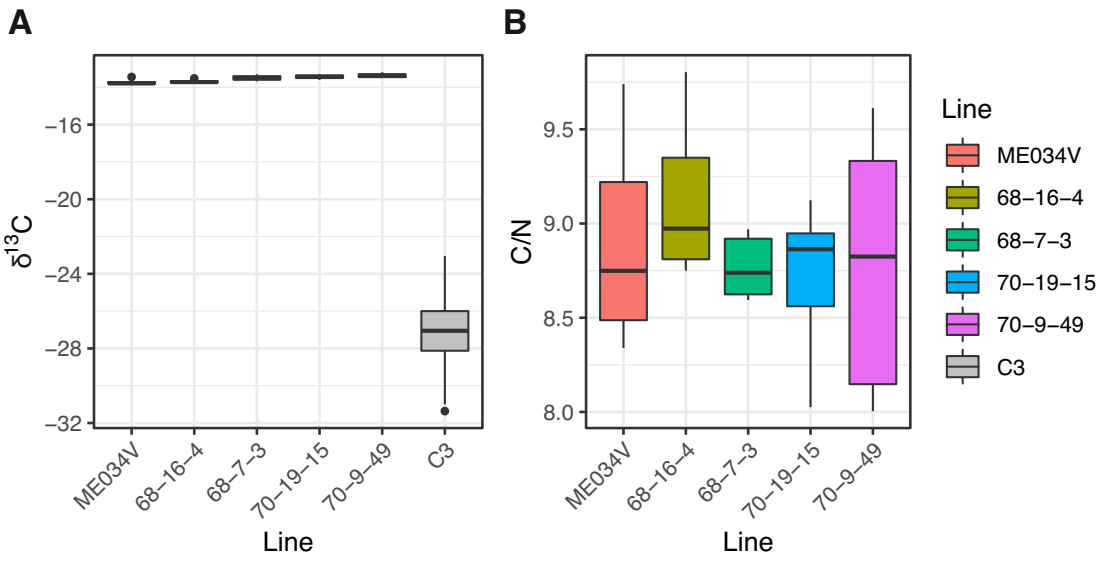


Figure 3.3: Isotope analysis of *hatch1* single- and *hatch1,hatch3* double-knock-out *S. viridis* lines. **A**) $\delta^{13}C$ ratio, calculated according to the Pee Dee Belemnite (PDB) standard. A range of C_3 values are shown for comparison (taken from Kohn (2010)[215]) **B**) Carbon / Nitrogen ratio. For both measurements, a leaf sample for 6 plants were taken per line.

3.2.3 *hatch1* single- and *hatch1,hatch3* double-knock-out lines show inhibited growth compared to wild-type

Wild-type and mutant plants were grown under normal greenhouse conditions, with positional randomisation to ensure a lack of positional effects on plant growth. Measurements of various physical traits were regularly taken throughout the plants' lifecycle. It would be expected that a significant disruption to C_4 photosynthesis would reduce the ability of these plants to generate biomass, and thus negatively impact their growth. Figure 3.4 A shows representative plants from each line. Wild-type plants grew faster, as determined by plant height measurements, than plants from all four mutant lines, which showed a similar growth rate (Figure 3.4 B). At 20 days after germination,

wild-type plants were taller (from soil to base of youngest fully expanded leaf on main stem) (Ø: 11.17 cm ± 1.29 (SD, N = 15) compared to all mutant lines (68-16-4 Ø: 7.87 cm ± 0.86 (SD, N = 15), 68-7-3 Ø: 8.03 cm ± 0.72 (SD, N = 15), 70-9-49 Ø: 8.20 cm ± 0.84 (SD, N = 15), 70-19-15 Ø: 8.30 cm ± 0.53 (SD, N = 15) (Figure 3.4 C). Post-hoc comparison using Dunnett's test showed that these height values compared to wild-type were significantly different for all mutant lines (68-16-4: $p<1x10^{-10}$, 68-7-3: $p<1x10^{-10}$, 70-9-49: $p<1x10^{-10}$, 70-19-15: $p<1x10^{-10}$). There was no significant difference in any of the growth measurements between *hatch1* single- and *hatch1,3* double-knock-out lines. The fact that knocking out *HATCH1* leads to reduced growth rate is consistent with a role for HATCH1 in the C_4 cycle. However, as additionally knocking out *HATCH3* does not lead to a further reduction in growth rate on top of that observed in the *hatch1* single mutant, the hypothesis that residual C_4 cycle activity in *hatch1* mutant lines could be explained by residual HATCH3 activity is not supported.

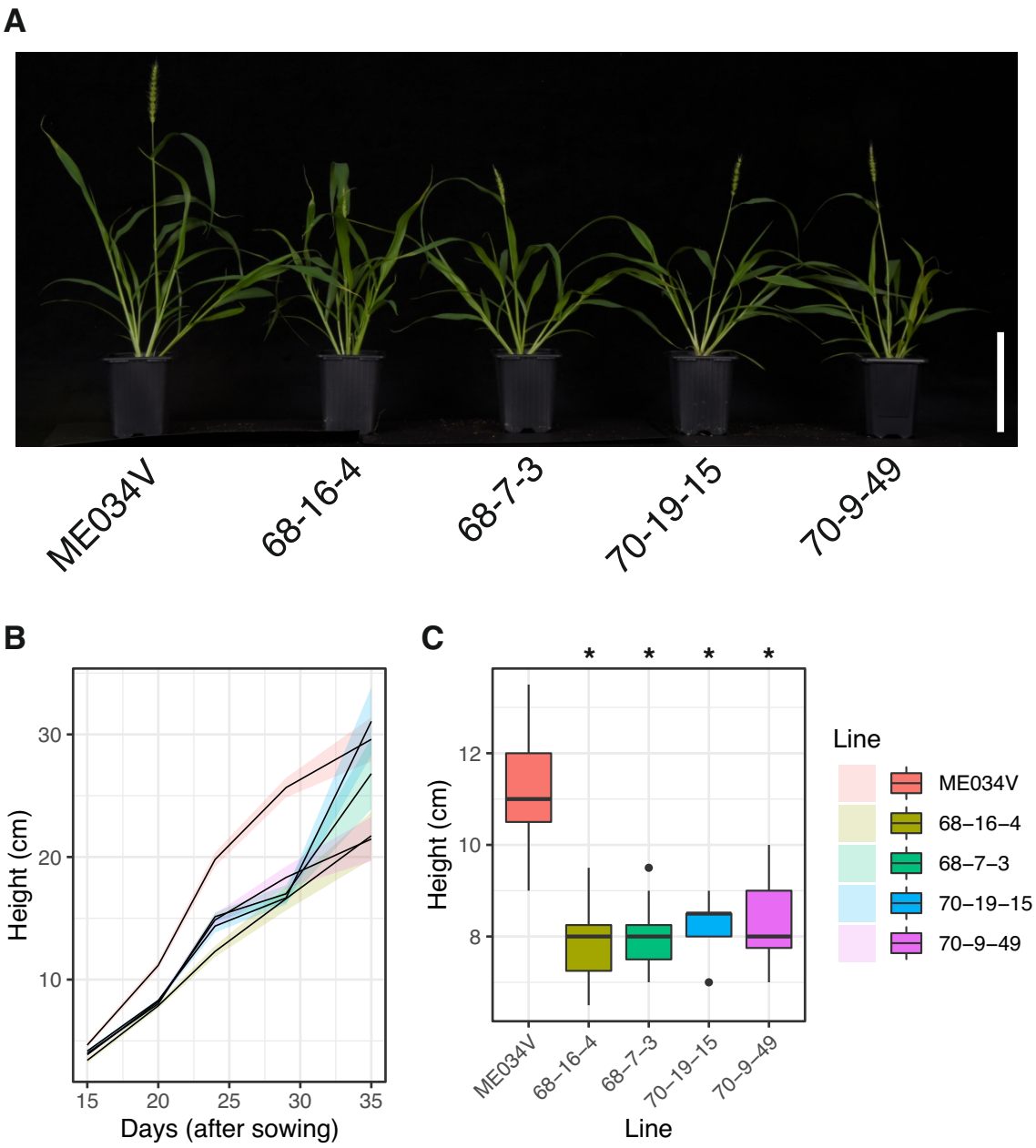


Figure 3.4: Plant height measurements of *hatch1* single- and *hatch1,hatch3* double-knock-out *S. viridis* lines. **A**) Photo of a representative plant from each line, taken at 23 days after germination. Scale bar = 10 cm. **B**) Plant height measurements (cm) (to base of youngest fully expanded leaf on main stem) over time. **C**) Plant height measurements (cm) at 20 days after germination. Asterisks indicate significant differences ($p \leq 0.05$) for mutant lines compared to wild-type according to post-hoc Dunnett's test.

Other plant growth measurements, including tiller number (Figure 3.5 A), date of panicle emergence (Figure 3.5 B), and flag leaf length and width on this date (Figure 3.5 C, D) were not consistently different between wild-type and mutant lines according to post-hoc comparisons using Dunnett's test.

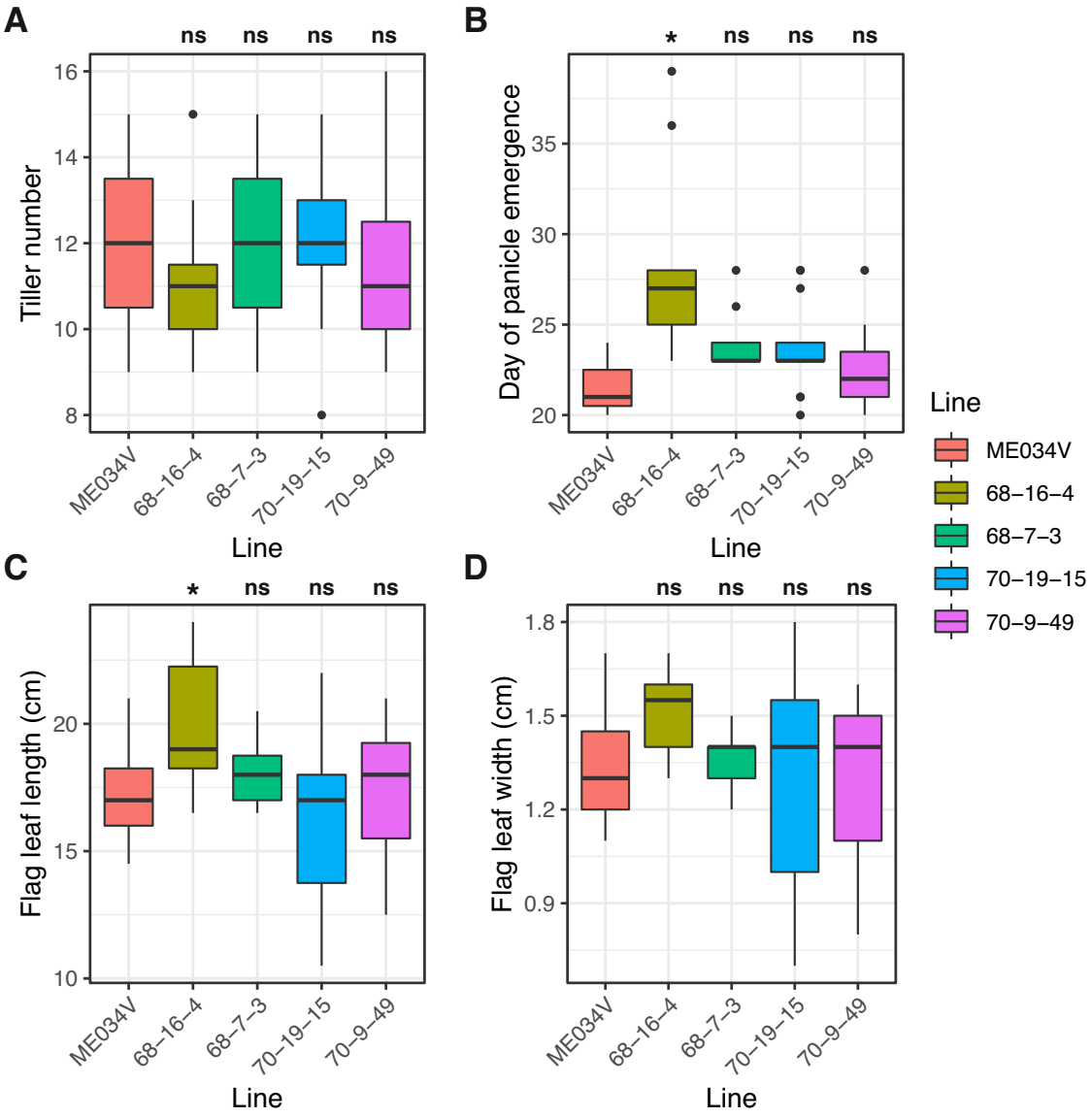


Figure 3.5: Other plant growth measurements of *hatch1* single- and *hatch1,hatch3* double-knock-out *S. viridis* lines. **A**) Plant tiller number at 29 days after germination. **B**) Day (after germination) of panicle emergence. **C**) and **D**) Length and width (cm) respectively of youngest fully expanded leaf on main stem on date of panicle emergence. Asterisks indicate significant differences ($p \leq 0.05$) for mutant lines compared to wild-type according to post-hoc Dunnett's test. ns = not significant.

3.2.4 *hatch1* single- and *hatch1,hatch3* double-knock-out lines show reduced CO_2 assimilation

To confirm whether *hatch1* and/or *hatch1,hatch3* knock-outs caused a reduction in CO_2 assimilation, as might be expected given disruption to the NADP-ME subtype C_4 cycle, LI-6800 Portable Photosynthesis Systems were used to measure a number of photosynthetic gas-exchange and fluorescence parameters at varying CO_2 concentrations. The net rate of CO_2 assimilation (A) was lower across the range of CO_2 concentrations tested in all mutant lines compared to wild-type (Figure 3.6 A). At approximately ambient CO_2 concentration (400 ppm) wild-type plants showed higher A values (Ø: 34.101 μmol m^{-2} s^{-1} ± 4.851 (SD, N = 10)) compared to all the mutant lines (68-16-4 Ø: 26.233 μmol m^{-2} s^{-1} ± 4.523 (SD, N = 10), 68-7-3 Ø: 27.900 μmol m^{-2} s^{-1} ± 5.412 (SD, N = 10), 70-9-49 Ø: 23.479 μmol m^{-2} s^{-1} ± 3.438 (SD, N = 10), 70-19-15 Ø: 24.584 μmol m^{-2} s^{-1} ± 6.850 (SD, N = 10) (Figure 3.6 B). Post-hoc comparison using Dunnett's test showed that these A values compared to wild-type were significantly different for all mutant lines (68-16-4: $p=0.00386$, 68-7-3: $p=0.02861$, 70-9-49: $p<0.001$, 70-19-15: $p<0.001$). At the maximum CO_2 concentration tested (1600 ppm) the difference in A was even greater between wild-type plants (Ø: 38.367 μmol m^{-2} s^{-1} ± 4.343 (SD, N = 10)) and all the mutant lines (68-16-4 Ø: 29.115 μmol m^{-2} s^{-1} ± 2.781 (SD, N = 10), 68-7-3 Ø: 29.062 μmol m^{-2} s^{-1} ± 4.070 (SD, N = 10), 70-9-49 Ø: 26.980 μmol m^{-2} s^{-1} ± 3.641 (SD, N = 10), 70-19-15 Ø: 28.506 μmol m^{-2} s^{-1} ± 6.657 (SD, N = 10) (Figure 3.6 C). Again, post-hoc comparison using Dunnett's test showed these differences in A compared to wild-type to be significant for all mutant lines (68-16-4: $p=0.00013$, 68-7-3: $p=0.00018$, 70-9-49: $p<0.0001$, 70-19-15: $p<0.0001$). Over the same range of CO_2

concentrations, stomatal water conductance (gsw) was only slightly lower in the mutant lines compared to wild-type (Figure 3.6 D).

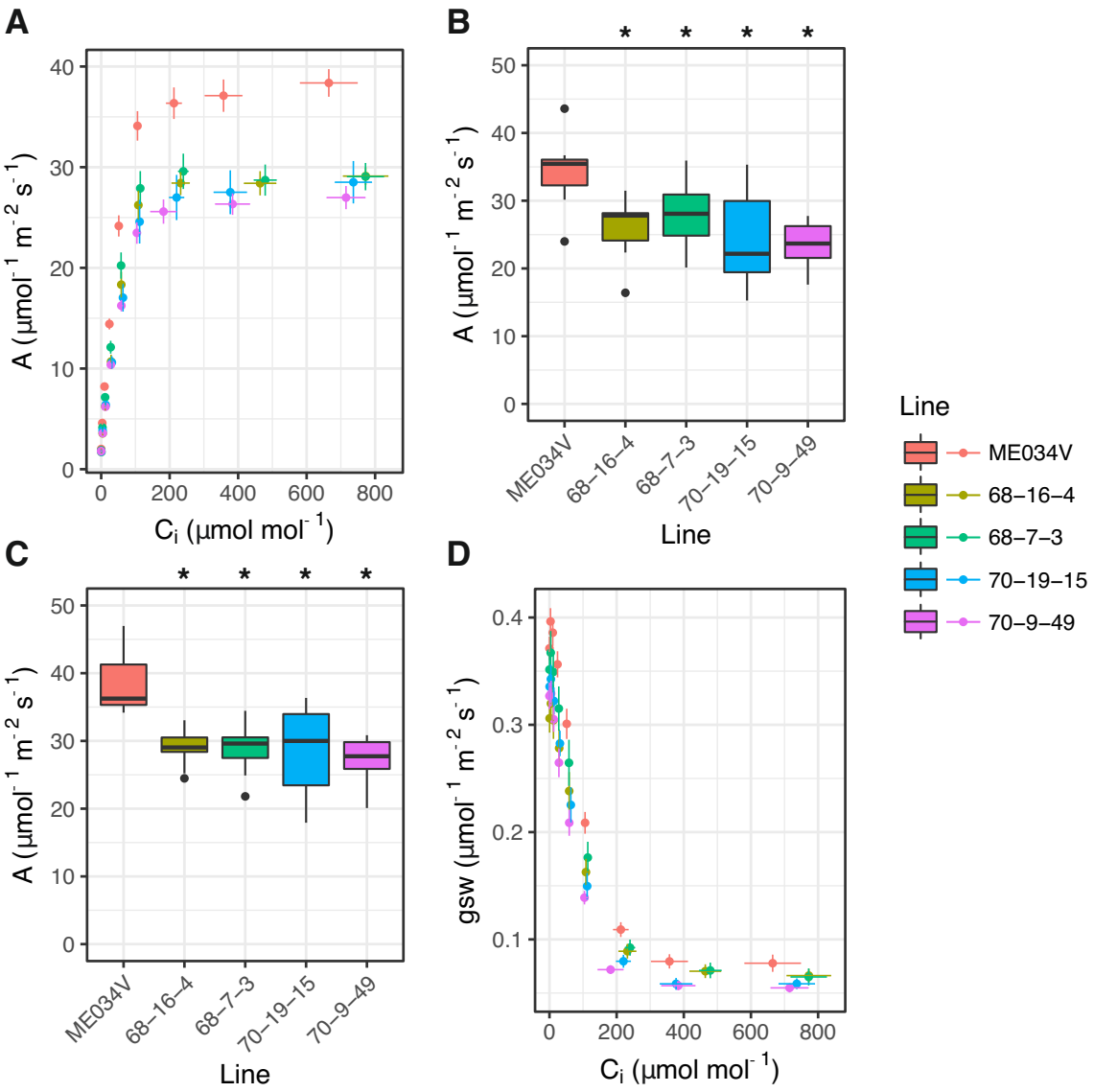


Figure 3.6: Characterisation of gas-exchange of *hatch1* single- and *hatch1,hatch3* double-knock-out *S. viridis* lines under varying CO_2 concentrations. **A**) Net CO_2 assimilation rate (A) over a range of intercellular CO_2 concentrations (C_i). Each point represents the average of 10 plants per line. Error bars represent standard error of the mean. **B**) Net CO_2 assimilation rate at approximately atmospheric CO_2 concentration (400 ppm). Asterisks indicate significant differences ($p \leq 0.05$) for mutant lines compared to wild-type according to post-hoc Dunnett's test. **C**) Net CO_2 assimilation rate at maximum tested CO_2 concentration (1600 ppm). **D**) Stomatal water conductance (gsw) over a range of intercellular CO_2 concentrations (C_i). Each point represents the average of 10 plants per line. Error bars represent standard error of the mean.

To further interrogate the impact of knocking out *hatch1* or *hatch1* and *hatch3* on photosynthesis, the A-C_i curve data was fitted with a biochemical model of C_4 photosynthesis[185,218]. This model allows for the estimation of various photosynthetic parameters, including the maximum activities of PEPC (V_{pmax}) and rubisco (V_{cmax}). Wild-type plants had higher V_{pmax} values (Ø: 194.469 µmol m^{-2} s^{-1} ± 52.276 (SD, N = 10)), compared to the single mutant lines 68-16-4 (Ø: 131.970 µmol m^{-2} s^{-1} ± 40.709 (SD, N = 10)), 68-7-3 (Ø: 152.837 µmol m^{-2} s^{-1} ± 41.044 (SD, N = 10)) and 70-19-15 (Ø: 120.739 µmol m^{-2} s^{-1} ± 39.727 (SD, N = 10)), and double mutant line 70-9-49 (Ø: 141.757 µmol m^{-2} s^{-1} ± 53.124 (SD, N = 10)). (Figure 3.7 A). Post-hoc comparison using Dunnett's test showed that these differences in V_{pmax} compared to wild-type were significant for the single mutant lines 68-16-4 (p=0.01351), and 70-19-15 (p=0.00291), and double mutant line 70-9-49 (p=0.04509), but not single mutant line 68-7-3 (p=0.14750). Similarly, wild-type plants had higher V_{cmax} values (Ø: 44.486 µmol m^{-2} s^{-1} ± 6.734 (SD, N = 10)), compared to the single mutant lines 68-16-4 (Ø: 36.511 µmol m^{-2} s^{-1} ± 2.803 (SD, N = 10)), 68-7-3 (Ø: 36.687 µmol m^{-2} s^{-1} ± 6.558 (SD, N = 10)) and 70-19-15 (Ø: 36.340 µmol m^{-2} s^{-1} ± 8.402 (SD, N = 10)), and double mutant line 70-9-49 (Ø: 33.060 µmol m^{-2} s^{-1} ± 5.379 (SD, N = 10)) (Figure 3.7 B). Post-hoc comparison using Dunnett's test showed that these differences in V_{cmax} compared to wild-type were significant for the single mutant lines 68-16-4 (p=0.01347), and 70-19-15 (p=0.00301), and double mutant line 70-9-49 (p=0.04501), but not single mutant line 68-7-3 (p=0.14747). Thus, *HATCH1* knock-out disrupts the activity of both carboxylation enzymes involved in C_4 photosynthesis.

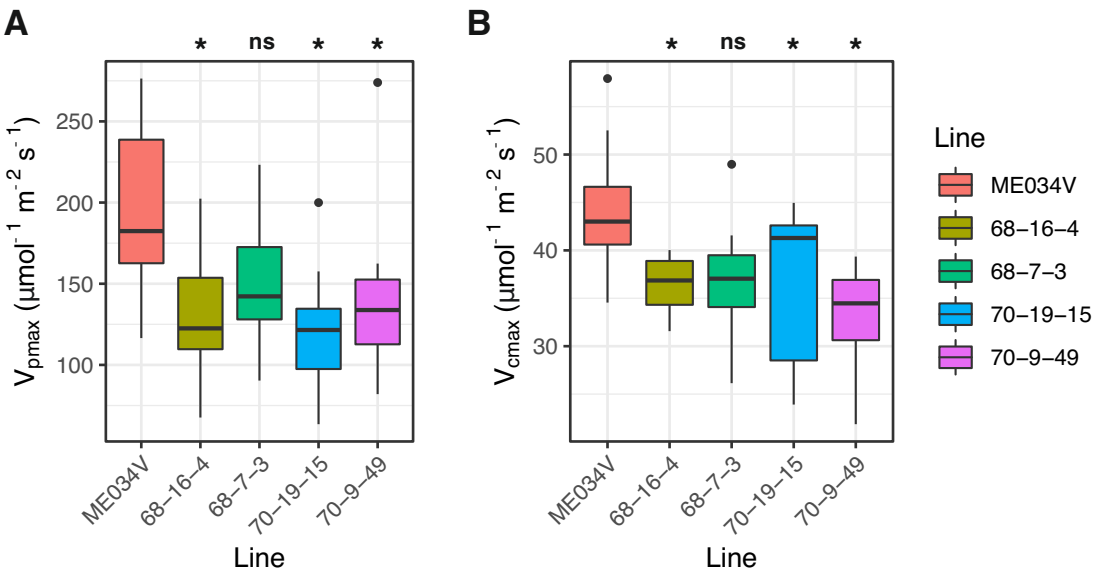


Figure 3.7: Estimation of the rate of the two carboxylation reactions in the C_4 cycle of *hatch1* single- and *hatch1,hatch3* double-knock-out *S. viridis* lines. **A**) Estimated maximum PEPC activity (V_{pmax}). **B**) Estimated maximum rubisco activity (V_{cmax}). Asterisks indicate significant differences ($p \leq 0.05$) for mutant lines compared to wild-type according to post-hoc Dunnett's test. ns = not significant.

The value of C_i (intercellular CO_2 concentration) where photosynthetic CO_2 assimilation and respiratory CO_2 production are in balance is called the CO_2 compensation point (Γ). C_4 plants concentrate CO_2 around rubisco in order to minimise photorespiration. Therefore, C_4 plants are characterised by very low CO_2 compensation points (~0-10 μmol mol^{-1}), compared to C_3 plants (~25-100μmol mol^{-1})[219]. The experimentally collected A-C_i curves, did not reach low enough CO_2 concentrations for A<0 for any of the plants. Thus, the x-intercept (C_i value for which A=0, i.e. Γ), could not be directly determined. Instead, the modelled A-C_i curves generated by fitting the steady-state C_4 photosynthetic model allowed for modelled estimates of Γ. Strangely, the calculated CO_2 compensation points

were all negative (Figure 3.8 A). This can be further illustrated by looking at the low CO_2 A-C_i curve measurements, which clearly intercept the y axis before the x axis as C_i decreases (Figure 3.8 B). As a negative C_i value is impossible, this puzzling result is presumably explained by a minor mis-calibration within the LI-6800. It is clear however that all mutant lines have very low, C_4-like Γ values. This further supports the conclusion that *HATCH1* (and *HATCH3*) knock-outs have not caused a reversion to C_3 photosynthesis, and the concomitant increase in photorespiration and CO_2 compensation point.

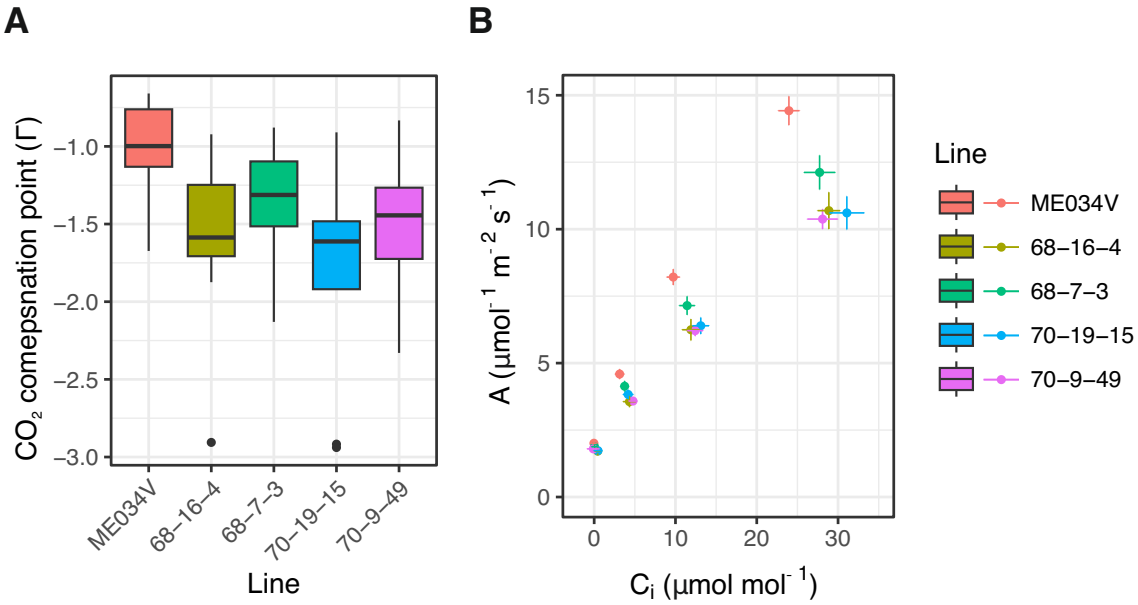


Figure 3.8: Estimation of CO_2 compensation point (Γ) of *hatch1* single- and *hatch1,hatch3* double-knock-out *S. viridis* lines. **A**) Estimated Γ values. **B**) Lower C_i end of the A-C_i curve previously shown in Figure 3.6 A.

Fluorescence measurements taken across the range of CO_2 values also allowed for the estimation of several parameters relating to the photochemistry of photosynthesis. Electron transport rate (ETR) was found to be lower in the mutant lines compared to wild-type across the range of CO_2 concentrations tested (Figure 3.9 A). As expected, this was associated with higher levels of non-photochemical quenching (NPQ) in the mutant lines (Figure 3.9 B), and a lower quantum efficiency of PSII (ΦPSII) (Figure 3.9 C).

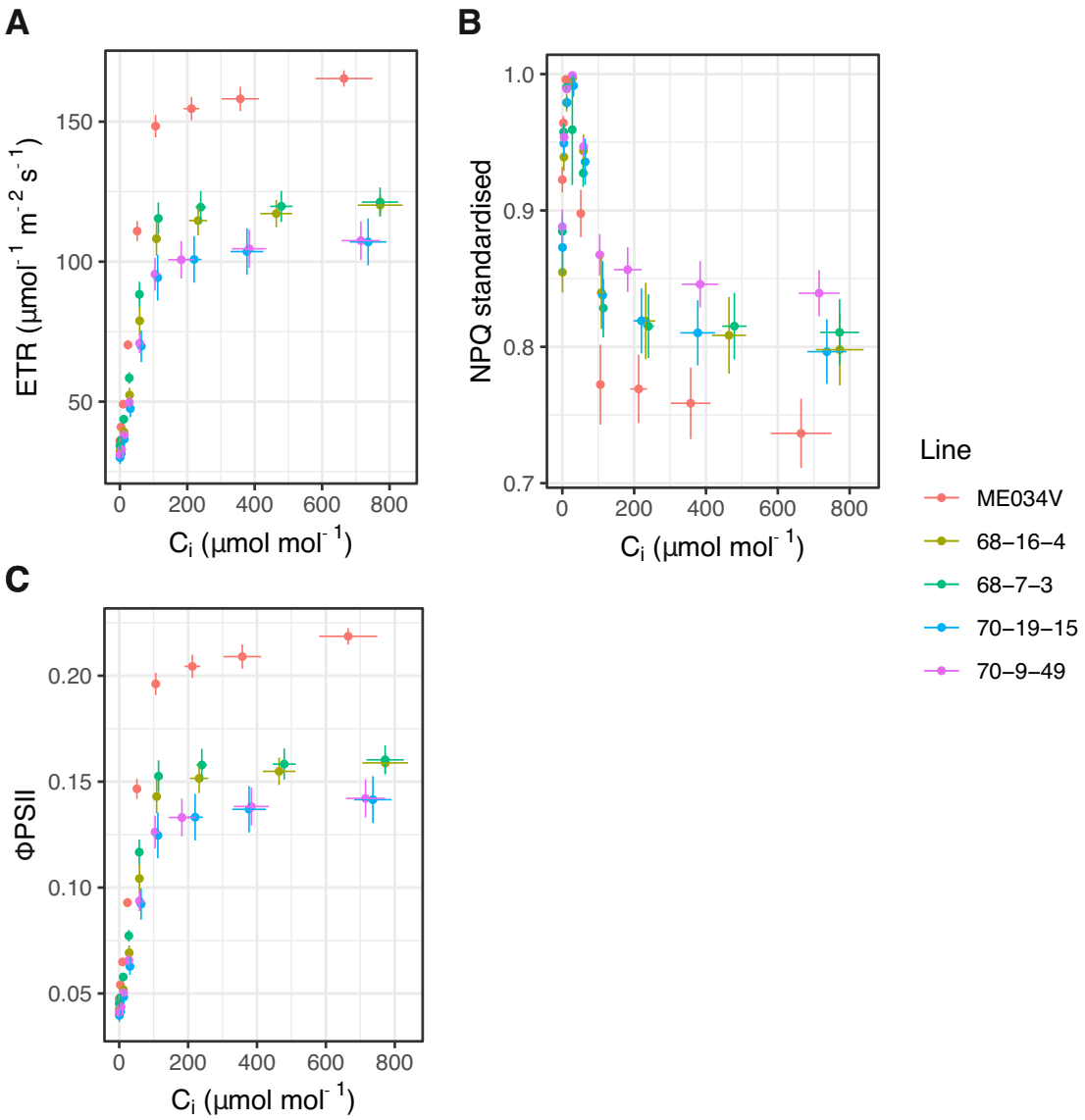


Figure 3.9: Characterisation of photochemistry of *hatch1* single- and *hatch3* double-knock-out *S. viridis* lines under varying CO_2 concentrations. **A**) Electron transport rate (ETR), **B**) Non-photochemical quenching (NPQ), and **C**) ΦPSII, each over a range of intercellular CO_2 concentrations (C_i). Each point represents the average of 10 plants per line. Error bars represent standard error of the mean.

In order to investigate the effect of varying light intensity on photosynthetic rate in mutant plants compared to wild-type, light response curves were also conducted (Figure

3.10 A). At low light intensities there was no significant difference in net CO_2 assimilation rate, and indeed the light compensation point (light intensity at which photosynthesis and respiration are in balance) was not significantly different between wild-type and mutant lines (Figure 3.10 C). However, with increasing light intensity comes an increasing gap between net CO_2 assimilation rates of wild-type and mutant plants from all lines. At the maximal photosynthetic rate at the maximum light intensity tested (1800 μmol m^{-2} s^{-1} PPFD), wild-type plants showed higher A values (Ø: 39.785 μmol μmol m^{-2} s^{-1} ± 4.376 (SD, N = 10)) compared to all the mutant lines (68-16-4 Ø: 28.788 μmol m^{-2} s^{-1} ± 1.947 (SD, N = 10), 68-7-3 Ø: 29.178 μmol m^{-2} s^{-1} ± 4.068 (SD, N = 10), 70-9-49 Ø: 27.938 μmol m^{-2} s^{-1} ± 6.440 (SD, N = 10), 70-19-15 Ø: 29.270 μmol m^{-2} s^{-1} ± 4.935 (SD, N = 10) (Figure 3.10 B). Post-hoc comparison using Dunnett's test showed that these A values compared to wild-type were significantly different for all mutant lines (68-16-4: $p<0.0001$, 68-7-3: $p<0.0001$, 70-9-49: $p<0.0001$, 70-19-15: $p<0.0001$). In agreement with the A-C_i curve data, the mean maximal photosynthetic rate is ~40 μmol m^{-2} s^{-1} for wild-type plants, but only ~30 μmol m^{-2} s^{-1} for mutant plants from all lines. As expected, given that photosynthesis, but not respiration, has been disrupted, dark respiration rate (R_d) was unaffected in the mutant lines compared to wild-type (Figure 3.10 D). Thus, *HATCH1* knock-out has no effect on respiration, or on photosynthesis at low light intensities, when photosynthesis is presumably limited by light rather than by the operation of the C_4 cycle. However, as light intensity increases and becomes non-limiting, *HATCH1* knock-out plants are unable to sustain the photosynthetic rates observed in wild-type.

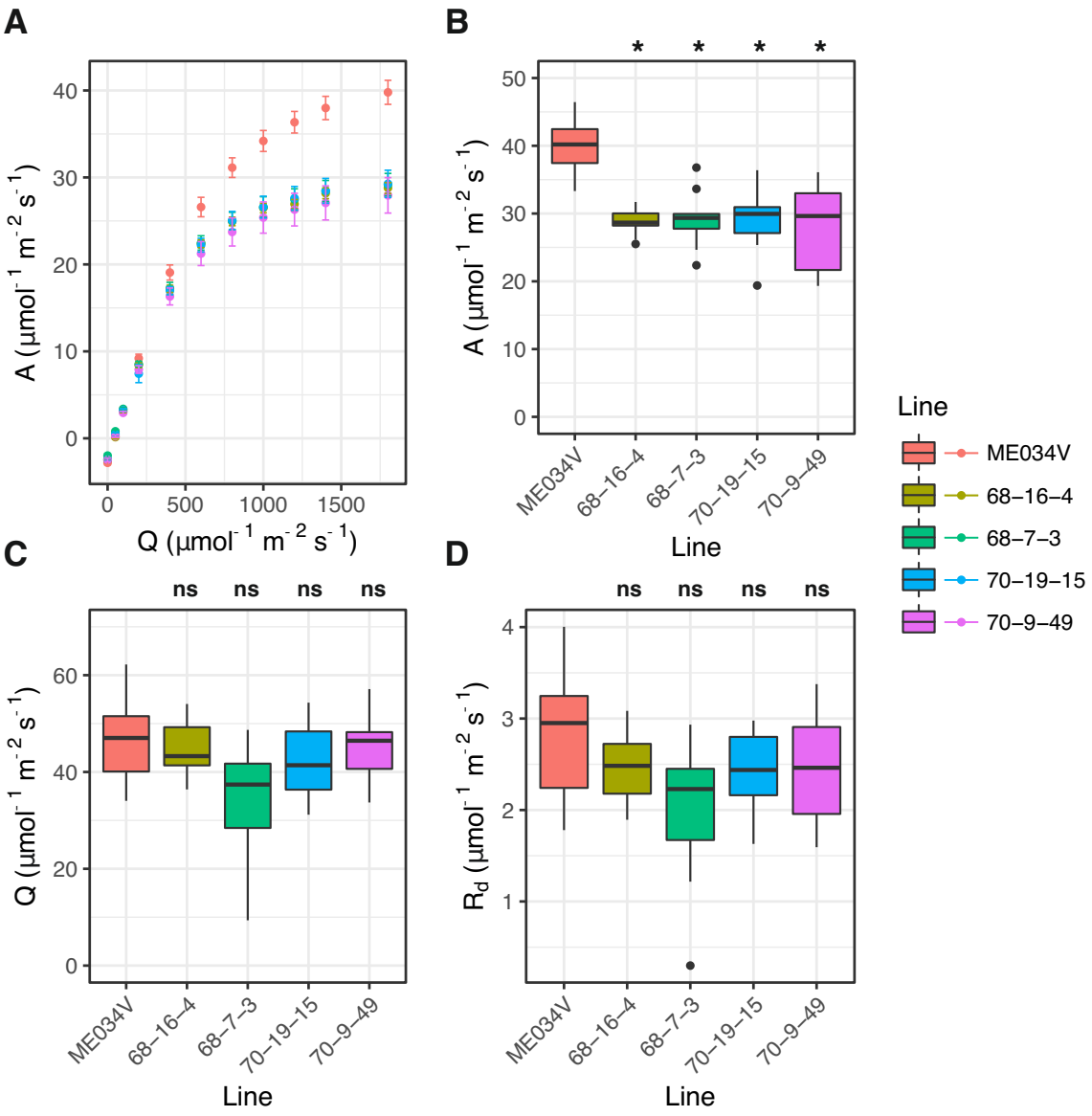


Figure 3.10: Characterisation of gas-exchange of *hatch1* single- and *hatch1,hatch3* double-knock-out *S. viridis* lines under varying light intensities. **A**) Light response curve. Net CO_2 assimilation rate (A) is plotted against light intensity (Q). Each point represents the average of 10 plants per line. Error bars represent standard error of the mean. **B**) Net CO_2 assimilation rate at maximum tested light intensity (1800 µmol m^{-2} s^{-1} PPFD). **C**) Light compensation point (light intensity at which photosynthesis is equal to respiration, and so A=0) **D**) Dark respiration rate (R_d) (-A when Q=0). Asterisks indicate significant differences ($p \leq 0.05$) for mutant lines compared to wild-type according to post-hoc Dunnett's test. ns = not significant.

3.2.5 Transcriptomic analysis of *hatch1* single- and *hatch1,hatch3* double-knock-out lines reveals compensatory transcriptome changes but no evidence for upregulation of supplementary C_4 cycles

To investigate whether knocking out *HATCH1* and *HATCH3* in *S. viridis* led to changes in the expression of other genes, a transcriptomic analysis was performed on the wild-type and mutant lines using RNA-seq. RNA samples were isolated from pooled central leaf sections and transcript abundance was compared between wild-type and *hatch1* single- and *hatch1,hatch3* double-mutant plants. As we had three independent *hatch1* single-knock-out lines, but only one *hatch1,hatch3* double-knock-out line, I focussed in the analysis on the single mutants, but will comment on differences observed with the double mutant where appropriate. Lists of significantly ($p_{adj} \leq 0.05$) up- and down-regulated genes can be found in appendix A.9.2 tables A.83 and A.84 respectively. In total, 538 genes were significantly up-regulated, and 356 genes were significantly down-regulated, in the *hatch1* single-mutant lines when compared to wild-type.

To investigate the cell-type specific expression patterns of up- and down-regulated genes, genes with significantly mesophyll- or bundle sheath-specific expression patterns were identified from a *S. viridis* cell-type specific RNA-seq dataset[56]. Out of the 538 up-regulated genes, 202 are preferentially expressed in the bundle sheath, and 148 are preferentially expressed in the mesophyll (Figure 3.11 A). Hypergeometric tests showed that the set of up-regulated genes is significantly enriched for bundle sheath-preferential genes ($p = 1.236 \times 10^{-5}$), but is not significantly enriched or depleted for mesophyll-preferential genes. Out of the 356 down-regulated genes, 78 are

preferentially expressed in the bundle sheath, and 179 are preferentially expressed in the mesophyll (Figure 3.11 B). Hypergeometric tests showed that the set of down-regulated genes is significantly depleted for bundle sheath-preferential genes ($p = 0.001182$) and significantly enriched for mesophyll-preferential genes ($p = 1.747 \times 10^{-18}$). Thus, the knock-out of *HATCH1*, which itself has bundle sheath-preferential expression, appears to lead to a broad compensatory upregulation in other bundle sheath-preferential genes, whilst down-regulated genes tend to be those that exhibit mesophyll-preferential expression.

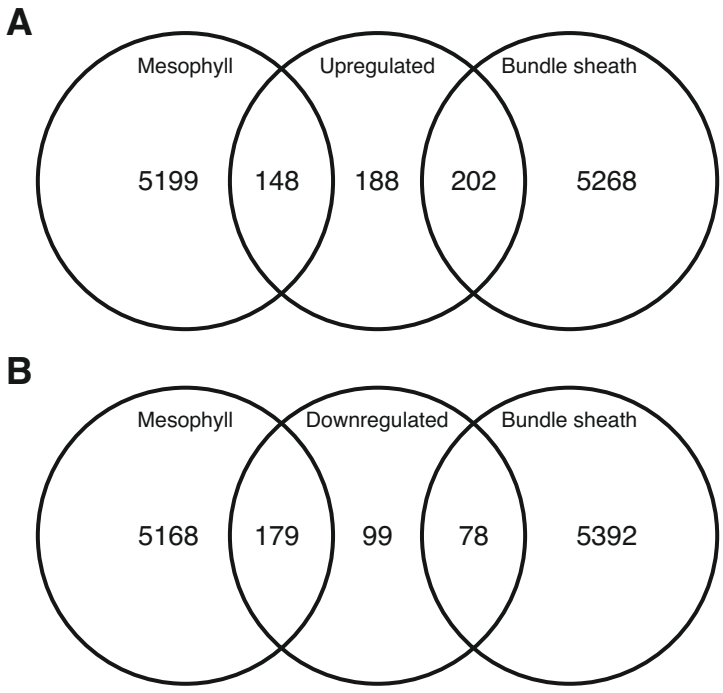


Figure 3.11: Venn diagrams representing the intersections between the sets of mesophyll- and bundle sheath-preferential genes with the sets of **A**) up-regulated genes in *hatch1* single-mutants compared to wild-type, and **B**) down-regulated genes in *hatch1* single-mutants compared to wild-type.

To investigate the sub-cellular localisation patterns of up- and down-regulated genes, the presence of various targeting peptides in the proteins encoded by these genes was predicted using TargetP-2.0[142]. Out of the 538 up-regulated genes, 71 encode a protein that is predicted to contain a chloroplast transit peptide (cTP) (Figure 3.12 A). A hypergeometric test showed that this represents a significant enrichment for chloroplast-localised proteins within the set of up-regulated genes ($p = 0.0004548$). Out of the 356 down-regulated genes, 23 encode a protein that is predicted to contain a cTP (Figure 3.12 B). Hypergeometric tests showed that this does not represent a significant enrichment or depletion for chloroplast-localised proteins within the set of down-regulated genes. Thus, the knock-out of *HATCH1*, which is itself encodes a chloroplast-localised protein, appears to lead to a broad compensatory upregulation in the expression of genes encoding other chloroplast-localised proteins.

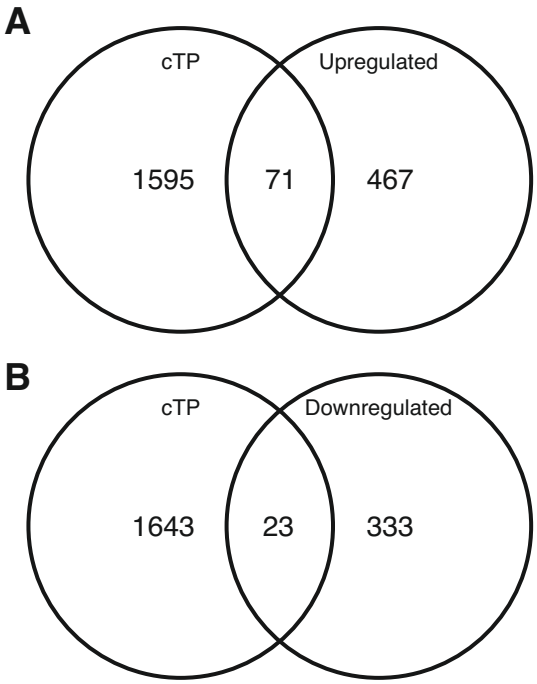


Figure 3.12: Venn diagrams representing the intersections between the set of genes with predicted chloroplast localisation and the sets of **A**) up-regulated genes in *hatch1* single-mutants compared to wild-type, and **B**) down-regulated genes in *hatch1* single-mutants compared to wild-type.

To investigate the broad molecular and biological functions of the up- and down-regulated genes, a GO (gene ontology)-term enrichment analysis was performed. Lists of GO terms significantly ($p \leq 0.05$) enriched within the sets of up- and down-regulated genes can be found in appendix A.9.3 tables A.85 and A.86 respectively. Below is a discussion of some interesting significantly enriched GO terms, as well as some of the individual up- and down-regulated genes within these categories.

Amongst the up-regulated genes the cellular compartment GO term 'chloroplast' (GO:0009507) is significantly enriched (42 genes, expected 26.54, p=0.0011), in agreement with the previously mentioned significant enrichment for genes encoding

proteins containing predicted cTPs. Of particular note amongst the upregulated genes encoding chloroplast targeted proteins is CP12 (chloroplast protein of 12 kDa). CP12 is a Calvin cycle regulatory protein that forms a complex with the Calvin cycle enzymes glyceraldehyde-3-phosphate dehydrogenase and phosphoribulokinase under oxidising (non-photosynthetic) conditions and inhibits their activity[220]. It is possible that reduced NADPH production by NADP-ME in the mutant lines leads to more oxidising conditions, and thus upregulated CP12 represents a mechanism of Calvin cycle inhibition in the mutant plants and may help tune the rate of the Calvin-Benson-Bassham cycle to match the reduced flux through the C_4 cycle. In this context it is also noteworthy that the biological process GO term 'oxidation-reduction process' (GO:0055114) is also significantly enriched (43 genes, expected 28.93, p=0.00479) in the set of up-regulated genes. Among these are multiple NAD(P)-binding Rossmann-fold and oxidoreductase superfamily proteins (Sevir.9G041200, Sevir.9G037100, Sevir.9G301900, Sevir.5G010800, Sevir.9G540700, Sevir.3G419200, Sevir.9G299500), as well as NAD(P)-dependent aldehyde dehydrogenase/oxidase genes (Sevir.2G227800, Sevir.8G063000, Sevir.2G047000, Sevir.1G312800). This indicates a disruption to redox metabolism that would be consistent with altered NADP-ME / Calvin cycle activity in the mutant plants. Furthermore, aldehydes are commonly generated as a result of stress-induced ROS[221], which are likely being produced to a greater extent in mutant plants, as evidenced by their altered photochemistry (see Figure 3.9). Another 'oxidation-reduction process' gene that is upregulated encodes glucose-6-phosphate dehydrogenase (Sevir.9G371200), which catalyses the first step in the oxidative pentose phosphate pathway: a major source of NADPH[222], and which therefore may be upregulated to

compensate for reduced flux through (NADPH-producing) NADP-ME. A final upregulated 'oxidation-reduction process' gene encodes a PGR5-like (proton gradient regulation 5-like) protein (Sevir.2G115600) involved in photosynthetic electron transport. In *A. thaliana*, PGR5 is involved in the regulation of PSI to avoid over-reduction and photoinhibition under conditions where electron sink activities are low, either by facilitating cyclic electron flow around PSI[223] or increasing efficiency of electron transfer to sinks[224]. Upregulation of a protein with this type of activity is consistent with predicted reduced Calvin cycle activity to generate electron sinks (NADP) in mutant lines.

In addition to the GO terms related to oxidation/reduction, several biological process GO terms relating to amino acid metabolism are also significantly enriched in the set of up-regulated genes: e.g. 'alpha-amino acid metabolic process' (GO:1901605) (10 genes, expected 3.84, p=0.00533) and 'alpha-amino acid catabolic process' (GO:1901606) (5 genes, expected 0.85, p=0.00156). Of particular interest are two up-regulated genes encoding chloroplast-localised 2-isopropylmalate synthase (Sevir.7G327500, Sevir.8G028400). This enzyme catalyses a key step in the leucine biosynthesis pathway, a pathway which begins with pyruvate[225]. It is possible that the disruption to chloroplast pyruvate metabolism caused by the knocking out of a chloroplast pyruvate exporter leads to an accumulation of chloroplastic pyruvate and thus an upregulation of pathways, such as leucine biosynthesis, that utilise chloroplastic pyruvate.

Amongst the down-regulated genes, *HATCH1* itself is significantly down-regulated, despite the gene being knocked-out rather than knocked-down. The knock-outs are caused by point mutations that result in frame shifts and premature stop codons, and

thus down-regulation of *HATCH1* expression presumably occurs via a process such as the translation-coupled feedback mechanism nonsense-mediated decay[226]. The biological process GO term 'sucrose biosynthetic process' (GO:0005986) is significantly enriched (3 genes, expected 0.08, $p=5.10x10^{-5}$) in the set of down-regulated genes, consisting of three sucrose-6^F-phosphate phosphohydrolase family proteins (Sevir.2G396500, Sevir.5G174500, Sevir.3G060400). This is consistent with reduced photosynthetic rate leading to reduced sucrose synthesis in mutant plants. These sucrose biosynthesis genes are also contained within the broader biological process GO term 'phosphate-containing compound metabolic process' (GO:0006796), which is also significantly enriched (45 genes, expected 26.52, p=0.00018) in the set of down-regulated genes. Several related molecular function GO terms, including 'ATP binding' (GO:0005524), 'ADP binding' (GO:0043531), and 'kinase activity' (GO:0016301) are also significantly enriched in the set of down-regulated genes. This may represent a response to reduced cellular ATP levels resulting from reduced photosynthetic rates.

As well as a broad analysis of up- and down-regulated genes across the whole genome, we also wanted to investigate the gene expression patterns of specific genes of interest, particularly those involved in C_4 photosynthesis. Figure 3.13 shows the core genes involved in the NADP-ME subtype C_4 cycle. One interesting pattern is the consistent downregulation of all four genes encoding the first four components of the cycle that act between the point of CO_2 entry into the mesophyll and HATCH1: CA (Sevir.5G247800), PEPC (Sevir.4G143500), OMT1 (Sevir.3G366500), and NADP-MDH (Sevir.6G256400). This suggests a possible mechanism of feed-back regulation whereby reduced flux through HATCH1 leads to the down-regulation of the genes preceding it in

the C_4 cycle. Strangely, the gene encoding PEPC kinase (PPCK) (Sevir.1G244600), which activates PEPC[227], shows significant upregulation in the mutant lines, apparently counter to the minor down-regulation of PEPC itself. Whilst *PPCK* expression in C_4 plants is known to be primarily light-regulated, some degree of negative, metabolic regulation by malate, as is the case in CAM plants, has been suggested[228]. Although we don't know how *HATCH1* knock-out would affect cellular malate levels, it is plausible that reduced flux through the C_4 cycle, particularly through the downregulation of PEPC itself, would lead to decreased malate levels in the mesophyll, which could therefore lead to an increased in *PPCK* expression in the mutant lines. The genes encoding the other core components of the NADP-ME subtype C_4 cycle - NADP-ME (Sevir.5G132500), BASS2 (Sevir.5G250900), PPDK (Sevir.3G253900), PPT (Sevir.6G128500) - do not show consistent alterations in expression between wild-type and mutant lines.

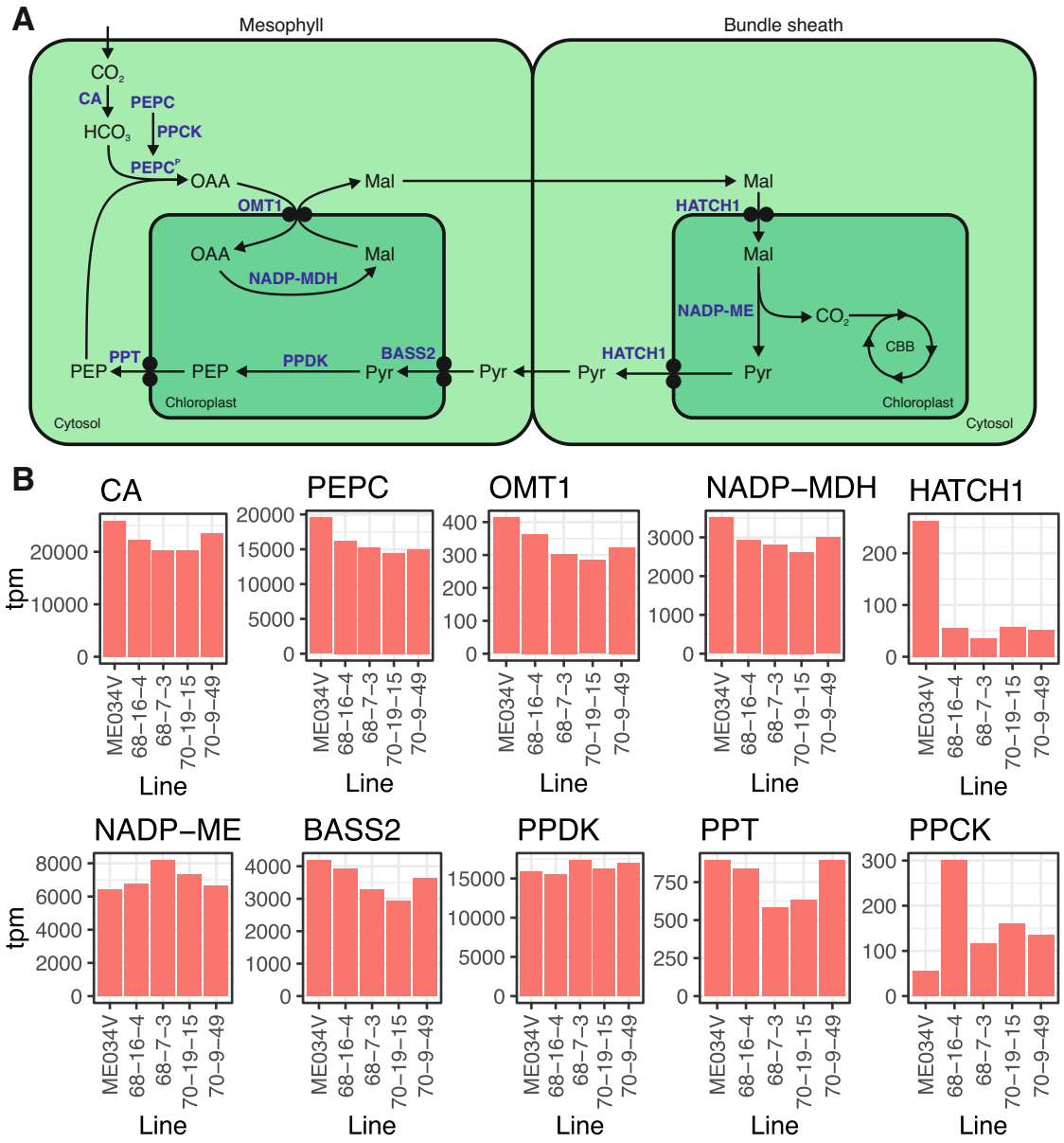


Figure 3.13: Effect of *hatch1* single- and *hatch1,hatch3* double-knock-out on the expression of NADP-ME C_4 cycle genes in *S. viridis*. **A**) Simplified schematic representation of NADP-ME C_4 cycle in *S. viridis*. Metabolites are shown in black. Enzymes/transporters are shown in blue. **B**) Mean transcripts per million (tpm) for relevant NADP-ME C_4 cycle genes across wild-type (ME034V), *hatch1* single knock-out (68-16-4, 68-7-3, 70-19-15), and *hatch1,hatch3* double knock-out *S. viridis* lines. For abbreviations see appendix A.1, table A.1.

In contrast to the consistent changes in the expression of genes encoding components of the NADP-ME C4 cycle above, there is no clear, consistent upregulation of genes encoding components of the other known C_4 cycles: NAD-ME and PCK (Figures 3.14). For both these alternate cycles, in the mesophyll, metabolite flux would be expected to be diverted away from the NADP-ME cycle by the action of DCT and chloroplast-localised aspartate amino transferase (AspAT), which convert OAA into aspartate rather than malate. Both *DCT* genes (Sevir.9G380500, Sevir.2G247400, Sevir.6G194000) and chloroplastic *AspAT* (Sevir.1G359000) show no consistent upregulation in the mutant lines compared to wild-type. Nevertheless, given the downregulation of *OMT1* and *NADP-MDH*, constant expression of *DCT1* and *AspAT* may be sufficient to divert flux from the NADP-ME pathway into the NAD-ME/PCK pathways. Furthermore, a bundle sheath-localised aminotransferase (Sevir.4G020000) is significantly upregulated in mutant plants. The enzyme encoded by this gene is uncharacterised in *Setaria*. However, the *A. thaliana* ortholog (AT1G62960) encodes 1-aminocyclopropane-1-carboxylate (ACC) synthase 10 (ACS10), which rather than ACC synthase activity actually shows aminotransferase activity, with specificity for aromatic amino acids and aspartate (likely reaction: aspartate/aromatic amino acid + 2-oxoglutarate ⇌ oxaloacetate/aromatic 2-oxo-acid + glutamate)[229]. Thus, it is possible that this gene in *S. viridis* encodes a bundle sheath-localised AspAT that could act to convert aspartate coming from the mesophyll back into OAA. Despite this, no other gene encoding a protein thought to be involved in NAD-ME or PCK cycles (NAD-MDH, NAD-ME, UCP, DIC, DTC, MPC, PCK, AlaAT – gene accession numbers found in Appendix A.9.1 table A.82) shows significant upregulation in the *hatch1* single-mutant lines compared to wild-type. Overall therefore, the RNA-seq

data does not support the hypothesis that either auxiliary NAD-ME or PCK cycles are upregulated to compensate for reduced NADP-ME cycle flux. However, given that *Setaria* has previously been shown to have high PCK activity levels even in the absence of high transcript levels[204], it may be that *S. viridis* has sufficient PCK activity to facilitate growth in the absence of HATCH1 and HATCH3. Additionally, there could yet be additional routes for malate to enter and pyruvate to exit the bundle sheath chloroplast that have yet to be characterised.

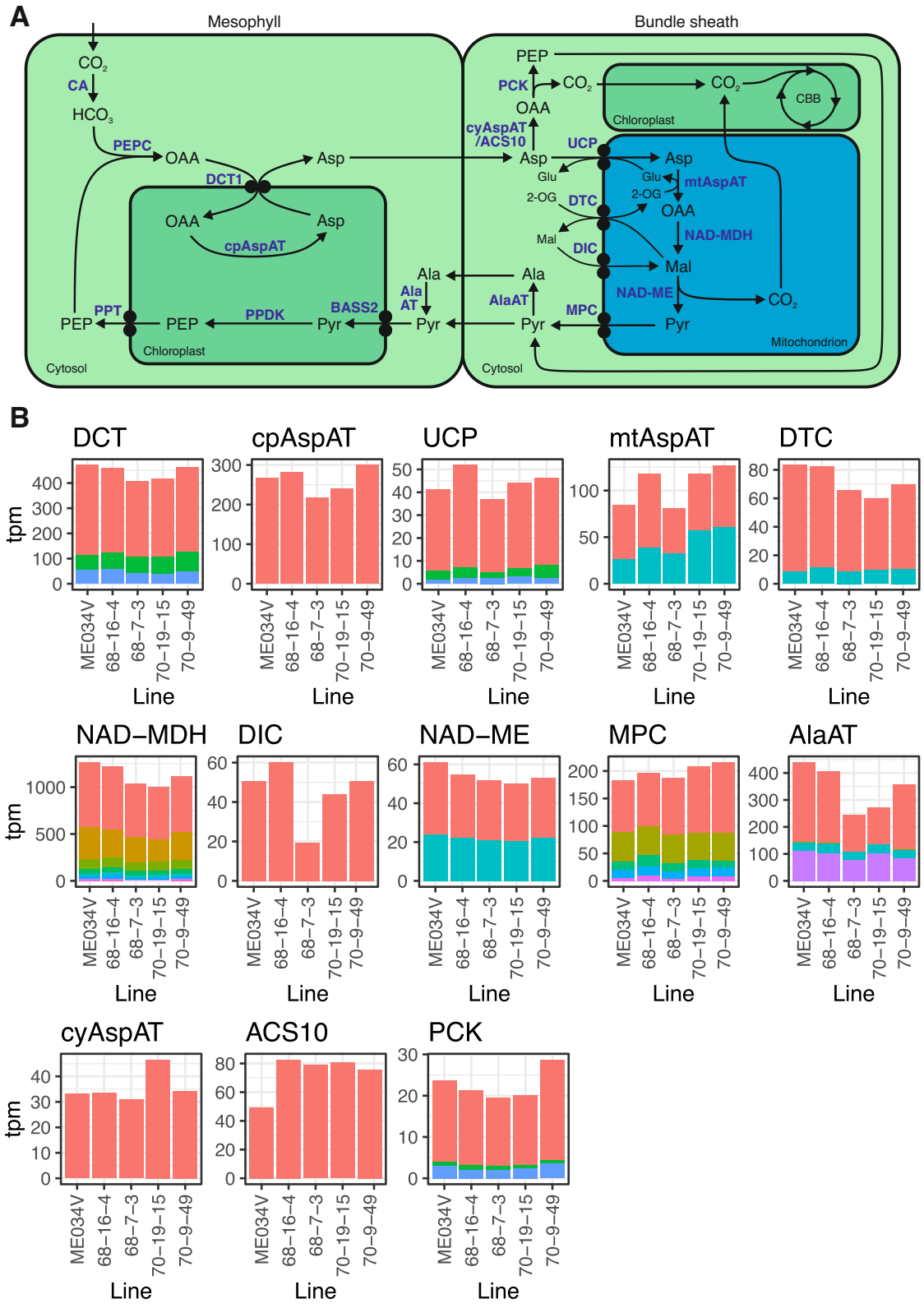


Figure 3.14: Effect of *hatch1* single- and *hatch1,hatch3* double-knock-out on the expression of potential NAD-ME and PCK C_4 cycle genes in *S. viridis*. **A**) Simplified schematic representation of hypothesised NAD-ME and PCK C_4 cycle in *S. viridis*.

Metabolites are shown in black. Enzymes/transporters are shown in blue. **B**) Mean transcripts per million (tpm) for relevant NAD-ME and PCK C_4 cycle genes across wild-type (ME034V), *hatch1* single knock-out (68-16-4, 68-7-3, 70-19-15), and *hatch1,hatch3* double knock-out *S. viridis* lines. Different colour bars represent different paralogs where the C_4-specific paralog has not been identified. For abbreviations see appendix A.1, table A.1.

3.3 Chapter conclusion

In this chapter, I further investigated the requirement for HATCH family proteins for C_4 photosynthesis in *S. viridis*. This was achieved though the generation of three independent *hatch1* single knock-outs lines and one *hatch1,hatch3* double knock-out line.

hatch1 knock-out *S. viridis* lines were shown to have significantly lower growth rates under greenhouse conditions than wild-type plants, which is consistent with the observed lower rates of net CO_2 assimilation under a range of CO_2 concentrations and light intensities. Decreased CO_2 assimilation rate is associated with slightly decreased maximal rates of both carboxylation enzymes, PEPC and rubisco *in planta* estimated from the A/C_i curves. This is consistent with the proposed role for HATCH1 in the import of malate and export of pyruvate from the bundle sheath chloroplast as part of the C_4 cycle. A reduced rate of pyruvate export from the bundle sheath chloroplast would lead to reduced production of PEP in the mesophyll, and thus reduced substrate supply to PEPC, reducing V_{pmax} *in planta*. Similarly, reduced rate of malate import into the bundle sheath chloroplast would lead to reduced NADP-ME-mediated CO_2 release, and thus

reduced CO_2 supply to rubisco, reducing V_{cmax} *in planta*. Whilst this leads to a lower photosynthetic rate, there is no indication of reversion to C_3 photosynthesis. This is presumably because diffusion of atmospheric CO_2 and O_2 into the bundle sheath chloroplast is still effectively prevented, meaning that a) CO_2 fixation is still carried out almost exclusively by PEPC (as evidenced by high, wild-type-like $\delta^{13}C$ values in the mutant plants), and b) photorespiration remains low (as evidence by low, wild-type-like carbon compensation points in the mutant plants). Reduced photosynthetic rate in the mutant lines will be associated with reduced flux through the Calvin cycle, and thus reduced consumption of NADPH and ATP generated by photosynthetic electron transport. This decrease in photochemical electron acceptors would lead to the observed reduced rate of photosynthetic electron transport in the mutant lines. In order to prevent backing-up of the electron transport chain, which can lead to the production of reactive oxygen species and photosystem damage, energy must instead by dissipated by increased non-photochemical quenching (NPQ) in the mutant lines. This decrease in the proportion of the light energy absorbed by chlorophyll that is used in photochemistry, instead being diverted to non-photochemical energy dissipation processes, causes the observed decreased ΦPSII in the mutant lines.

Compared to the *HATCH1* knock-down plants discussed in the previous chapter, the *hatch1* knock-out lines showed more significant growth defects. However, the reduction in CO_2 assimilation rate, as determined by A-C_i and light response curves, was similar between knock-down and knock-out plants (if not slightly more in the knock-down). One explanation for this discrepancy is that Michael Niklaus grew the *HATCH1* knock-down lines under conditions of lower light intensity than I grew the *hatch1* knock-out lines.

Light response curves collected on both sets of plants (Figures 2.4 C and 3.10 A) indicate that at low light intensities there is minimal difference in CO_2 assimilation between knock-down or knock-out plants and wild-type. Thus, it is possible that the knock-out plants, but not the knock-down plants, were grown at sufficient light intensity for a reduced growth phenotype to manifest. On the other hand, it is possible that the reduced growth in *hatch1* knock-out compared to *HATCH1* knock-down plants is explained by maintenance of significant C_4 cycle flux in the *HATCH1* knock-down plants through the remaining ~10% *HATCH1* expression. In this case, the apparently large reduction in CO_2 assimilation rate in *HATCH1* knock-down plants (Figure 2.4 B) is more difficult to explain. However, it should be noted that the A-C_i curves carried out by Michael Niklaus on the *HATCH1* knock-down lines were only N=3, and thus likely not as reliable as the N=10 A-C_i curves I collected on the *hatch1* knock-out lines. Whilst I favour the hypothesis that *hatch1* knock-out leads to reduced C_4 cycle flux compared to knock-down, a side-by-side comparison of knock-down and knock-out plants would be required to confirm this.

The generation of a *hatch1,hatch3* double knock-out line allowed for assessment of the second hypothesis to explain minimal phenotypic disruption to *HATCH1* knock-down plants: compensation of reduced *HATCH1* expression through expression of its parolog, *HATCH3*. There were no significant growth or photosynthetic differences between *hatch1* single- and *hatch1,hatch3* double-knock-out lines, and therefore this hypothesis is not supported.

In order to further investigate the role of *hatch1* (and *hatch3*) knock-out, a transcriptomic analysis was carried out using RNA-seq. No evidence was found for compensatory upregulation of supplementary C_4 cycles (based on NAD-ME or PCK), as had been hypothesised to explain the absence of a growth phenotype in the *HATCH1* knock-down plants. However, it is still possible that some C_4 cycle flux is maintained through even low levels of *PCK* expression, given that potentially high PCK activity that has been observed in the absence of high transcript levels[204]. Despite a lack of evidence for upregulation of supplementary C_4 cycles, *hatch1* knock-out led to significant transcriptome changes in other areas. Broad changes in gene expression patterns, such as compensatory upregulation of bundle sheath- and chloroplast-localised proteins, are consistent with the localisation of *HATCH1*. The upregulation of genes encoding proteins involved in oxidation-reduction processes, including components of the Calvin cycle, oxidative pentose phosphate pathway, and photosynthetic electron transport chain, as well as the downregulation of genes encoding proteins involved in sucrose metabolism and ATP/phosphate metabolism, are consistent with disruption to photosynthesis in the knock-out plants. Furthermore, downregulation of the genes encoding the initial steps of the C_4 pathway are consistent with a disruption to C_4 photosynthesis in particular. Next, in order to further confirm the role of HATCH1 in C_4 photosynthesis, seeds from all four mutant lines have been sent to John Lunn and Regina Feil at the Max Planck Institute for Plant Physiology in Golm. At the time of writing, these plants are being grown for metabolic analysis using pulse-chase labelling, which will allow for a more direct assessment of the changes in metabolite flux caused by *hatch1* knock-out.

The above evidence strongly supports the idea that HATCH1 plays a role in the C_4 cycle in *S. viridis*. However, the fact that *hatch1* (and *hatch1,hatch3*) knock-out plants are still fully viable indicates that HATCH proteins are not solely responsible for C_4 cycle bundle sheath chloroplast malate/pyruvate transport in this species. In maize, *HATCH* genes are preferentially expressed in the mesophyll rather than the bundle sheath, and the transporter(s) responsible for bundle sheath chloroplast malate import and pyruvate export remain unidentified. These transporters missing from our understanding of the C_4 cycle in maize may also be present in *Setaria*, and account for the remaining C_4 cycle flux upon *hatch1* and *hatch3* knock-out in *S. viridis.* In the next chapter I will attempt to identify these missing transporters in maize.

3.4 Materials and methods

3.4.1 *S. viridis* growth conditions

Plants were grown as in section 2.4.14. For T2 plants, which were being grown for phenotypic analysis, positional effects were avoided through positional randomisation of plants. Trays were rotated 90° every day, and the positions of both pots within trays and trays relative to each other were randomised using a random number generator once a week.

3.4.2 Generating CRISPR/Cas9 constructs

Four guides each were designed to target the first half (first and second exons) of *S. viridis HATCH1* and *HATCH3*. First, the online tool CRISPOR[230] was used to generate a list

of candidate guides with high specificity scores and minimal potential off-target sites within the *S. viridis* genome. Cas-OFFinder[231] was used to identify and eliminate candidate guides with potential off-targets with either 5'-NGG-3' or 5'-NAG-3' PAM. ViennaRNA RNAfold[232,233] was used to identify guides which, when combined with the sgRNA backbone, gave a predicted sgRNA secondary structure that retained the three stem loops that are associated with high editing efficiency[234]. The selected guide sequences can be found in Appendix A.7.1, table A.64.

Cloning of level 0, level 1, and level 2 plasmids for plant transformation was carried out according to Hahn et al. (2020)[235]. The selected guide sequences and reverse complement sequences were synthesised as single-stranded oligonucleotides and annealed as described in Hahn et al. (2020)[235]. Assembly of annealed guide oligonucleotides into level 0 acceptor vectors (see Appendix A.7.2, table A.65) was carried out by Golden Gate reactions using the recipe and thermocycler settings at Appendix A.7.2, tables A.66 and A.67. After *E. coli* transformation and plasmid isolation, correct assembly was confirmed by Sanger sequencing (Source Bioscience) using the primer in Appendix A.7.2, table A.68. Assembly of individual guides and tRNA-sgRNA backbones into separate *HATCH1*-targeting and *HATCH3*-targeting polycistronic tRNA-sgRNA arrays within level 1 acceptor vectors (see Appendix A.7.3, tables A.69-71) was carried out by Golden Gate reactions using the restriction enzyme BsaI-HFv2 (NEB), and using the recipe and thermocycler settings at Appendix A.3.1, tables A.11 and A.12. After *E. coli* transformation and plasmid isolation, correct assembly was confirmed by Sanger sequencing (Source Bioscience) using the primers in Appendix A.7.3, table A.72. Assembly of either a single or both level 1 *HATCH1*-targeting and *HATCH3*-targeting

polycistronic tRNA-sgRNA arrays, along with level 1 Hygromycin and spCas9 expression cassettes, into a level 2 acceptor vector (see Appendix A.7.4 tables A.73-75) was carried out by Golden Gate reactions using the restriction enzyme BbsI-HF (NEB), and using the recipe and thermocycler settings at Appendix A.3.1, tables A.11 and A.12. After *E. coli* transformation and plasmid isolation, correct assembly was confirmed by Sanger sequencing (Source Bioscience) using the primers in Appendix A.7.3, table A.76.

3.4.3 Transformation of *S. viridis* ME034V callus and plant regeneration

Agrobacterium-mediated transformation of *S. viridis* ME034V was carried out according to the protocol developed by Joyce Van Eck and Kerry Swartwood. *S. viridis* seed coats were removed by rubbing between two wooden blocks covered in corrugated plastic. All subsequent steps were performed in a flow hood to ensure sterility. De-hulled seeds were sterilised for 15 mins in sterilisation solution (10% sodium hypochlorite, 0.1% TWEEN), washed three times in water, and dried on sterile filter paper. In order to initiate callus generation, seeds were then plated on CIM plates (see Appendix A.2.3, table A.8), 16 per plate, sealed with micropore tape, and stored at 24°C for 4-6 weeks. Once the calli had matured (as indicated by the development of opaque, white regions), mature callus regions were separated from immature gelatinous regions using sterile needles under a dissecting microscope and subcultured onto fresh CIM plates, 25 per plate. The plates were sealed again with micropore tape and stored at 24°C for a further 2-3 weeks. The calli were subcultured in the same way onto fresh CIM plates and stored at 24°C for a further 2 weeks.

Three colonies each of *A. tumefaciens* strain AGL1 transformed with the three CRISPR constructs (pOM068, pOM069, pOM070) were grown overnight in 10 ml LB (28°C, 125 rpm). When the cultures reached OD_{600} = 0.6, the cells were collected by centrifugation (8000 rpm, 20°C, 10 mins). The supernatant was removed and the cells were resuspended in 10 ml liquid CIM (Appendix A.2.3, table A.8), and acetosyringone and pluronic F-68 (Sigma-Aldrich) were added to final concentrations of 0.2 mM and 0.1% respectively.

For each of the three transformations, around 100 mature *S. viridis* calli were transferred to a 50 ml tube, completely covered with 10 ml agrobacterium suspension, and incubated for 5 mins at room temperature with occasional rocking. The agrobacterium suspension was removed from the calli, and the calli were transferred onto sterile filter paper on top of fresh CIM plates. The plates were sealed with parafilm and stored in the dark at 22°C for 3 days. To select for transformants, the calli were subcultured onto fresh CIM selective plates (Appendix A.2.3, table A.8), 15 per plate. The plates were sealed with micropore tape and stored in the dark at 24°C for 16 days. To continue selection for transformants, and induce root and shoot development, the calli were subcultured onto PRM selective plates (Appendix A.2.3, table A.9), 12 per plate, with each callus slightly pushed down into the medium. The plates were sealed with micropore tape and incubated in the light (16-hour photoperiod) at 24°C for 2-3 weeks. Every 2-3 weeks developing shoots were transferred to RM selective plates (see Appendix A.2.3, table A.10), and the remaining callus was subcultured onto fresh PRM selective plates. This process was repeated three times. After two weeks on RM selective, surviving plantlets with well-developed root systems were transferred to soil.

3.4.4 Plant DNA extraction

2 cm^2 of leaf tissue was taken per plant, and placed into a deep 96-well plate containing a metal bead. 500 μl DNA extraction buffer (Appendix A.3.6, table A.21) was added, and the samples were lysed in a tissue lyser (4 mins, 28 s^{-1}) (Qiagen TissueLyser II). The plates were centrifuged (10 mins, 6000 rpm), and 200 μl supernatant was added to 200 μl isopropanol in a fresh 96-well plate, before being left for 10 minutes to allow DNA to precipitate. The plates were centrifuged again (10 mins, 4000 rpm), and the supernatant was discarded. 100 μl ice-cold ethanol was added, and the plates were centrifuged again (10 mins, 4000 rpm), and the supernatant was discarded. After any remaining ethanol had been left to evaporate, 100 μl ddH_2O was added, and the samples were left overnight at 4°C to allow the DNA to dissolve.

3.4.5 Genotyping and sequencing of transgenic plants

T_0 plants were generated as in section 3.4.3, and DNA extracted as in section 3.4.4. The plants were then genotyped by PCR using GoTaq polymerase (Promega) (Appendix A.3.2) with primers designed to amplify Sevir.2G209800 (a positive control gene), the hygromycin resistance gene in the construct, and the Cas9 gene in the construct (for primers see Appendix A.8.1, tables A.77-79). Plants showing bands of the expected height for all three of these PCRs were taken forward to amplification using PCR with primers designed to amplify *HATCH1* (Appendix A.8.2, table A.80) using phusion polymerase (NEB) (Appendix A.3.4) and/or *HATCH3* (Appendix A.8.2, table A.81) using q5 polymerase (NEB) (Appendix A.3.5) as appropriate. PCR products were purified, and Sanger sequenced (Source Bioscience). Specific mutations were identified using visual

inspection of Sanger sequencing traces aligned to the wild-type gene sequence. DNA extraction, PCR genotyping and sequencing were repeated. For several promising plants, seed was collected and taken forward to the next generation.

The same process was carried out in the T_1 generation, except plants were first selected based on absence of the hygromycin resistance gene and the Cas9 gene (but with successful amplification of the positive control gene Sevir.2G209800 to confirm successful DNA extraction). During subsequent PCR amplification and sequencing of *HATCH1* and/or *HATCH3* genes, homozygous mutants were identified. For these construct-negative, homozygous mutant plants, seed was collected and taken forward to phenotypic characterisation in the next generation.

3.4.6 Analysis of isotopic composition of *S. viridis*

Leaf samples were taken in the morning, from four-week-old plants. For each plant line, six equivalent 6 cm samples of the 2nd fully expanded leaf on the main stem (from six separate plants) were taken, and immediately flash frozen in liquid nitrogen. The frozen tissue was freeze-dried before being ground using a tissue lyser (Qiagen TissueLyser II). ~2 mg of each sample was weighed out. Stable isotope analysis was carried out using a service provided by the School of Archaeology, University of Oxford.

3.4.7 *S. viridis* growth measurements

Plant height (from base of plant to base of youngest fully expanded leaf on main culm) and tiller number were taken every week. Panicle emergence date (from date of planting) was taken, and on this date the length and width of the flag leaf were also measured.

3.4.8 *S. viridis* photosynthetic gas-exchange and fluorescence measurements

Photosynthetic gas-exchange and fluorescence measurements were carried out using a LICOR 6800 (LICOR Bioscience). For all measurements the youngest fully expanded leaf on the main culm was used. Leaves were first clamped into the LICOR 6800 and equilibrated for 20 minutes to 400 μmol mol^{-1} CO_2, 500 μmol s^{-1} flow rate, 0.2 kPa over pressure, 50% relative humidity, 32°C leaf temperature, and irradiance of 1800 μmol photons m^{-2} s^{-1}. For A-C_i curves, gas exchange and fluorescence measurements were subsequently taken at 4-minute intervals at the following CO_2 concentrations: 400, 10, 25, 50, 100, 200, 400, 400, 800, 1200, 1600, 400 μmol mol^{-1}, with all other parameters kept constant. For light response curves, gas exchange and fluorescence measurements were subsequently taken at 4-minute intervals at the following light intensities: 1800, 1400, 1200, 1000, 800, 600, 400, 200, 100, 50, 0, with all other parameters kept constant.

3.4.9 Photosynthetic parameter estimation from A-C_i data

Photosynthetic parameters V_{pmax} and V_{cmax} were estimated by fitting the updated steady-state model of C_4 photosynthesis[185,218] to the A-C_i curve data. This was achieved by modifying the spreadsheet for generating modelled A-C_i curves from von Caemmerer (2021), to perform curve-fitting with experimentally determined A-C_i curve data. Input photosynthetic parameter values at 25°C were also taken from von Caemmerer (2021): K_c (Michaelis constant of rubisco for CO_2) = 1210 μbar[236], K_o (Michaelis constant of rubisco for O_2) = 292000 μbar[236], K_p (Michaelis constant of PEPC for CO_2) = 82 μbar[237], R_d (leaf mitochondrial respiration) = 1 μmol m^{-2} s^{-1}, R_m (mesophyll mitochondrial

respiration) = 0.5 μmol m^{-2} s^{-1}, γ* (half reciprocal of rubisco specificity) = 0.000382, a_o (ratio of solubility and diffusivity of O_2 to CO_2) = 0.047[238,239], g_m (mesophyll conductance to CO_2) = 1 μmol m^{-2} s^{-1} bar^{-1} [240]. Temperature was set at 32°C, and values for the aforementioned parameters were predicted at this temperature using Arrhenius functions. Further temperature-independent input photosynthetic parameters were: α (percentage of PSII activity in bundle sheath) = 0, O_m (mesophyll O_2 partial pressure) = 210000 μbar, and g_{bs} (bundle sheath conductance to CO_2) = 0.003 mol m^{-2} s^{-1} bar^{-1} [241]. Irradiance was set at 1800 μmol m^{-2} s^{-1}. Curve fitting was performed by solving using the GRG Nonlinear method to minimise the sum of squared differences between the modelled and measured A-C_i curves, whilst allowing V_{pmax} and V_{cmax} to vary.

In order to calculate a modelled value for carbon compensation point (Γ), equation 4.48 from Caemmerer (2000)[185] was solved for Γ. This equation is:

$$\Gamma^2 + \Gamma\left(\frac{V_{pmax} - R_m}{g_s} + K_p - \Gamma_s\right) - K_p\left(\frac{R_m}{g_s} + \Gamma_s\right) = 0$$

Where the bundle sheath CO_2 partial pressure at compensation point (Γ_s) is:

$$\Gamma s = \frac{\gamma * O_m + K_c(1 + \frac{O_m}{K_o})(\frac{R_d}{V_{cmax}})}{1 + R_d/V_{cmax}}$$

Light compensation point and dark respiration rate R_d were calculated directly from the collected light response curves. Light compensation point was estimated by fitting a linear equation through the A-Q points at 0 μmol m^{-2} s^{-1} and 50 μmol m^{-2} s^{-1} PPFD, and calculating the x-intercept (value for A when Q=0). Rd was calculated as minus the value

for A when Q=0 μmol m^{-2} s^{-1} PPFD (negative net CO_2 assimilation at Q=0 is the dark respiration rate).

3.4.10 Total RNA extraction from *S. viridis* leaves

Leaf samples were taken in the morning, from four-week-old plants. For each plant line, three pooled samples of the middle 2 cm of the 2nd fully expanded leaf on the main stem from three separate plants were taken, and immediately flash frozen in liquid nitrogen. The frozen tissue was ground using a tissue lyser (Qiagen TissueLyser II). The frozen leaf tissue powder was resuspended in 1 ml TRIzol reagent (Thermo Fisher Scientific) using a vortex, and then incubated at room temperature for 5 minutes. The cell debris was collected by centrifugation (15,000 xg, 10 min, 4°C), before 700 μl supernatant was transferred to a fresh 1.5 ml Eppendorf tube. 200 μl chloroform was added, mixed using a vortex, and incubated at room temperature for 3 minutes. The solution was again centrifuged (15,000 xg, 15 min, 4°C), and 350 μl of the upper, aqueous phase was transferred to a fresh 1.5 ml Eppendorf tube. 500 μl isopropanol was added, mixed by inversion, and the solution was incubated at -20°C for 10 minutes. Next, the solution was centrifuged (15,000 xg, 10 min, 4°C) to collect the extracted RNA in a pellet. The supernatant was poured off, and the pellet was washed twice by adding 1 ml 75% ethanol, mixing using a vortex, centrifuging (15,000 xg, 5 min, 4°C), and pouring off the supernatant. After removing any excess supernatant through pipetting and air-drying for 10 minutes, the RNA pellet was resuspended in 50 μl DEPC-treated water by incubation for 10 minutes at 60°C, before storage at -70°C. The RNA was shipped to Novogene for cDNA library construction and paired-end Illumina sequencing.

3.4.11 RNA-seq data analysis

Raw, paired-end sequencing read files were downloaded from Novogene. In order to process the reads, quantify them, and carry out differential expression analysis between lines in a reproducible manner, I developed a computational pipeline using Snakemake[174]. This pipeline can be accessed on GitHub (https://github.com/Olliematttinson/RNA_seq_read_quantification_pipeline). Briefly, read quality was first assessed using FastQC (v 0.11.8) (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Salmon (v 0.14.2)[242] was then used to generate an index for the most recent version of the *S. viridis* transcriptome (Sviridis_500_v2.1_transcript.fa) downloaded from Phytozome V13[243]. Salmon was subsequently used to map and quantify the RNA-seq reads against this index transcriptome. This resulted in a quantification file for each RNA-seq replicate containing transcript abundances in both direct read counts and transcripts per million (tpm). Python scripts were written to merge these separate quantification files into one, and then sum the counts and tpm values for multiple transcript variants of the same gene to produce single transcript abundance estimates for each gene. The gene-level counts data was used as input for differential expression analysis using DESeq2 (v 1.28.0)[244] running in R (v 4.0.0). DESeq2's 'Independent Filtering' option was used to remove genes with a mean counts level below an automatically-determined threshold, to maximise the number of genes with higher expression levels that are differentially expressed between conditions, at a significance level (after Benjamini-Hochberg correction for multiple comparisons) of $p_{adjusted} \leq 0.05$. Boxplots for genes of interest were generated using ggplot2 in R.

In order to obtain predictions of cell-type gene expression specificity, the above pipeline was also run using *S. viridis* mesophyll- and bundle sheath-specific reads[56]. Reads were obtained from European Nucleotide Archive, project PRJEB5074 (mesophyll samples ERR385861, ERR385862, ERR385863, bundle sheath samples: ERR385864, ERR385865, ERR385866). In order to obtain predictions of sub-cellular localisation of gene products, the *S. viridis* proteome containing only the primary transcript (Sviridis_500_2.1_protein_primaryTranscriptOnly.fa) was downloaded from Phytozome V13[243], and input into TargetP-2.0[142]. Hypergeometric tests were performed in R to confirm if mesophyll- or bundle sheath-localised, or proteins with a predicted chloroplast transit peptide, were significantly over- or under-represented in the sets of differentially expressed genes between lines. GO term enrichment analysis was carried out using the online tool at PlantRegMap[245].

Chapter 4: A novel screen for C_4 cycle transporters in *Zea mays*

4.1 Chapter introduction and background

4.1.1 Transporters missing from our understanding of the C_4 cycle in *Z. mays*

As stated in the introduction to this thesis, there are three transport steps currently missing from our understanding of C_4 photosynthesis in maize (Figure 4.1). First, in the mesophyll cell chloroplast, whilst most C_4 plants utilise an Na^+-dependent mechanism of pyruvate import[94] catalysed by BASS2[99], maize is a member of a restricted lineage of grasses that utilises an H^+-dependent pyruvate uptake system[94] that remains unidentified. Second, the transporters responsible for bundle sheath chloroplast malate uptake and pyruvate export also remain unidentified. Whilst malate uptake is proposed to be facilitated by DCT2[52], this transporter is only capable of dicarboxylate exchange, and thus cannot support the net influx of four-carbon dicarboxylates required for the C_4 cycle to function. Therefore, the most parsimonious explanation is the existence of either two separate undiscovered transporters facilitating chloroplast malate import and pyruvate export respectively, or a single transporter that can facilitate both these transport reactions.

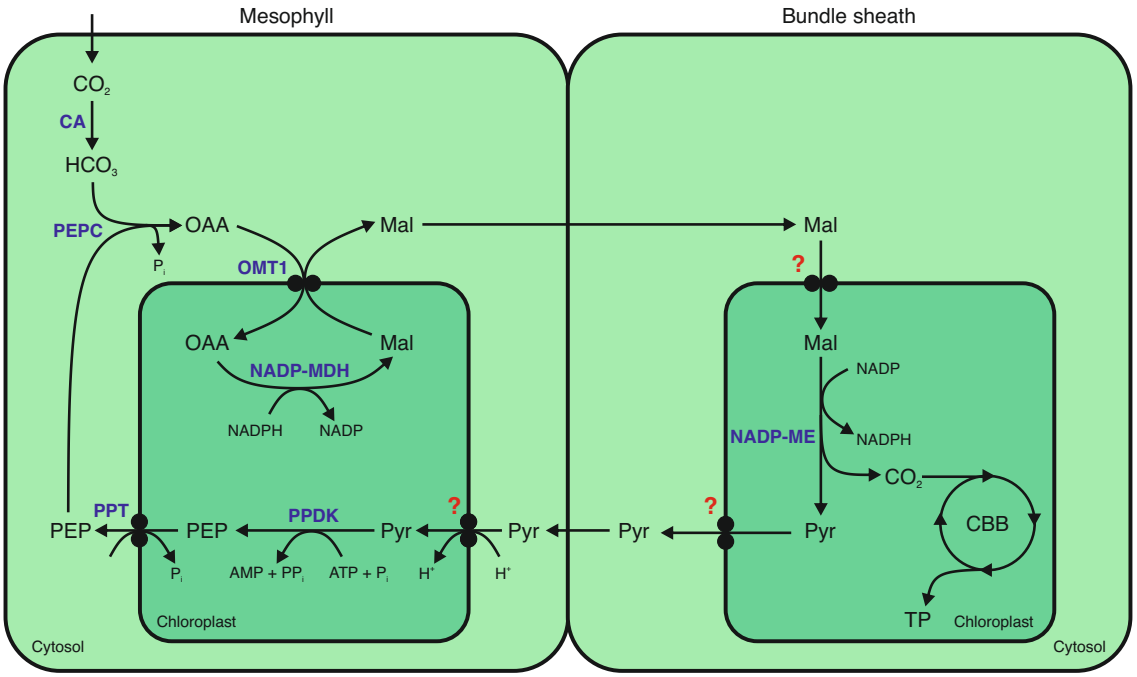


Figure 4.1: Schematic representation of our current understanding of the NADP-ME subtype C_4 cycle in *Z. mays*. Metabolites are shown in black. Known enzymes/transporters are shown in blue. Unknown transporters are represented by red question marks. For abbreviations see appendix A.1, table A.1.

4.1.2 Using bacteria to screen for metabolite transport activity

The use of exogenous expression in bacteria to characterise transporters of key metabolites such as malate and pyruvate has already been discussed in chapter 2, section 2.1.3, and demonstrated in that chapter for the characterisation of *HATCH1* from *S. viridis*. Owing to the comparative ease of using the bacterial system, it was deemed feasible to use the system to screen multiple candidate maize transporters for the missing transport activities from the C_4 cycle (malate importer, pyruvate importer and pyruvate exporter). In brief, exogenously expressing a maize chloroplast malate importer in the *E. coli* malate uptake null mutant *ΔdctA* would be expected to restore

growth on minimal media containing malate as a sole carbon source (Figure 4.2 A). Similarly, expressing a maize chloroplast pyruvate importer in an *E. coli* pyruvate uptake null mutant would be expected to restore growth on pyruvate as a sole carbon source (Figure 4.2 B). Finally, expressing a maize chloroplast pyruvate exporter in wild-type *E. coli* grown on glucose as a sole carbon source would be expected to lead to some reduction in growth (Figure 4.2 C), owing to the export of a key central metabolite from the cell, as was previously observed for *Setaria HATCH1*. Even if pyruvate was not the exported metabolite, mass spectrometry analysis of the supernatant in which the bacteria was grown would identify which metabolites were exported, and thus potentially lead to the identification of a different chloroplast-localised metabolite transporter.

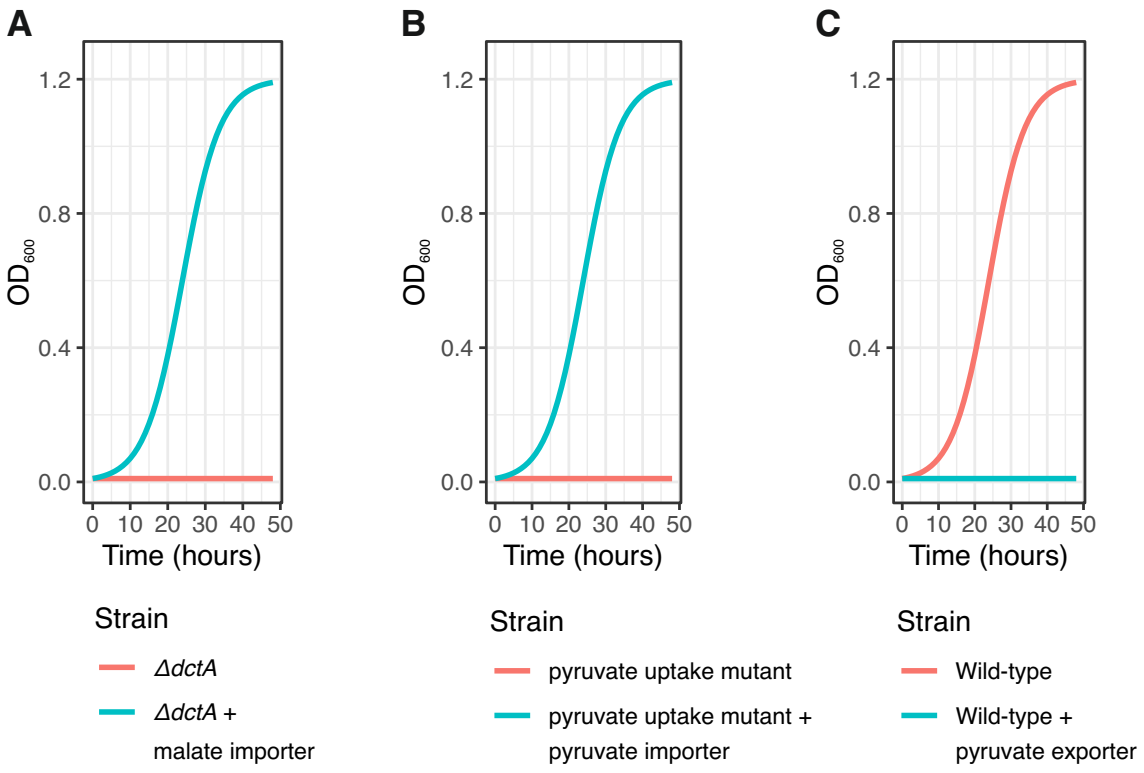


Figure 4.2: Mock data representation of the hypotheses behind the three bacterial screens for metabolite transport activity. **A**) Predicted restoration of growth upon expression of a chloroplast malate importer in *E. coli ΔdctA* on malate as a sole carbon source. **B**) Predicted restoration of growth upon expression of a chloroplast pyruvate importer in an *E. coli* pyruvate uptake null mutant on pyruvate as a sole carbon source. **C**) Predicted inhibition of growth upon expression of a chloroplast pyruvate exporter in wild-type *E. col* (e.g. BW25113) on glucose as a sole carbon source.

4.1.3 Pyruvate uptake systems in *E. coli*

In order to use *E. coli* to screen candidate genes for pyruvate uptake activity, it was necessary to have an *E. coli* strain that is defective in pyruvate uptake. Despite the centrality of pyruvate to bacterial central metabolism, the pyruvate uptake systems in *E. coli* have not yet been fully characterised. At the start of this project, an *E. coli* mutant that is defective in pyruvate uptake and thus unable to grow on pyruvate as a sole

carbon source was not yet available, and therefore I had to make one. The following section explains our understanding of pyruvate uptake systems in *E. coli*, which informed my selection of genes to knock out to attempt to generate a pyruvate uptake null mutant.

A screen for *E. coli* strains capable of growing on the toxic pyruvate analogue 3-fluoropyruvate (3-FP) suggested the existence of at least two pyruvate uptake systems: one constitutive and one inducible[153]. A candidate for the inducible pyruvate uptake system was identified in the gene *YjiY*, which encodes a transmembrane protein and whose expression is controlled by the pyruvate-responsive histidine kinase/response regulator system BtsS/BtsR (Bts stands for Brenztraubensäure, the German word for pyruvate)[246]. This gene was subsequently renamed *BtsT*, and shown to encode a pyruvate/H^+ symporter, based on pyruvate uptake assays on intact *ΔbtsT E. coli*, as well as on proteoliposomes containing reconstituted BtsT[247]. Further screening for 3-FP-resistant *E. coli* mutants led to the identification of the constitutive pyruvate uptake system: CstA[248]. *CstA* (Cst stands for Carbon Starvation) had previously been identified as a putative peptide transporter involved in the carbon starvation response[249]. However, pyruvate uptake assays on *ΔcstA E. coli* demonstrated that CstA is a pyruvate/H^+ symporter[248,250]. *ΔcstA ΔbtsT* double mutants showed lower pyruvate uptake rates than either single mutant, indicating a cumulative effect of knocking out both pyruvate uptake systems[248]. However, the cells were still able to grow on pyruvate and thus alternative uptake mechanisms must also exist.

In addition to the systems above, another candidate pyruvate transporter-encoding gene is *YhjX.* Expression of *YhjX* is induced by another pyruvate-responsive histidine kinase/response regulator system, YpdA/YpdB (recently renamed PyrS/PyrR)[251,252]. YpdA/YpdB also increases activation of BtsS/BtsR (which controls expression of *BtsT*)[251,252]. YhjX belongs to the major facilitator superfamily. One study that initially hypothesised that YhjX might be a pyruvate *exporter*, actually found that *ΔyhjX,* grown in media that induces *YhjX* expression, led to a smaller decrease in extracellular pyruvate concentration compared to wild-type[253]. *YhjX* expression was also increased in an *E. coli* pyruvate kinase mutant, and a *ΔyhjX* exported more pyruvate than wild-type, indicating that YhjX is involved in pyruvate reassimilation in this strain[254]. Furthermore, *ΔyhjX E. coli* showed slightly reduced pyruvate uptake compared to wild-type[248], and *ΔcstA ΔbtsT ΔyhjX* showed only very slow growth on pyruvate, whilst all combinations of single and double mutants did not[250]. Finally, *YhjX* (alongside both *CstA* and *BtsT*) showed significantly increased expression in *E. coli* grown on minimal media containing pyruvate compared to glucose[255]. All this evidence points to YhjX very likely being a third pyruvate uptake system in *E. coli*.

Finally, a fourth candidate pyruvate uptake system from *E. coli* is another major facilitator superfamily member, YnfM. This protein was initially identified as a candidate pentose sugar efflux transporter based on analysis of intracellular arabinose concentrations in *YnfM* deletion and over-expression lines[256]. Subsequently, *YnfM* was also identified as a candidate pyruvate transporter based on a large-scale analysis of microarray data indicating that it was significantly more highly expressed during growth on pyruvate compared to acetate, and likely regulated by the pyruvate-dependent

transcriptional regulator PdhR (pyruvate dehygrogenase regulator protein)[257]. Overall, YnfM is also therefore a candidate pyruvate uptake system in *E. coli*.

It may be possible to identify further pyruvate uptake systems in *E. coli* based on homology to other known pyruvate uptake systems in other species. Several candidate or characterised pyruvate transporters from bacteria do not have *E. coli* orthologs, including the TRAP-T (tripartite ATP-independent periplasmic transporter) family proteins *Rhodobacter sphaeroides* TakP[258] and *Anabaena* sp. Alr3026[259], as well as *Rhizobium leguminosarum* MctP[260]. The genes encoding the eukaryotic mitochondrial pyruvate carrier[107,109] and animal monocarboxylate transporters (MCTs)[261] also do not have *E. coli* orthologs. Other families containing genes that encode for pyruvate transporters do have *E. coli* members. For example, in plants, BASS2 facilitates pyruvate:Na^+ symport into the chloroplast[99]. The single *E. coli* BASS family member, YfeH, is therefore a candidate pyruvate transporter. The lactate/pyruvate:H^+ symporter, Jen1, from yeasts[262,263] has several *E. coli* orthologs. These include the already-characterised transporters of other substrates *ShiA* (shikimate), *KgtP* (2-OG), *NanT* (N-acetyl neuraminic acid), and *ProP* (proline/glycine betaine), as well as one uncharacterised transporter, *YhjE*, which is therefore a candidate pyruvate transporter. The *Corynebacterium glutamicum* pyruvate transporter MctC[264] is closely related to three already characterised *E. coli* transporters of other substrates: PutP (proline), DctA (dicarboxylates), and ActP (acetate/glycolate), and one uncharacterised transporter (YidK), although out of these only *MctC* is orthologous to *ActP*. Given the variety in transport substrates amongst transporters encoded by genes belonging to the same gene family, particularly when separated by large evolutionary distances, YfeH, YhjE, and

YidK can only be tentatively classified as candidate pyruvate transporters. Nevertheless, they are worth taking into consideration.

Overall, there are two confirmed pyruvate uptake systems in *E. coli* (CstA and BtsT), two strong candidates (YhjX and YnfM), and several other weaker candidates based on homology. Thus I selected these genes to knock out to attempt to generate a pyruvate uptake null mutant of *E. coli*.

4.1.4 Chapter aims and outline

The overall aim of this chapter was to attempt to identify the transporters missing from our understanding of the C_4 cycle in maize. Firstly, a bioinformatic screen was carried out to identify candidate C_4 cycle transporters based on cell-type specific transcriptomic and proteomic data, as well as computational predictions of protein structure. Secondly, the top candidate genes were exogenously expressed in *E. coli* and growth-based screens were developed and carried out to attempt to screen the proteins encoded by these genes for relevant transport activities. The development of screens for malate import and metabolite export were fairly straightforward, but the development of a screen for pyruvate import necessitated further investigation into the pyruvate uptake systems of *E. coli*. Finally, one particularly strong candidate gene was taken forward for analysis by gene knock-out in maize, with some preliminary analysis of the resulting lines being discussed here.

4.2 Results and discussion

4.2.1 A bioinformatic screen identifies candidate C_4 cycle transporters in *Z. mays*

A bioinformatic pipeline was developed to screen the maize genome for candidate genes that could encode for C_4 cycle transporters (Figure 4.3). First, as we were searching for chloroplast-localised transporters, candidate genes should encode proteins containing predicted chloroplast transit peptides (cTPs). Genes encoding proteins containing predicted mitochondrial transit peptides (mTPs) were also allowed at this stage due to the high degree of similarity between cTPs and mTPs and the occurrence of many dual-targeted proteins encoded in the genome. In order to avoid false negatives, multiple subcellular localisation prediction programs were used: TargetP-2.0[265,266], WoLF PSORT[267], and iPSORT[268]. Second, as transporters are transmembrane proteins, containing hydrophobic transmembrane domains (TMDs), candidate genes should encode proteins containing at least one predicted TMD. Again, in order to avoid false negatives, multiple TMD prediction programs were used: TMHMM-2.0[269], Phobius[270], Philius[271], PureseqTM[272]. Third, transcripts encoding the candidate transporter proteins should be highly expressed in maize bundle sheath and/or mesophyll cells in cell-type-specific transcriptomic data-sets[48,273,274]. Fourth, the protein itself should also be abundant in quantitative proteomic isolates from these cell types[49]. Finally, the comparative genomics tool OrthoFinder[275] was used to assign *A. thaliana* orthologs to each candidate maize gene, and genes with clearly unrelated, characterised molecular functions (e.g. components of photosystems or known transporters with known different substrates) were manually excluded.

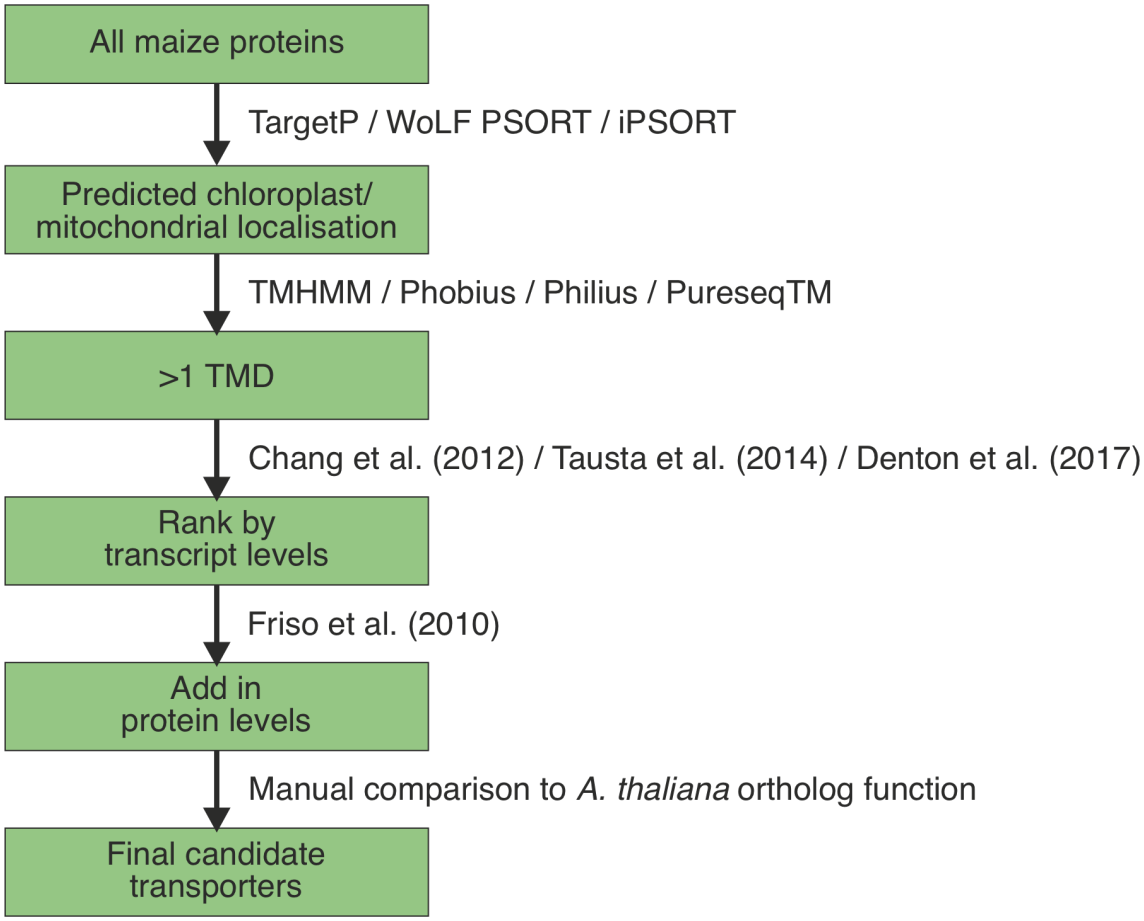


Figure 4.3: Bioinformatic workflow for generating a list of candidate *Z. mays* C_4 cycle transporter-encoding genes

The 11 most promising genes were taken forward as candidate genes for further analysis (Table 4.1).

Candidate #	Accession	Name	Localisation	TMDs	Mesophyll RNA (tpm)	Bundle sheath RNA (tpm)	Mesophyll protein (ppm)	Bundle sheath protein (ppm)
1	GRMZM2G138258	MEP3	C	0-3	348.25	2625.14	3450	5600
2	GRMZM2G099367	APE1	C	2	517.67	22.169	222	241
3	GRMZM2G088138		C/M	2-4	474.37	249.02	825	386
4	GRMZM2G327686	HATCH3.1	C/M	3-4	120.87	67.53	604	0
5	GRMZM2G145500	PIC1/TIC21	C/M	2-4	138.14	110.63	410	45.2
6	GRMZM2G004352		C	3	259.15	102.86	172	192
7	GRMZM2G074053	AVT3	C/S/-	11	104.85	107.47	ND	ND
8	GRMZM2G015033		C	2-3	55.64	38.22	115	989
9	GRMZM2G078118		C/M	2-3	134.59	101.76	407	543
10	GRMZM2G179292	HATCH1	C/M	3-5	119.58	55.21	124	0
11	GRMZM2G133400	HATCH3.2	C/M	3-4	6.17	3.69	604	0

Table 4.1: List of candidate *Z. mays* C_4 cycle transporter-encoding genes. Localisation: All predictions by different prediction programs are shown. C = chloroplast transit peptide, M = mitochondrial transit peptide, S = signal peptide, - = no transit peptide. TMDs: Range of predicted TMD numbers by different prediction programs are shown. Mesophyll and bundle sheath cell-type-specific transcriptomic and proteomic data from Chang et al. (2012)[48] and Friso et al. (2010)[49] respectively. Names come from the *A. thaliana* ortholog.

Further information on each of these candidate genes is given below:

Candidate 1 (GRMZM2G138258) is of particular interest. It was originally named *MEP3* (*mesophyll envelope protein 3* and it is the ortholog of *A. thaliana RER4* (*Reticulata-related 4*). Mutation in *RER4*, as well as other members of the gene family *RE* and *RER3*,

leads to leaf lamina reticulation and reduced plant growth[276]. This mutant phenotype, and the further metabolic alterations observed in the leaves of *re* and *rer3* mutant plants (metabolomics was not performed on *rer4* plants), were similar to those found in *cue1* mutant plants defective in the chloroplast PEP transporter PPT[80,81]. It was initially suggested that RER proteins could modulate the activity of PPT, but attempts to co-immunoprecipitate PPT and RER3 were unsuccessful[276]. Therefore, it is plausible that RER proteins are transporters of similar metabolites to PPT, such as pyruvate. In maize, the protein encoded by GRMZM2G138258 has also been named MEP3 (mesophyll envelope protein 3) due to its enrichment in the maize chloroplast proteome compared to that of pea[51]. Candidate 1 shows significant transcript and protein levels in both mesophyll and bundle sheath, although it is preferentially expressed in the bundle sheath. It is therefore a candidate for all three missing transporters, although most likely the bundle sheath chloroplast malate importer or pyruvate exporter.

Candidate 2 (GRMZM2G099367) is orthologous to *A. thaliana APE1* (Acclimation of the Photosynthesis to the Environment 1) (AT5G38660). *AtAPE1* was initially identified in a mutant screen for altered chlorophyll fluorescence characteristics in plants grown in low light before transfer to high light[277]. Interestingly, another mutant line identified in this screen, *ape2*, was found to carry a mutation in *TPT* (encoding for the chloroplast triose phosphate translocator). Whilst *APE1* remained uncharacterised, it was plausible that it also encoded for a chloroplast metabolite transporter. As candidate 2 is highly expressed in maize mesophyll cells (according to transcriptomics) and in both mesophyll and bundle sheath cells (according to proteomics), it could encode any of the three missing maize transporters. Since I carried out this work however, the ortholog of this protein in

the green alga *Chlamydomonas reinhardtii* has been found to encode a thylakoid-localised protein that interacts with PSII, and functions in PSII supercomplex remodelling to modulate photosynthetic electron transport and prevent photoinhibition during acclimation to high light[278]. These findings would appear to eliminate candidate 2 from contention as a C_4 cycle transporter. This was not known when the work described below was completed.

Candidate 3 (GRMZM2G088138) has no *A. thaliana* ortholog. As such, the gene family to which this candidate belongs is entirely uncharacterised. Candidate 3 has high expression in both maize mesophyll and bundle sheath cells, although with preferential expression in the mesophyll (in both transcriptomics and proteomics). It is therefore a candidate for all three missing transporters, although most likely the mesophyll chloroplast pyruvate importer.

Candidates 4 (GRMZM2G327686), 10 (GRMZM2G179292), and 11 (GRMZM2G133400) are of particular interest as they belong to the *HATCH* gene family. Candidate 10 is the ortholog of *Setaria HATCH1*, whilst Candidates 4 and 11 are both orthologous to *Setaria HATCH3*. Whilst characterisation of *Setaria HATCH1* and *HATCH3* was the focus of chapters 2 and 3 of this thesis, the maize genes have not yet been characterised. Maize *HATCH3.2* shows low transcript levels, and identical protein levels to *HATCH3.1*, which may therefore merely result from mis-matching of peptides from *HATCH3.1* to *HATCH3.2*. Maize *HATCH3.1* and *HATCH1* however show mesophyll-preferential transcript and protein accumulation, in contrast to the bundle sheath-preferential and non-preferential expression of *Setaria HATCH1* and *HATCH3* respectively. Thus, these

proteins likely carry out different roles in maize to *Setaria,* perhaps facilitating pyruvate import into the mesophyll chloroplast.

Candidate 5 (GRMZM2G145500) is orthologous to an *A. thaliana* gene AT2G15290, which has been described to encode either a putative iron permease (PIC1 – Permease In Chloroplasts 1)[279] or a component of the TIC (Translocon on the Inner Chloroplast membrane) complex (TIC21)[280]. As the function of this protein is unknown, candidate 5 was taken forward with the idea that high, mesophyll-specific expression, particularly proteomically, might indicate a changed role in maize as the mesophyll chloroplast pyruvate importer.

Candidate 6 (GRMZM2G004352) is orthologous to an uncharacterised *A. thaliana* gene (AT2G21960). No information could be found about this gene or its orthologs in any species. Thus, as candidate 6 is expressed in both maize mesophyll and bundle sheath cells, it is a candidate for all three missing C_4 cycle transporters.

Candidate 7 (GRMZM2G074053) is potentially interesting, as its *A. thaliana* orthologs (*AVT3A-C*: AT5G65990, AT4G38250, and AT2G42005 respectively) are members of the Amino acid Vacuolar Transporter (AVT) family, with one (AT5G65990) being shown to transport neutral amino acids such as alanine[281]. In *A. thaliana* N-terminal GFP fusions of AVT1-3 proteins showed vacuolar membrane localisation[281]. The use of an N-terminal protein fusion in that study was a mistake, as this would disrupt the N-terminal-located chloroplast transit peptide, and thus result in an erroneous localisation of the protein. For maize candidate 7, the targeting peptide prediction software used in our pipeline (TargetP, WoLF PSORT, and iPSORT) give differing predictions (no target peptide, signal

peptide, cTP respectively). However, the online resource Aramemnon[282] (which utilises 18 separate targeting peptide prediction algorithms) gives a strong consensus prediction of chloroplast localisation. Candidate 7 shows significant transcript levels in both mesophyll and bundle sheath cells. The protein was not detected in the proteomic study, which may be related to the high number of TMDs. Therefore, based on the transcriptomic data, this gene could encode for any of the three missing transporters. It may alternatively encode a transporter of C_4 cycle-related amino acids, such as aspartate and alanine.

Candidate 8 (GRMZM2G015033) is orthologous to an uncharacterised *A. thaliana* gene (AT3G15110). Whilst it shows only moderate transcript abundance in maize mesophyll and bundle sheath, the measured protein abundance is higher, particularly in the bundle sheath. Thus, candidate 8 could encode for the maize bundle sheath chloroplast malate importer or pyruvate exporter.

Candidate 9 (GRMZM2G078118) is orthologous to an *A. thaliana* gene (AT5G37360), which is annotated as an 'ammonium transporter 1-like protein', but is uncharacterised. Substantial levels of transcript and protein are found in both maize mesophyll and bundle sheath cells, and thus candidate 8 could encode any of the three missing transporters.

After bioinformatically selecting a shortlist of candidate transporter genes, the next step in the high-throughput screen for transport activity was to express these genes in *E. coli*. To this end, the gene sequences were modified through the removal of the predicted cTP (which would not be cleaved off when expressed in *E. coli* and may interfere with

protein folding) and codon optimisation for expression in *E. coli*. The sequences were then cloned into an inducible expression vector and transformed into the appropriate *E. coli* cell lines for the experiments below.

4.2.2 A bacterial screen for malate import reveals no candidates

In order to determine if expression of any of the candidate genes restored growth of the *ΔdctA* cell line on minimal media containing malate as a sole carbon source, *ΔdctA* transformed with candidate gene expression plasmids were pre-grown, and either supplied with the inducer or not. These cultures were then inoculated to OD 0.01 in M9 minimal media containing malate (33.3 mM, pH 7) as a sole carbon source, and OD_{600} was measured continuously in a plate reader. BW25113, the corresponding wild-type strain to *ΔdctA* (as a positive control), and *ΔdctA* (as a negative control) were also included. As a further positive control for the functionality of the inducible expression system, *ΔdctA* containing the inducible plasmid containing *DctA* itself was also included. Unfortunately, there are no known plant chloroplast transporters capable of malate uptake, and thus a positive control using a chloroplast malate uptake system could not be included. Finally, for this screen for maize chloroplast malate importers, given that DCT2 is the only transporter that has to date been hypothesised to fulfil this role[52], it was also decided to include *ΔdctA* containing the expression plasmid containing *DCT2*. If expression of *DCT2* were to restore *ΔdctA* growth on malate as a sole carbon source, this would indicate that, contrary to previous reports[30], DCT2 is capable of net malate

uptake, and thus could fulfil the role of maize bundle sheath chloroplast malate importer. The results for the screen are shown in (Figure 4.4).

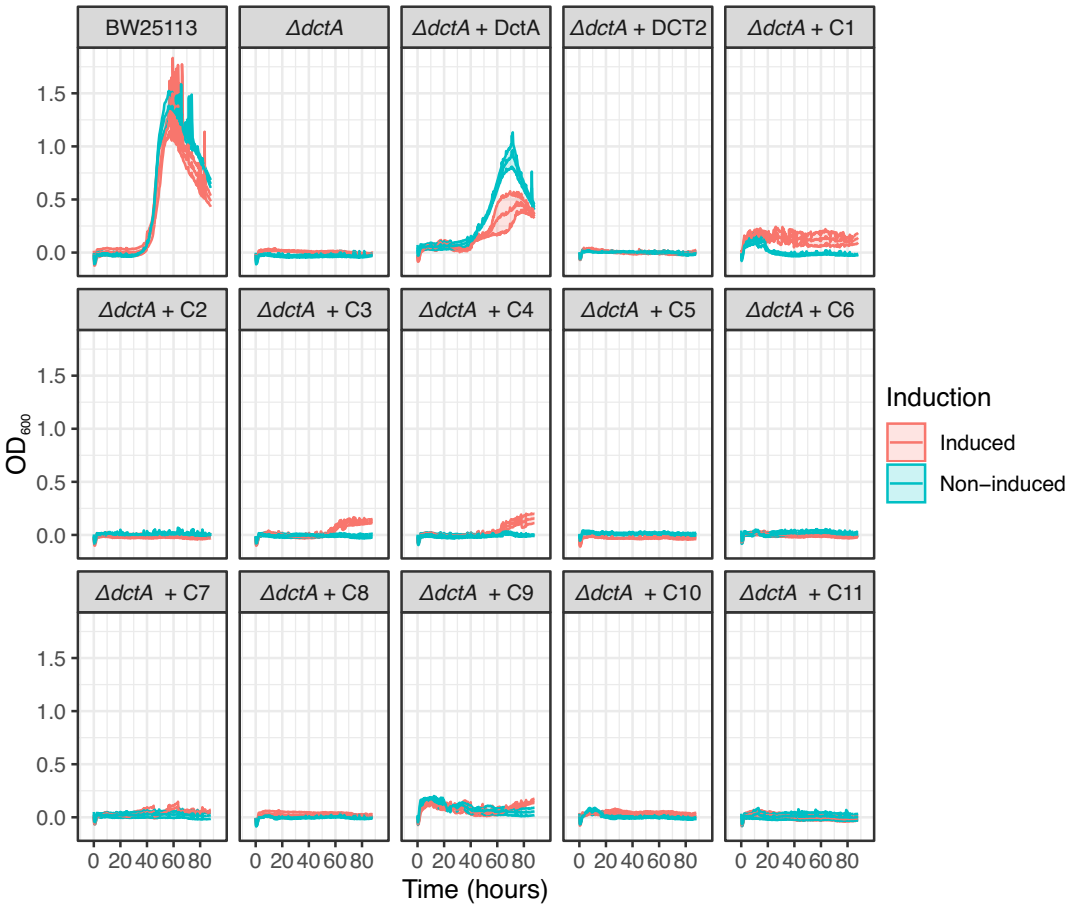


Figure 4.4: Expression of candidate genes in *E. coli ΔdctA* on minimal media containing malate as a sole carbon source. Each panel shows the change in OD_{600} of one strain over time. Error bars show standard error of the mean (N = 3). C1-C11 = candidate genes 1-11.

As expected, BW25113, which contains a functional *DctA* gene, was able to grow on minimal media containing malate as a sole carbon source, whilst *ΔdctA* could not.

Somewhat surprisingly, *DctA* was able to restore *ΔdctA* growth on malate in both the presence and the absence of the inducer. In fact, growth rate was faster in the non-induced cultures. This indicates that the inducible expression system is somewhat leaky, and that the level of *DctA* leaky expression in the absence of inducer is still sufficient to complement the *ΔdctA* mutant. Thus, for the remainder of the screen, a positive result for a candidate gene would mean that gene restored *ΔdctA* growth on malate as a sole carbon source, at *either* level of expression provided by presence *or absence* of the inducer.

In line with previous characterisation of DCT2 as an obligate dicarboxylate exchanger[30], *DCT2* was not able to restore *ΔdctA* growth on malate as a sole carbon source. This validates the search for a further transporter capable of malate uptake. Unfortunately, none of the candidate genes tested in this screen were able to restore growth of *ΔdctA* on malate as a sole carbon source.

4.2.3 Generation of an *E. coli* pyruvate uptake null mutant

In order to carry out an analogous screen to the malate importer screen to look for pyruvate importers, it was necessary to have an *E. coli* strain that is defective in pyruvate uptake, and thus cannot grow on minimal media containing pyruvate as a sole carbon source. Exogenous expression of a chloroplast pyruvate transporter would be expected to restore the ability of this strain to grow on pyruvate. To this end, I successively knocked out candidate pyruvate transporters in *E. coli*, testing the growth of each successive strain on minimal media containing pyruvate as a sole carbon source.

To generate the multiple candidate pyruvate uptake null mutant strains, I started with the *ΔyfeH* knock-out line from the Keio collection[283] (in which *YfeH* is replaced with a kanamycin resistance cassette). This strain was named 1m, for single mutant. Subsequently I used recombineering to knock out the four other candidate pyruvate uptake systems, to generate the strains 2m (*ΔyfeH ΔcstA*), 3m (*ΔyfeH ΔcstA ΔbtsT*), 4m (*ΔyfeH ΔcstA ΔbtsT ΔynfM*), and 5m (*ΔyfeH ΔcstA ΔbtsT ΔynfM ΔyhjX*). The process of recombineering involves replacing a target gene with an antibiotic resistance gene expression cassette using homologous recombination. To generate 2m, *CstA* was replaced with an ampicillin resistance cassette. To generate 3m, *BtsT* was replaced with a spectinomycin resistance cassette. Subsequently, due to running out of antibiotic resistance genes to use, the kanamycin resistance gene was removed from 3m using recombination, thus freeing up kanamycin for use in further recombineering. To generate 4m, *YnfM* was replaced with a kanamycin resistance cassette. Once again, the kanamycin resistance cassette was removed. Finally, to generate 5m, *YhjX* was replaced with a kanamycin resistance cassette. At each stage, successful gene knock-out and replacement with antibiotic resistance cassette, or antibiotic resistance cassette removal, were confirmed using PCR. To further verify quintuple gene knock-out in the 5m line, the whole genome was sequenced.

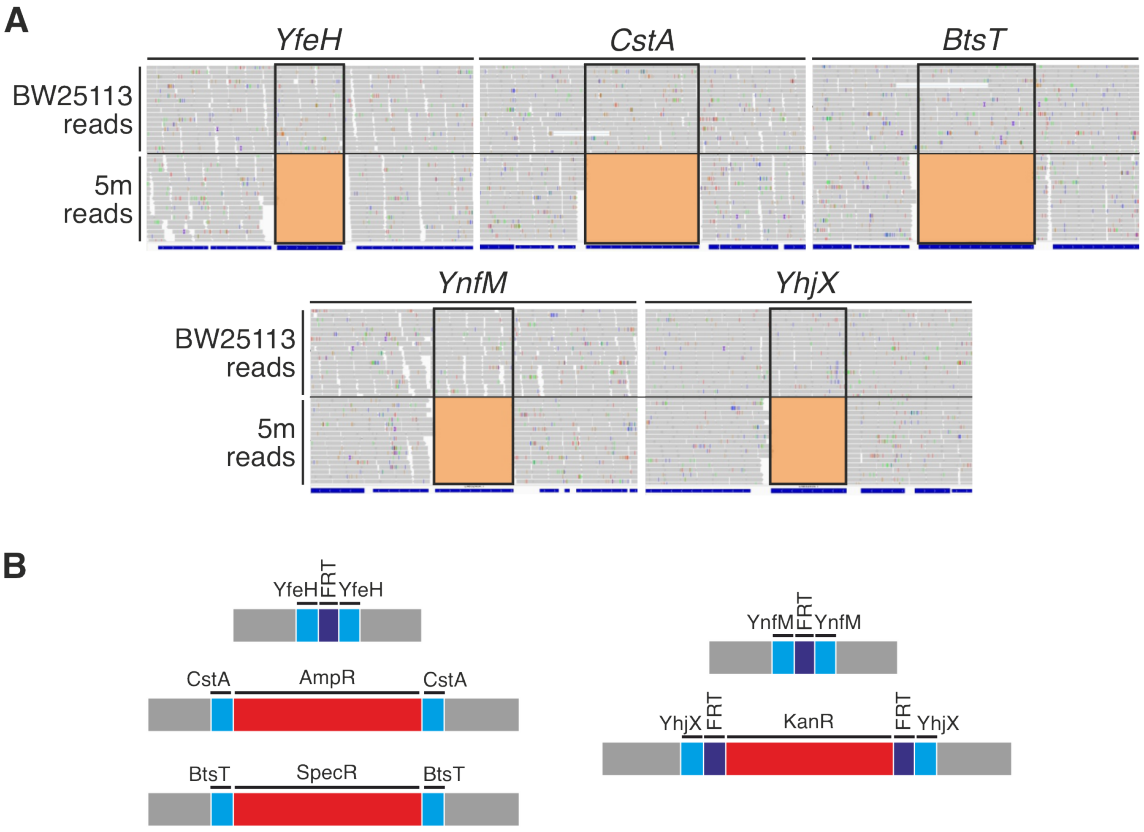


Figure 4.5: Whole genome sequencing of BW25113 (wild-type) and quintuple mutant (*ΔyfeH ΔcstA ΔbtsTΔynfM ΔyhjX*) *E. coli* lines. **A**) The top and bottom panels show reads from BW25113 and 5m respectively mapped against the genome of another *E. coli* wild-type strain, K12. Orange boxes show locations where no reads map to the K12 genome, indicating replacement of the gene of interest. **B**) Cartoon representations of the five genes in question showing inserted sequences. Blue = gene of interest. Red = antibiotic resistance gene. Purple = FRT (FLP recognition target) sites used to remove Kanamycin resistance cassettes from *YfeH* and *YnfM* sites.

The five mutant lines, as well as the corresponding wild-type BW25113, were inoculated in M9 minimal media containing pyruvate (44.4 mM, pH 7) as a sole carbon source (Figure 4.6 A). Analysis of the growth curves by fitting a logistic equation allowed for the estimation of intrinsic growth rate for each line (Figure 4.6 B). BW25113 had a mean

growth rate of 0.248 h^{-1} ± 0.037 (SD, N=3). 1m had a mean growth rate of 0.236 h^{-1} ± 0.062 (SD, N=3). 2m had a mean growth rate of 0.232 h^{-1} ± 0.022 (SD, N=3). 3m had a mean growth rate of 0.208 h^{-1} ± 0.024 (SD, N=3). 4m had a mean growth rate of 0.122 h^{-1} ± 0.028 (SD, N=3). 5m had a mean growth rate of 0.031 h^{-1} ± 0.018 (SD, N=3). Thus, each successive knock-out had detrimental effect on mean growth rate for each strain, but this effect was only marginal for 1m, 2m, and 3m. In contrast, 4m (additional knock-out of *YnfM*) had significantly decreased growth rate compared to wild-type, whilst 5m (additional knock-out of *YhjX*) had almost complete growth arrest. The differences in growth rate between wild-type and each mutant line were analysed using the post-hoc Dunnett's test (which is used for multiple comparisons with a single baseline). 1m, 2m, and 3m did not show significantly decreased growth rates compared to wild-type ($p = 0.99234$, $p = 0.96753$, $p = 0.53091$ respectively). 4m did show significantly decreased growth rate compared to wild-type ($p = 0.00383$), and 5m showed a more significant decrease in growth rate still ($p = < 0.001$).

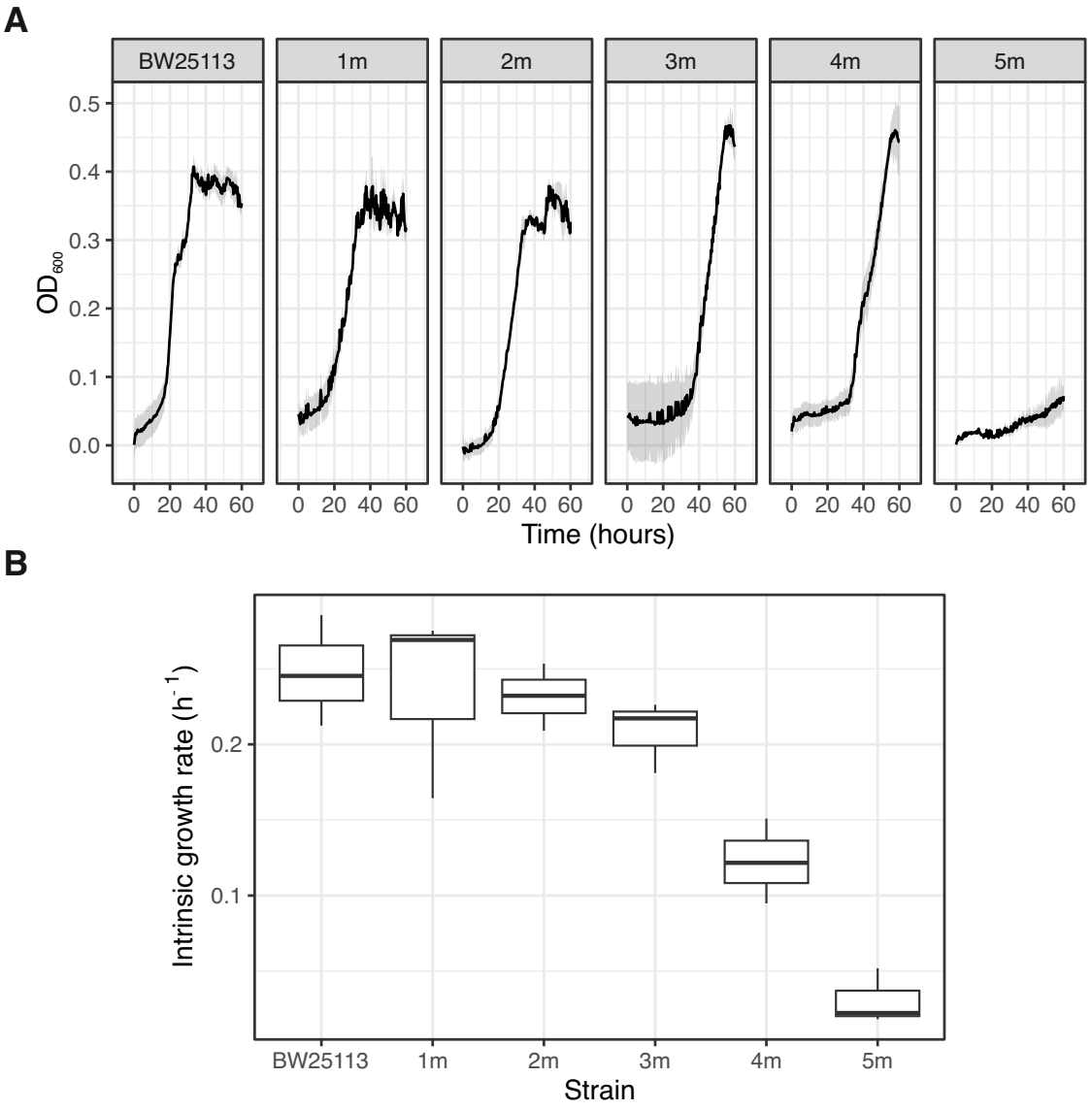


Figure 4.6: Growth of pyruvate transporter knock-out *E. coli* lines on minimal media containing pyruvate as a sole carbon source. **A**) Growth curves. Each panel shows the change in OD_{600} of one strain over time. Error bars show standard error of the mean (N = 3). **B**) Estimated intrinsic growth rate. (N=3).

To attempt to confirm the roles of the candidate pyruvate transporters in facilitating pyruvate uptake into *E. coli*, an inducible expression plasmid was constructed for each candidate transporter, and the candidate transporters were individually expressed in

the 5m line, grown on pyruvate as a sole carbon source (Figure 4.7). No significant growth restoration was observed for any of the transporters, except for CstA and BtsT. Interestingly, growth restoration was only observed with the *CstA* and *BtsT* expression plasmids when no inducer was added, but not when 50 μM inducer was added. This is consistent with the previous observation that growth restoration of *ΔdctA* containing the *DctA* expression plasmid on malate as a sole carbon source was stronger in the absence of inducer. Thus, leaky expression of these transporters from the inducible expression plasmid is sufficient to restore 5m growth on pyruvate as a sole carbon source.

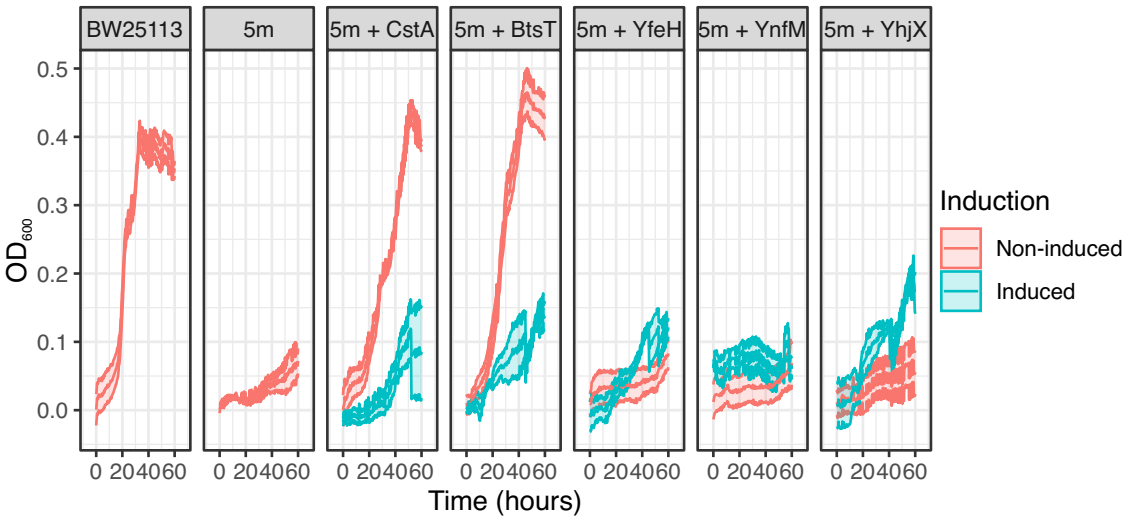


Figure 4.7: Growth of quintuple knock-out line re-expressing individual candidate pyruvate transporters on minimal media containing pyruvate as a sole carbon source. Each panel shows the change in OD_{600} of one strain over time. Error bars show standard error of the mean (N = 3).

Overall, these data provide no evidence that YfeH facilitates pyruvate uptake in *E. coli*, but does support roles for CstA, BtsT, YnfM, and YhjX, as pyruvate uptake systems in *E. coli*. CstA and BtsT do not cause a decrease in growth rate on pyruvate as a sole carbon source when sequentially knocked out, but in the quintuple mutant their re-expression leads to a restoration in growth on pyruvate as a sole carbon source. In contrast, YnfM and YhjX do not cause a restoration in growth upon their re-expression in the quintuple mutant on pyruvate as a sole carbon source, but do cause a decrease in growth rate on pyruvate as a sole carbon source when sequentially knocked out in the triple mutant background. It is likely that CstA and BtsT are the primary pyruvate uptake systems in *E. coli*, but that knocking them out does not lead to a decrease in growth rate on pyruvate as a sole carbon source due to the remaining YnfM and YhjX transport systems. YnfM and YhjX are secondary uptake systems that are not sufficient to alone support *E. coli* growth on pyruvate as a sole carbon source when re-expressed in the quintuple mutant, at least utilising the expression system used here. Knocking out *YnfM* and *YhjX* does lead to a decrease in growth rate on pyruvate as a sole carbon source when the primary pyruvate uptake systems (CstA and BtsT) are not present.

After performing this work, another study found that triple mutant *ΔcstA ΔbtsT ΔyhjX E. coli* showed only very slow growth on minimal media containing pyruvate as a sole carbon source, but that all combinations of single and double mutants showed no significant difference compared to wild-type[250]. Further, growth of this triple mutant on pyruvate as a sole carbon source could be restored through re-expression of *CstA*[250]. This is in agreement with the work presented here. As that study did not also knock out

YnfM, we do not know if a further growth defect would have been observed, but the results presented here suggest that this would have been the case.

The quintuple mutant showed extremely reduced growth on pyruvate as a sole carbon source, but growth was not completely eliminated. This suggests either that other, low efficiency pyruvate uptake systems exist in *E. coli*, or simply that a small amount of pyruvate can enter the cell via passive diffusion (likely due to a small amount of pyruvate being present as pyruvic acid at pH 7) even in the absence of specific transport systems. Overall, aside from interesting investigation into the pyruvate uptake systems of *E. coli*, the main purpose of knocking out these systems was to generate a mutant that carried out no or very slow pyruvate uptake, and therefore showed no or very slow growth on pyruvate as a sole carbon source. This was successfully achieved, and moreover growth of this strain can be restored by expression of known pyruvate uptake systems. Thus, the quintuple mutant could be taken forward to use as a chassis for testing maize proteins for pyruvate uptake activity.

4.2.4 A bacterial screen for pyruvate import reveals no candidates

In order to determine if expression of any of the candidate genes restored growth of the 5m cell line on minimal media containing pyruvate as a sole carbon source, 5m transformed with candidate gene expression plasmids were pre-grown, and either supplied with the inducer or not. These cultures were then inoculated to OD 0.01 in M9 minimal media containing pyruvate as a sole carbon source, and OD_{600} was measured continuously in a plate reader. BW25113, the corresponding wild-type strain to 5m (as

a positive control), and 5m (as a negative control) were also included. Restoration of 5m growth on pyruvate as a sole carbon source through inducible expression of bacterial pyruvate transporters (*CstA* and *BtsT*) had already been demonstrated (Figure 4.7). Therefore, as a further positive control to test the ability of a known plant chloroplast pyruvate importer to restore 5m growth on pyruvate, 5m containing the inducible plasmid containing *S. viridis BASS2* was also included. The results for the screen are shown in Figure 4.8.

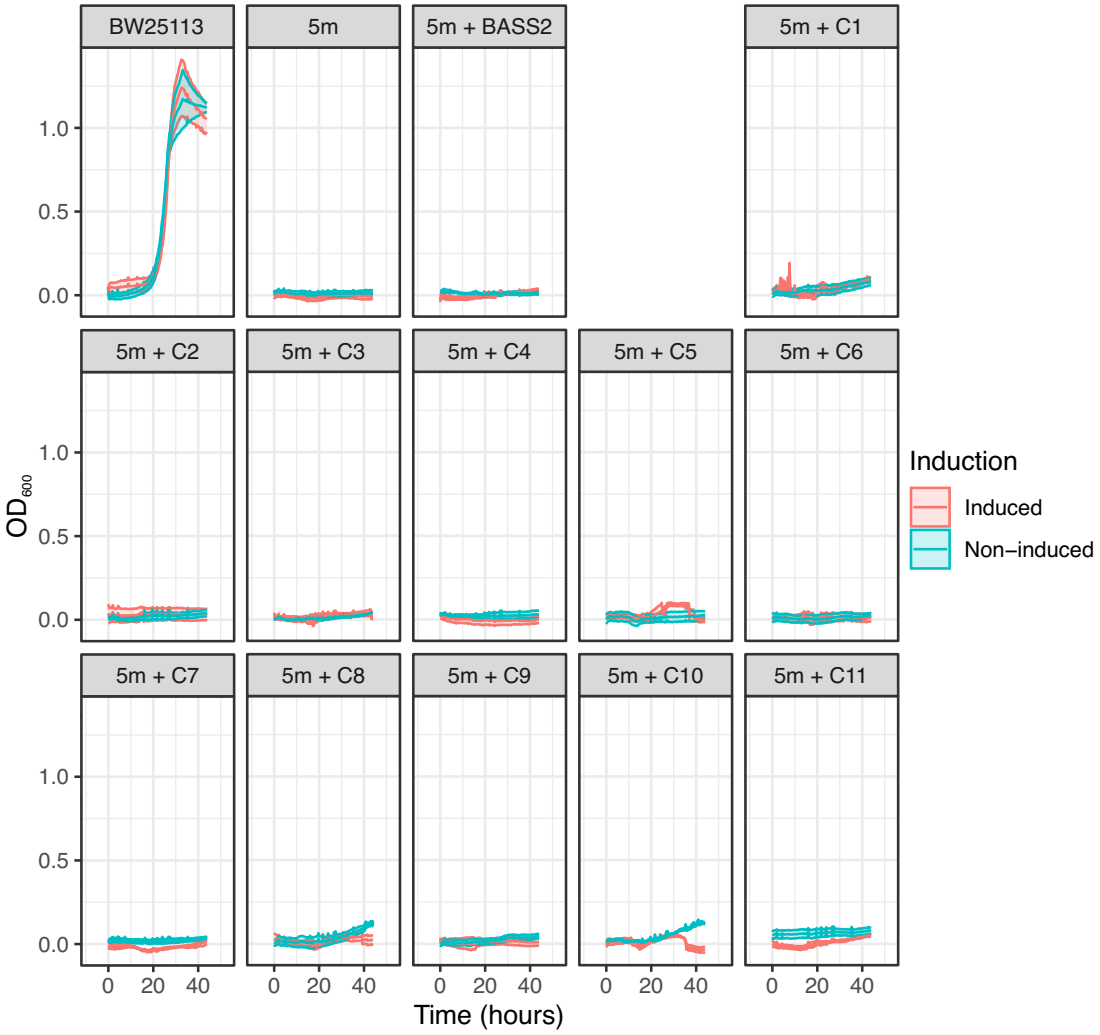


Figure 4.8: Expression of candidate genes in *E. coli* 5m on minimal media containing pyruvate as a sole carbon source. Each panel shows the change in OD_{600} of one strain over time. Error bars show standard error of the mean (N = 3). C1-C11 = candidate genes 1-11.

As expected, BW25113, which contains functional pyruvate uptake systems, was able to grow on minimal media containing pyruvate as a sole carbon source, whilst 5m could not. Unfortunately, none of the candidate genes tested in this screen were able to restore significant growth of 5m on pyruvate as a sole carbon source. However, the

positive control, *SvBASS2*, also did not restore growth on pyruvate. The implication of this failed positive control is that some plant proteins, such as *BASS2*, do not function properly in this exogenous expression system. This may be due to several reasons, including a lack of post-translational modification in the bacterial system, and potential protein mis-folding in a foreign cellular environment. Even if the protein is expressed, it may localise to internal liposomes, rather than as desired to the plasma membrane. The bacterial expression system used here was chosen with this limitation of potential false negatives known, as a method of screening for specific transport activity in a higher throughput manner than would otherwise be possible. Thus, in the absence of any restoration of 5m growth on pyruvate, we cannot make any strong conclusions about the possible pyruvate transport activities of any of the candidate genes.

4.2.5 A bacterial screen for metabolite export reveals some candidates

In order to determine if the expression of any candidate genes led to a decrease in growth rate of the BW25113 cell line on minimal media containing glucose as a sole carbon source, possibly due to export of a central metabolite such as pyruvate, BW25113 transformed with candidate gene expression plasmids were pre-grown, and either supplied with the inducer or not. These cultures were then inoculated to OD 0.01 in M9 minimal media containing glucose (22.2 mM, pH 7) as a sole carbon source, and OD_{600} was measured continuously in a plate reader. The bacterial growth curves are shown in Figure 4.9 A. Analysis of the growth curves by fitting a logistic equation allowed for the estimation of intrinsic growth rate for each line, which was normalised to show

the proportional growth rate of induced compared to non-induced cultures for each line (Figure 4.9 B). Reductions in growth rate in induced compared to non-induced cultures was tested for statistical significance using t-tests with Benjamini-Hochberg correction for multiple comparisons. A significant reduction in growth rate in induced compared to non-induced cultures was observed for lines containing candidate 1 (BH-adjusted p = 0.0228), candidate 2 (BH-adjusted p = 0.0193), candidate 3 (BH-adjusted p = 0.0195), candidate 4 (BH-adjusted p = 0.0221), candidate 7 (BH-adjusted p = 0.0100), candidate 10 (BH-adjusted p = 0.00515), and candidate 11 (BH-adjusted p = 0.00306), but not lines containing candidate 5 (BH-adjusted p = 0.189), candidate 6 (BH-adjusted p = 0.511), candidate 8 (BH-adjusted p = 0.455), or candidate 9 (BH-adjusted p = 0.173).

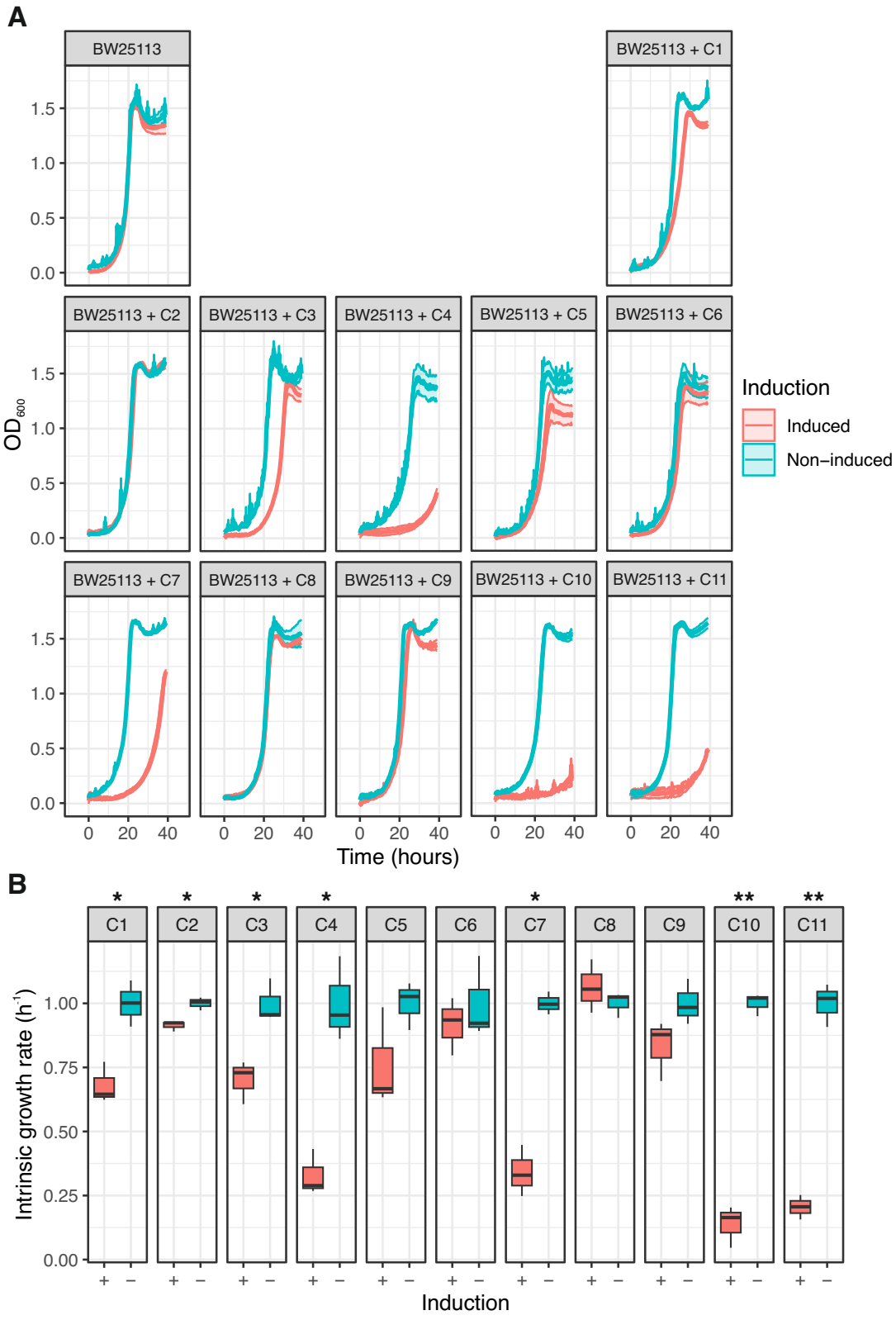

A
BW25113
BW25113 + C1
BW25113 + C2
BW25113 + C3
BW25113 + C4
BW25113 + C5
BW25113 + C6
BW25113 + C7
BW25113 + C8
BW25113 + C9
BW25113 + C10
BW25113 + C11
OD600
Time (hours)
Induction
Induced
Non−induced
B
C1
C2
C3
C4
C5
C6
C7
C8
C9
C10
C11
Intrinsic growth rate (h-1)
Induction

Figure 4.9: Expression of candidate genes in BW25113 *E. coli* on minimal media containing glucose as a sole carbon source. **A**) Growth curves. Each panel shows the change in OD_{600} of one strain over time. Error bars show standard error of the mean (N = 3). **B**) Estimated intrinsic growth rate, standardised to mean of non-induced intrinsic growth rate for each line, for each plasmid-containing line (N=3). Stars indicate significance between induced and non-induced for a strain (t-test with Benjamini-Hochberg correction for multiple (11) comparisons, * for adjusted p-value ≤ 0.05, ** for adjusted p-value ≤ 0.01). C1-C11 = candidate genes 1-11.

A significant decrease in growth rate for *E. coli* lines expressing candidate 4 (*HATCH3.1*), candidate 10 (*HATCH1*), and candidate 11 (*HATCH3.2*) is consistent with the similar decrease in growth rate for *E. coli* expressing *S. viridis HATCH1* (chapter 2). This suggests that molecular function may be conserved both between species and across members of the *HATCH* gene family. Significant decreases in *E. coli* growth rate upon induction of candidates 1, 2, 3, and 7 provide the first experimental evidence that these proteins may act as metabolite transporters.

4.2.6 Metabolite export profiles of promising candidates

In order to identify any metabolites being exported by *E. coli* expressing the most promising candidate maize transporters, the genes encoding some of these transporters were again expressed in *E. coli*, and the supernatant was collected and analysed by mass spectrometry. The three *HATCH* genes (candidates 4, 10, and 11) will be discussed more in the next chapter. Candidates 1, 3, and 7 are discussed below.

Induction of candidate 1 (*MEP3*) expression appeared to cause a minor increase in export rate of pyruvate and succinate compared to non-induced cultures (Figure 4.10). This is consistent with the subtle but significant decrease in growth rate of induced compared to non-induced cultures (Figure 4.9). Pyruvate export activity is consistent with the bundle sheath-preferential expression of this gene. However, these differences in metabolite export rate between induced and non-induced cultures were not statistically significant either after or before correcting for multiple comparisons, and so no hard conclusions can be drawn here. It is possible that MEP3 is a pyruvate exporter, but that it functions only ineffectively when exogenously expressed in *E. coli*.

In contrast, induction of candidate 3 expression led to a small but significant increase in export of succinate and fumarate (Figure 4.10). This could suggest a C_4 role in dicarboxylate transport, although preferential expression of this gene in the mesophyll, where dicarboxylate transport is already accounted for by OMT1/DCT1, argues against this idea. It is possible that the low levels of dicarboxylate transport here are an experimental artifact. The highest level of metabolite transport caused by expression of candidate 3 was actually pyruvate, and there was significantly higher pyruvate export rate from induced compared to non-induced cells before correcting for multiple comparisons ($p = 0.021$). However, comparatively large variation in pyruvate export rate between induced samples meant that the increase in pyruvate export rate upon induction was no longer significant after correction for multiple comparisons. As with candidate 1, this is consistent with the comparatively small but significant decrease in growth rate of induced compared to non-induced cultures for this candidate (Figure 4.9). As a pyruvate transport role (import into the chloroplast) would be consistent with

the mesophyll-preferential expression of this gene, I believe that candidate 3 warrants further investigation.

Induction of candidate 7 (*AVT3*) appeared to cause a large increase in pyruvate export rate compared to the non-induced control (Ø: 27,700 molecules pyruvate $cell^{-1}$ s^{-1} ± 10,100 (SD, N = 3) induced compared to 340 molecules pyruvate $cell^{-1}$ s^{-1} ± 340 (SD, N = 3) non-induced). This is consistent with the larger disruption in growth rate of induced compared to non-induced cultures for this candidate in comparison to candidates 1 and 3 (Figure 4.9). However, again owing to large experimental variation, this difference in export rate of pyruvate was not significant after or before correction for multiple comparisons (Figure 4.10). It is worth noting that this experiment, which was carried out as part of a larger screen, did not test for the export of alanine. This is the most likely transport substrate for candidate 7, given previous investigation into its *A. thaliana* ortholog *AVT3*[281]. However, possible export of pyruvate, a similar 3-carbon molecule, would be consistent with the hypothesis for a primary role in alanine transport across bundle sheath and mesophyll chloroplast envelopes as part of auxiliary C_4 cycles in maize.

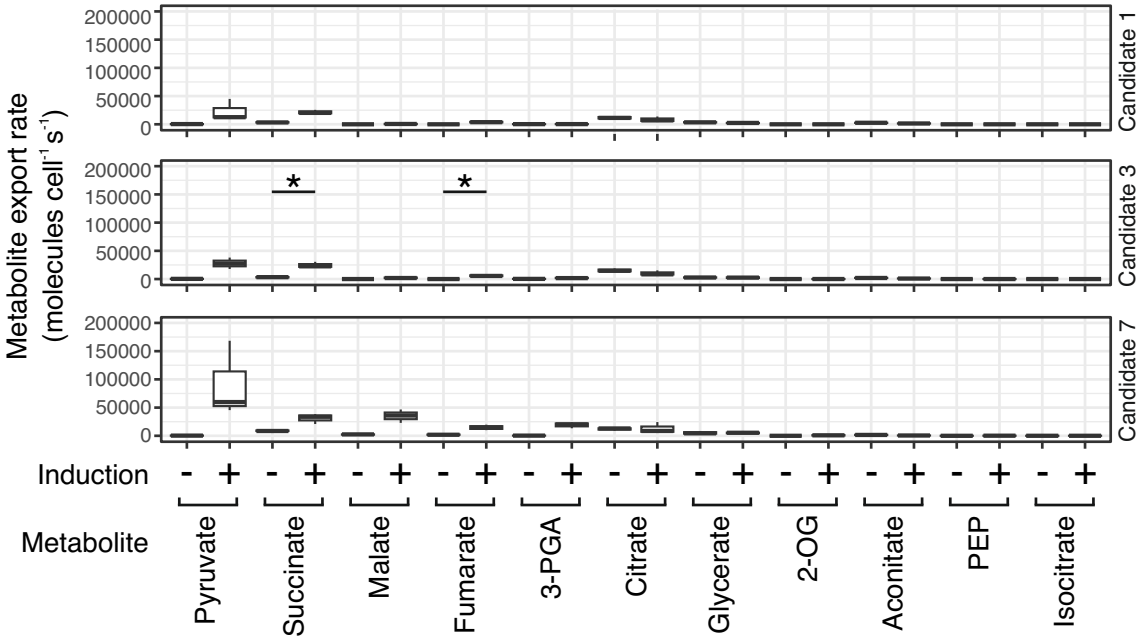


Figure 4.10: Metabolite profiling of cell culture supernatant of BW25113 *E. coli* after induction (+) or not (-) of the expression of candidate maize transporters. Asterisks indicate significant differences ($p < 0.05$) between induced and non-induced samples using t-tests with Benjamini-Hochberg correction for multiple comparisons as appropriate.

4.2.7 Generation and preliminary analysis of maize *mep3* mutant lines

Despite a lack of conclusive evidence for metabolite transport upon expression in *E. coli*, based on the bioinformatic screen it was decided that *MEP3* (candidate 1) was the most promising candidate for playing a role in the C_4 cycle in maize. It was therefore decided to generate a CRISPR/Cas9-based knock-out for this gene in maize. This was carried out by the Flemish Institute for Biotechnology (VIB), and resulted in the generation of three independent MEP3 loss of function lines: 812-09, 812-05, and 802-01. These carried a

one bp insertion, four bp deletion, and 341 bp deletion in *MEP3* respectively, all of which lead to a reading frame shift and premature stop codon (Figure 4.11).

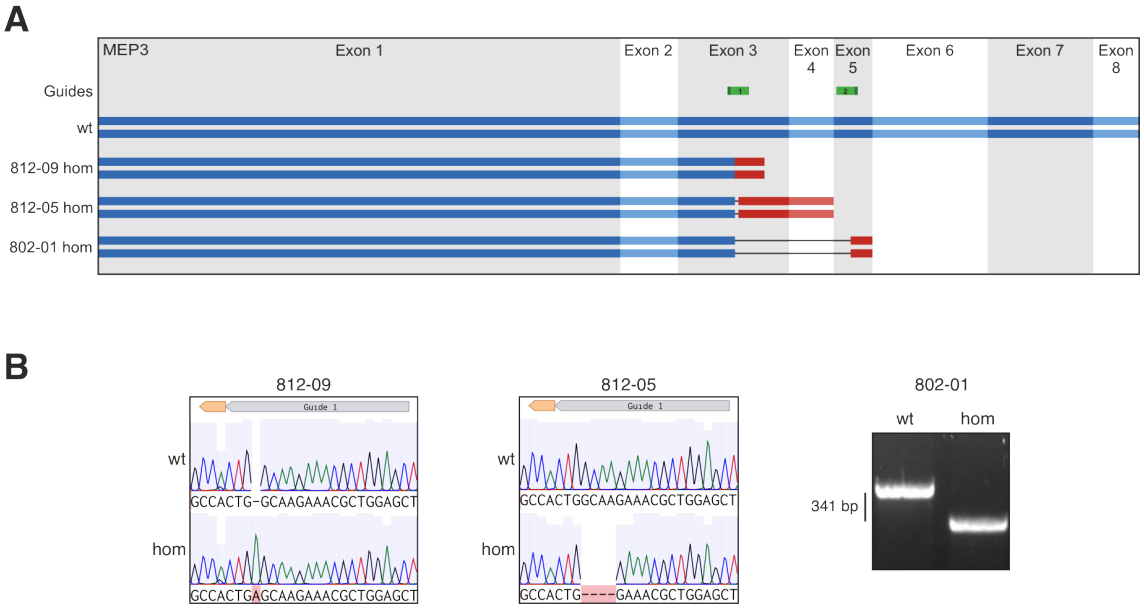


Figure 4.11: Mutation in *MEP3* observed in three independent mutant *Z. mays* lines. **A**) Cartoon representations of *MEP3* coding sequences (CDS) in segregating wild-type (wt) and homozygous mutant (hom) lines. Blue indicates regions encoding wild-type amino acid sequence. Red indicates regions encoding frame-shifted amino acid sequence, stopping where premature stop codon occurs. Green indicates predicted sgRNA binding sites. **B**) Confirmation of mutations. For lines 812-09 and 812-05, Sanger sequencing traces comparing *MEP3* sequences of segregating wild-type (wt) vs homozygous mutant (hom) plants from each mutant line at the predicted binding sites for the CRISPR sgRNA that has induced the mutation. For line 812-01, PCR gel electrophoresis comparing *MEP3* amplicons from both segregating wild-type (wt) and homozygous mutant (hom) plants. Sequencing was not carried out, as the 341 bp deletion was sufficiently large to be readily distinguished by PCR.

Maize transformation takes a long time, and at the time of writing these lines have only just been generated. Thus, the analysis presented here is only preliminary. However, in the T2 generation it was clear that *mep3* homozygous mutant plants showed a substantial phenotypic defect. Compared to segregating wild-type, *mep3* mutant plants from all three independent lines showed severe growth stunting (Figure 4.12 A), and did not successfully set seed. Furthermore, mutant plants had notably paler leaves than wild-type (Figure 4.12 B). Lack of time precluded collection of sufficient data for robust statistical analysis. However, the phenotypes are consistent and clear to the naked eye, and I would expect them to be statistically significant when sufficient data is collected.

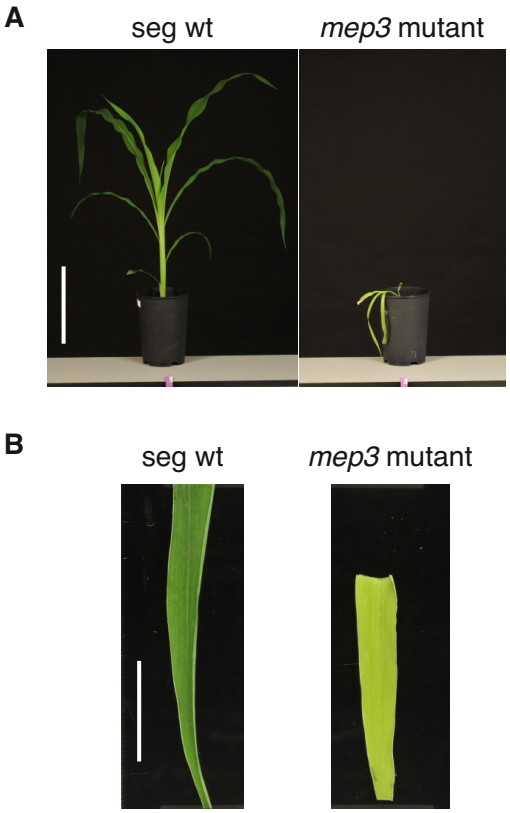


Figure 4.12: *mep3* knock-out *Z. mays* compared to segregating wild-type. Representative plants selected from a single line (2022-802-11) are shown. Pictures of representative plants from all three independent lines can be found in Appendix A.11.1. **A**) Whole plants. Scale bar = 20 cm. **B**) Close up of leaf 4. Scale bar = 2 cm.

4.3 Chapter conclusion

In this chapter I carried out a bioinformatic screen to identify a number of candidates for the missing transport steps from our understanding of the C_4 cycle in maize. The top 11 candidates were taken forward to a series of high throughput screens for metabolite transport activity, through expression in *E. coli*.

In order to screen for malate import activity, I expressed the candidate genes in the *E. coli* malate uptake null mutant *ΔdctA*, and looked for a restoration of growth on malate as a sole carbon source. In order to screen for pyruvate import activity in an analogous way I had to generate a pyruvate uptake null mutant, which did not previously exist: the quintuple mutant *ΔyfeH ΔcstA ΔbtsT ΔynfM ΔyhjX* (5m). Unfortunately, neither screens for malate or pyruvate import returned any positive results. Indeed, it seems that restoration of bacterial growth on a carbon source through exogenous expression of a transporter of that carbon source is not straightforward. Firstly, plant proteins expressed in *E. coli* may not be expressed to high levels, localise correctly to the plasma membrane, or undergo appropriate post-translational modifications and folding to function correctly. This could explain why BASS2, a known chloroplast pyruvate importer, was not able to restore 5m growth on pyruvate. In the future it would be preferable to rule out that negative screen results were not simply due to poor protein expression/localisation, rather than an actual lack of transport activity. This would require the generation of specific antibodies or protein tagging with an epitope. However, this kind of additional work was not possible within the context of a high throughput screen. Secondly, in order to restore bacterial growth, it is possible that the level of metabolite import must be fine-tuned. This could explain why greater levels of growth restoration; on malate of *ΔdctA* containing the *DctA* expression plasmid, and on pyruvate of 5m containing *CstA* or *BtsT* expression plasmids; were observed in the absence compared to presence of inducer: leaky expression from the inducible promoter caused sufficient transporter expression to restore growth, while high level of expression following induction caused a growth arrest. These inherent limitations to the

malate and pyruvate import screens were trade-offs for their relatively quick and high throughput nature. In future I would likely narrow down on fewer, stronger candidates, and test for the import of radio-labelled malate or pyruvate by overexpressed proteins reconstituted into liposomes.

In contrast, a screen for candidate metabolite exporters, carried out by looking for a reduction in growth upon candidate gene expression in *E. coli* grown on glucose, identified several candidates. Unsurprisingly, the three members of the *HATCH* family in maize caused significant growth inhibition upon expression in *E. coli*, indicating that the metabolite transport activity previously characterised (chapters 2 and 3) in *Setaria HATCH1* is conserved in maize. The *HATCH* family in general, with particular focus on the maize genes highlighted in this screen, will be discussed more in the next chapter. Four other candidate genes were identified in this screen as causing a significant decrease in growth rate upon induction of expression in *E. coli* (numbered candidates 1, 2, 3, and 7). Candidate 2 was the weakest of these candidates, owing to only causing a small reduction in *E. coli* growth rate, and the recent characterisation of an algal ortholog of this gene as a PSII-interacting thylakoid protein[278]. However, candidates 1, 3, and 7, were all considered strong candidates, which were taken forward to screening of metabolite export profiles.

Candidate 3 belongs to an uncharacterised gene family, and shows high, mesophyll-preferential expression. It was therefore primarily considered a candidate chloroplast pyruvate importer. Candidate 7 belongs to the *AVT* family, whose members have been shown to transport amino acids such as alanine. Candidate 7 was considered as a

candidate for any of the missing transport steps, owing to reasonably high expression levels in both mesophyll and bundle sheath cells. Whilst screening for the export of core C_4 cycle organic acids did not give conclusive results owing to relatively high levels of experimental variation, both candidate 3 and candidate 7 showed a potential capacity for pyruvate transport, and thus should be investigated further. Especially with regards to candidate 7, it would be interesting to screen for the export of amino acids, particularly aspartate and alanine, which would indicate a role in auxiliary C_4 cycles in maize, rather than the core NADP-ME C_4 cycle.

Finally, candidate 1, *MEP3*, was identified from the beginning as the strongest candidate for a role as a C_4 cycle transporter, owing to its extremely high, bundle sheath-preferential transcript and protein levels. It belongs to the *RER* family, whose *A. thaliana* members cause a reticulated leaf phenotype when mutated, similar to another C_4 cycle transporter, *PPT*. Expression of candidate 1 in *E. coli* led to a small but significant reduction in growth rate. Analysis of the resulting metabolite export profile revealed a correspondingly small rate of pyruvate export by induced cultures, although this was not significantly different compared to non-induced cultures. It is possible that this negative result can be explained by inefficient protein function upon exogenous expression in *E. coli*, or simply by comparatively large experimental variation. Despite this, due to the strong bioinformatic evidence in favour of MEP3 acting as a C_4 cycle metabolite transporter, it was decided to generate a knock-out mutant in maize. Whilst these lines have only just been generated, and thus the results presented here are only very preliminary, three independent *mep3* knock-out lines show clear growth stunting and pale leaves. It is interesting to note the similarity between *mep3* knock-out maize with

maize plants mutated in the proposed C_4 cycle transporter *DCT2*, which also showed yellow/pale green colouration, severe stunting, and reduced viability[52]. Further analysis of these *mep3* plant lines, for example through fluorescence and gas-exchange phenotyping to investigate disruptions to photosynthesis, pulse-chase labelling using radio-labelled CO_2 to investigate disruptions to metabolite flux, and investigation of metabolite transport by bundle sheath strand/chloroplasts, represent clear pathways forward to the potential identification of another C_4 cycle metabolite transporter in maize.

4.4 Materials and methods

4.4.1 A bioinformatic screen to identify candidate C_4 cycle transporters in *Z. mays*

The most recent version of the complete maize proteome was downloaded from Phytozome V12 (Zmays_284_Ensembl-18_2010-01-MaizeSequence.protein.fa)[172]. Protein subcellular localisation was predicted using TargetP-2.0[265,266], WoLF PSORT (command line v0.2)[267], and iPSORT[268]. The presence of TMDs was predicted using TMHMM-2.0[269], Phobius (v1.01)[270], Philius[271], PureseqTM[272]. Mesophyll and bundle sheath cell-type-specific transcriptomic data from Chang et al. (2012)[48], Tausta et al. (2014)[274], and Denton et al. (2017)[273] were obtained from Denton et al. (2017). Mesophyll and bundle sheath cell-type-specific proteomic data was obtained from Friso et al. (2010)[49]. *A. thaliana* orthologs for each maize gene was predicted using OrthoFinder[275]. Custom Python scripts were used to parse the output files from all prediction tools and subset the data to contain only genes whose gene products were

predicted to contain a cTP and at least one TMD by at least one prediction algorithm. Functions of *A. thaliana* orthologs were identified using The Arabidopsis Information Resource (TAIR)[284].

4.4.2 Construction of inducible vectors for over-expression of candidate proteins in *E. coli*

The coding sequences for maize candidate genes 1-11 (GRMZM2G138258, GRMZM2G099367, GRMZM2G088138, GRMZM2G327686, GRMZM2G145500, GRMZM2G004352, GRMZM2G074053, GRMZM2G015033, GRMZM2G078118, GRMZM2G086258, GRMZM2G179292, GRMZM2G133400), and maize *DCT2* (GRMZM2G086258) were acquired from Phytozome V12[172]. Gene models were manually verified by comparison with publicly available RNA-seq data[48] using IGV[177], and corrected if necessary. Chloroplast transit peptides were predicted using ChloroP[285], as well as by manual alignment with bacterial orthologs if present, and removed from the sequence. Codon optimisation for expression in *E. coli* was carried out using IDT Codon Optimization tool. Next, sequences were domesticated to remove BsaI and BbsI recognition sites and Golden Gate SC module adaptors were added (Appendix A.4.5-7 and A.4.13-21). The sequences were synthesised by Gene Universal (candidates 1-5 and *DCT2*) or Twist Bioscience (candidates 6-11). The coding sequence for *S. viridis BASS2* (Sevir.5G250900) had already been acquired and modified in the same way by Michael Niklaus.

These sequences were assembled into DAPG-inducible *E. coli* expression vectors (parts lists at Appendix A.4.28 tables A.29-31 and A.38-47) by Golden Gate reactions using the

restriction enzyme BsaI-HFv2 (NEB), and using the recipe and thermocycler settings at Appendix A.3.1, tables A.11 and A.12. After *E. coli* transformation and plasmid purification, correct assembly was confirmed by Sanger sequencing (Source Bioscience) using the primers in Appendix A.4.28, table A.54.

4.4.3 Construction of inducible vectors for over-expression of *E. coli* transporters in *E. coli*

The coding sequences for *E. coli DctA* and *YfeH* were amplified using primers designed to append Golden Gate SC module adaptors to each end (Appendix A.4.26 table A.23 and Appendix A.4.27 table A.25). The PCR products were purified by ethanol precipitation. Next, the PCR products were digested using BsaI-HFv2 (NEB) at 37°C for 1 hour, followed by heat inactivation at 80°C for 20 mins, and were then purified again by ethanol precipitation. To facilitate subsequent cloning into level 1 vectors, the digested PCR products were assembled into level 0 SC vectors (parts lists at Appendix A.4.26 table A.24 and Appendix A.4.27 table A.26) using the restriction enzyme BbsI-HF (NEB), and using the recipe and thermocycler settings at Appendix A.3.1, table A.11 and table A.12.

The coding sequences for *E. coli CstA, BtsT, YhjX,* and *YnfM* were acquired from Ecocyc[178], domesticated to remove BsaI and BbsI recognition sites, and Golden Gate SC module adaptors were added (Appendix A.4.22-25). The sequences were synthesised by Twist Bioscience.

The level 0 vectors were assembled into DAPG-inducible *E. coli* expression vectors (parts lists at Appendix A.4.28 tables A.48-53) by Golden Gate reactions using the restriction enzyme BsaI-HFv2 (NEB), and using the recipe and thermocycler settings at Appendix

A.3.1, table A.11 and table A.12. After *E. coli* transformation and plasmid isolation, correct assembly was confirmed by Sanger sequencing (Source Bioscience) using the primers in Appendix A.4.28, table A.54.

4.4.4 Generation of an *E. coli* quintuple mutant using recombineering

ΔyfeH was obtained from the Keio collection[283]. Recombineering for removal of *CstA*, *BtsT*, *YnfM*, and *YhjX*, was carried out according to the protocol in Sharan et al. (2009)[286], summarised below.

To generate knock-out oligonucleotides, primers were designed with ~20 bp 3' regions to amplify the desired gene expression cassette, and 50 bp homology regions to the ends of the gene being knocked out. The forward primer homology region consisted of the 47 bp prior to the start of the gene plus the start codon. The reverse primer homology region consisted of the final 21 bp of the gene plus 29 bp downstream. To generate the *CstA* and *BtsT* knock-out oligos, the *AmpR* and *SpecR* resistance cassettes respectively were amplified from the pEU-E01-His-tev-mcs-n1 (CFS cell-free Sciences Co., Ltd Matsuyama, Japan) and pOGG157 (gifted by Philip Poole) plasmids by PCR using GoTaq polymerase (Promega) (Appendix A.3.2). To prevent later contamination of the knock-out oligo PCR products with trace amounts of template plasmid, the pEU-E01-His-tev-mcs-n1 and pOGG157 plasmids were first linearised by digestion with EcoRI (NEB) and BsaI-HFv2 (NEB) respectively (37°C 60 mins, 65°C 20 mins), purified by ethanol precipitation, and diluted 1:1000 in ddH_2O. To generate the *YnfM* and *YhjX* knock-out oligos, the *KanR* resistance cassette with flanking FRT sites from *ΔyfeH E. coli* was amplified by colony PCR using GoTaq polymerase (Promega) (Appendix A.3.3). Primers

for all the above PCRs can be found at appendix A.10.1, table A.87. The amplified knock-out oligos were purified by ethanol precipitation.

To provide *E. coli* with the λ Red recombination genes, pSIM5 plasmid (gifted by Donald L. Court) was transformed into heat-shock competent cells and grown on LB agar plates containing 25 μg/mL Chloramphenicol at 32°C overnight. The next day a single colony was picked and inoculated into 5 mL LB containing 25 μg/mL Chloramphenicol and incubated overnight at 32°C, 200 rpm. The next morning, 0.25 mL of the overnight culture was transferred to 17.5 mL LB containing 25 μg/mL Chloramphenicol in a 250 mL Erlenmeyer flask and incubated at 32°C, 200 rpm until the cells reached an OD_{600} of 0.4-0.6. The λ Red recombination system was then induced by incubation in a 42°C water bath for 15 mins, with continuous shaking.

The cells were then made electrocompetent. First the cell suspension was chilled in an ice-water slurry for 10 minutes, before being transferred to 50 mL tubes and centrifuged (4600 xg, 10 mins, 4°C) (Beckman GS-15 R). The supernatant was poured off and the pellet was gently re-suspended in 1 mL ice-cold ddH_2O, before another 30 mL ice-cold ddH_2O was added. The cells were collected again by centrifugation (4600 xg, 10 mins, 4°C). Again the supernatant was discarded, and the pellet gently re-suspended in 1 mL ice-cold ddH_2O. The 1 mL suspension was transferred to a pre-chilled 1.5 mL Eppendorf tube and centrifuged (10,000 xg, 30 s, 4°C). The supernatant was carefully removed, and the pellet was re-suspended in 200 μl ice-cold water and kept on ice until use.

To introduce the knock-out oligo into the freshly made electrocompetent cells, 50 μl of competent cells and 100 ng of knock-out oligo were added to pre-chilled 0.1 cm

electroporation cuvettes (ThermoFisher). The following controls were also carried out: induced cells without DNA and un-induced cells with DNA. The cuvettes were transferred to an electroporator (Bio-Rad Gene Pulser) and electroporated (25 μF, 2.5 kV, 200 Ω). Immediately after electroporation, 500 μl LB was added to each electroporation cuvette, gently pipetted up and down to mix with the cell-DNA mixture, and transferred to 1.5 mL Eppendorf tubes. Cellular outgrowth was carried out at 32°C, 150 rpm, for 3 hours.

Next, 50 μl undiluted and 10x diluted cell suspension were plated out on LB agar containing the appropriate antibiotics (100 μg/mL Ampicillin for *CstA* knock-out, 50 μg/mL Spectinomycin for *BtsT* knock-out, 50 μg/mL Kanamycin for *YnfM* and *YhjX* knock-outs), and incubated at 32°C for 24 hours. After this, both control plates had no or very few colonies. The lack of colonies on the induced-cells-without-DNA plate indicated that the selection was working properly. The lack of colonies on the non-induced-cells-with-DNA plate indicated that there was no contaminating factor in the knock-out oligo DNA. Colonies on the induced-cells-with-DNA plates were screened using colony PCR (Appendix A.3.3). For *CstA, BtsT, YnfM, YhjX* knock-outs, the primers in Appendix A.10.1 tables A.88, A.89, A.91, A.93 were used. In each case, two primers were designed to bind within the flanking homology regions of the target gene. These would amplify bands of two different sizes if the target gene was still present or had been replaced with the antibiotic resistance cassette. One primer was further designed to bind within the antibiotic resistance cassette. This, along with one of the flanking primers, would amplify no band if the target gene was still present, but give a band of a particular size if the

target gene had been replaced with the antibiotic resistance cassette (Appendix A.10.2 figure A.2). For *CstA*, *BtsT, YnfM, YhjX* knock-outs, the PCR gels, along with expected band-sizes, can be found in Appendix A.10.2 figures A.3, A.4, A.6, A.8 respectively.

For successive knock-outs of *CstA*, *BtsT*, *YnfM*, and *YhjX*, the above procedure was carried out, and the next round was carried out starting with the confirmed mutant from the previous step. Additionally, before the removal of *YnfM* and *YhjX,* it was necessary to remove the *KanR* genes already existing in the bacterial genome (otherwise insertion of another copy of *KanR* in the genome could not be selected for). This was carried out utilising the pCMT-flp plasmid[287], which encodes FLP recombinase, according to the procedure[288], summarised below:

First pCMT-flp was transformed into the *E. coli* cell line (3m or 4m) using heat-shock, plated out on LB agar containing 100 μg/mL Ampicillin and 10 μg/mL Tetracycline, and incubated at 30°C for 24 hours. Next, a single colony was inoculated into 1 mL LB in an Eppendorf tube, and incubated in a benchtop thermomixer at 30°C, 800 rpm, for 3 hours. The temperature was then increased to 37°C for another 3 hours. This induces *FLP* expression to facilitate recombination between FRT sites in the genome, as well as inactivating the pCMT-flp plasmid's origin of replication and thus causing segregational loss of the plasmid. Aliquots of this culture were then streaked out on LB agar containing 100 μg/mL Ampicillin and incubated overnight at 37°C. The next day, individual colonies were picked and streaked first on LB agar containing either 100 μg/mL Ampicillin and 50 μg/mL Kanamycin, 100 μg/mL Ampicillin and 10 μg/mL Tetracycline, or 100 μg/mL Ampicillin only. The next day, patches that grew on Ampicillin only, but not in the

presence of Kanamycin (due to removal of *KanR* from genome) and not in the presence of Tetracycline (due to loss of pCMT-flp), were taken forward for confirmation of *KanR* loss using colony PCR (Appendix A.3.3). For removal of the *KanR* in place of *YfeH* (prior to *YnfM* mutation) and *YnfM* (prior to *YhjX* mutation), the primers in Appendix A.10.1 tables A.90 and A.92 respectively were used. In both cases, these two primers were designed to bind within the flanking homology regions of the target gene. These would amplify bands of two different sizes if the *KanR* cassette was still present or had been successfully removed by recombination between the two FRT sites. For removal of the *KanR* in place of *YfeH* and *YnfM*, the PCR gels, along with expected band-sizes can be found in Appendix A.10.2 figures A.5 and A.7 respectively.

4.4.5 Whole genome sequencing and read mapping of quintuple mutant *E. coli*

Whole genome sequencing, and read mapping and visualisation, were carried out as previously in section 2.4.8.

4.4.6 Expression of candidate proteins in *E. coli* grown on M9 minimal media and growth curve measurements

For screening growth on malate: Strains used were BW25113, *ΔdctA*, and *ΔdctA* + candidate gene-containing plasmids (see section 4.4.2). In the morning, a single colony of each bacterial strain was inoculated into 5 mL LB (+/- 10 μg/mL Gentamicin, depending on if strain contained plasmid), and incubated at 37°C, 200 rpm, for approximately 8 hours. Subsequently, 0.5 mL of this culture was inoculated into 5 mL M9 pyruvate (Appendix A.2.1) (+/- 10 μg/mL Gentamicin), and incubated at 37°C, 200 rpm, overnight. The next morning, OD_{600} measurements were taken, and the sufficient

volume of this culture was inoculated into two 5 mL M9 pyruvate (+/- 10 μg/mL Gentamicin, +/- 50 μM DAPG) cultures to give OD_{600} = 0.5. These cultures were incubated at 37°C, 200 rpm for two hours, for induction to occur. Subsequently, these cultures were washed twice in M9- (Appendix A.2.1) (+/- 10 μg/mL Gentamicin, +/- 50 μM DAPG) through centrifugation (3200 xg, 10 mins), decanting supernatant, and resuspending in ~1 mL M9- (+/- 10 μg/mL Gentamicin, +/- 50 μM DAPG) and repeating. OD_{600} measurements were taken, and sufficient volume of this culture was inoculated into 1 mL M9 malate (+/- 10 μg/mL Gentamicin, +/- 50 μM DAPG) to give OD_{600} = 0.01. 3x 100 μL of each culture were aliquoted into a clear-bottom, black-walled 96-well plate, covered in clear film, and incubated in a plate reader (BMG Labtech FLUOstar Omega), at 37°C, 500 rpm, with OD_{600} measurements taken for two days.

For screening growth on pyruvate: Strains used were BW25113, 5m, and 5m + candidate gene-containing plasmids The same protocol was followed, except the pre-growth media was M9 glucose (Appendix A.2.1) (rather than M9 pyruvate), and the experimental media (after washing) was M9 pyruvate (rather than M9 malate).

For screening growth on glucose: Strains used were BW25113, *ΔdctA*, 5m, and BW25113 + candidate gene-containing plasmids. The same protocol was followed, except M9 glucose was used for both pre-growth and experimental media.

4.4.7 Growth curve fitting

Growth curve parameters were calculated using the R package Growthcurver[289]. This package fits a logistic equation to the experimentally determined growth curve data. The logistic equation is defined by the following equation:

$$N_t = \frac{K}{1 + \left(\frac{K - N_0}{N_0}\right) e^{-rt}}$$

Where N_t = population size at timepoint t, N_0 = population size at timepoint 0, r = intrinsic growth rate (growth rate of population when resources are not limiting), and K = carrying capacity (maximum population size in given environment).

4.4.8 Generation of knock-out lines in *Z. mays* using CRISPR/Cas9 by VIB

The CRISPR guide design, construct cloning, *Z. mays* transformation, and genotyping of multiple generations of maize plants were carried out by VIB. Briefly, two guides were designed to target *MEP3,* which were predicted to bind in exon 3 and exon 5. These were incorporated into a construct and were transformed into inbred maize B104 via *A. tumefaciens*-mediated transformation of immature embryos[290,291]. Several transgenic T_0 plants were obtained and genotyped for the presence of mutations at the predicted Cas9 cut-sites by PCR amplification and sanger sequencing. Three independent T_0 plants (812-09, 812-5, 802-01) with promising mutations were backcrossed with wild-type B104. The resulting T_1 plants were checked for the presence of T-DNA through a hygromycin-resistance assay on detached leaf sections, and mutations were further confirmed using PCR amplification (for all three lines) and sanger sequencing (for 812-09 and 812-05). 1-2 T_1 plants from each line that were transgene-free and heterozygous for the mutation were self-crossed. The resulting T_2 seed was grown, and the plants were genotyped by PCR amplification and sanger sequencing as before to identify homozygous mutant and segregating wild-type plants. Photographs of plants from this generation were taken by VIB.

Chapter 5: An investigation into the HATCH family across plants, with particular focus on *Zea mays*

5.1 Chapter introduction and background

5.1.1 *HATCH* family proteins may play a role in the C_4 cycle in maize

In the previous chapter *HATCH3.1* and *HATCH1* were identified as candidates for encoding transporters involved in the C_4 cycle in *Zea mays* (maize). The *HATCH* family consists of three members in maize: *HATCH1*, and two closely related paralogs: *HATCH3.1* and *HATCH3.2*. *HATCH3.1* and *HATCH1* show moderately high, mesophyll-preferential expression, whilst *HATCH3.2* is not highly expressed (Figure 5.1 A). Whilst slightly lower, the mesophyll expression levels for *HATCH3.1* and *HATCH1* are similar to known C_4 cycle transporters in maize (Figure 5.1 B).

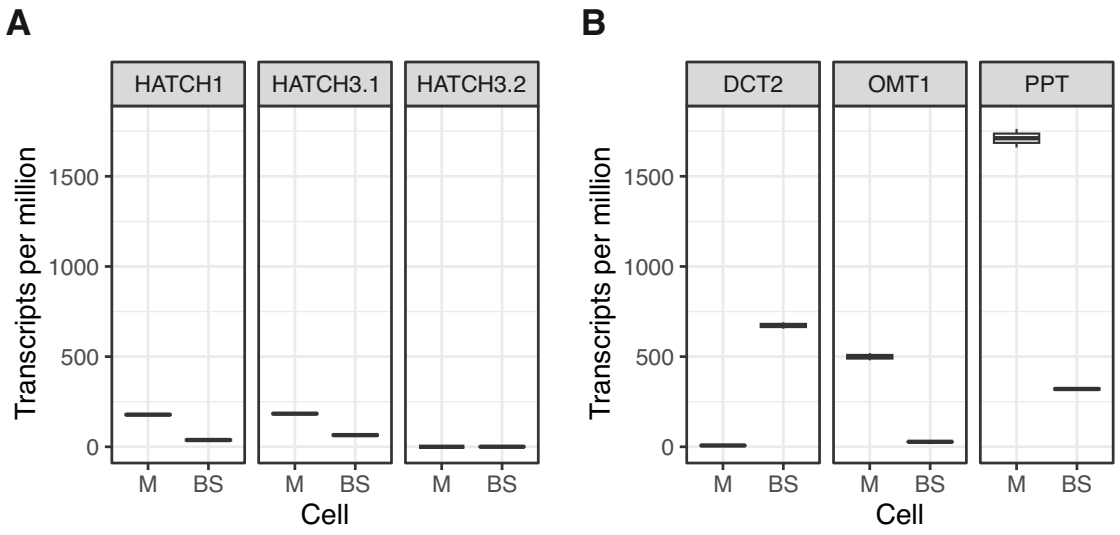


Figure 5.1: Mesophyll and bundle sheath cell-type specific transcript abundance for **A**) *HATCH* family genes and **B**) other known C_4 cycle transporters (OMT1, DCT2, PPT) in *Z. mays*. M = mesophyll. BS = bundle sheath.

It is interesting that *HATCH3.1* and *HATCH1* are preferentially expressed in the mesophyll in maize, in contrast to the bundle sheath-specific expression of *HATCH1* in *S. viridis*. Therefore, if *HATCH3.1* and/or *HATCH1* have roles in C_4 photosynthesis in maize, these are likely different to the role that *HATCH1* plays in *S. viridis*. Assuming a similar molecular function to the previously characterised *HATCH1* in *S. viridis* (see chapters 2 and 3) in transporting malate and pyruvate, some hypotheses for the function of *HATCH3.1* and *HATCH1* in maize present themselves. Firstly, HATCH3.1 and/or HATCH1 could run in the opposite direction to *S. viridis* HATCH1 and facilitate the import of pyruvate and export of malate from the mesophyll cell chloroplast (Figure 5.2 A). This kind of switching of transporter direction has precedent amongst chloroplast transporters with the change of PPT1 from chloroplast PEP importer in C_3 plants to PEP

exporter in C_4 plants (see thesis introduction). Thus *HATCH3.1* and/or *HATCH1* may encode the pyruvate importer missing from our understanding of C_4 photosynthesis in maize. *S. viridis* HATCH1 was already shown to be capable of malate export when exogenously expressed in *E. coli* (see Figure 2.3 A), indicating that transport in this transporter family may be bidirectional. Alternatively, HATCH3.1 and/or HATCH1 could run in the same direction as observed in *S. viridis* HATCH1 and facilitate malate import (and possibly pyruvate export) from the mesophyll cell chloroplast (Figure 5.2 B). This would appear to be the opposite direction to the flux of metabolites required for the C_4 cycle. However, the import of malate may be required to 'kick-start' the C_4 cycle during photosynthetic induction upon illumination. This could occur by facilitating malate import into the chloroplast to replace chloroplastic malate that would be depleted during the dark period, and which can then act as a counter-exchange substrate with oxaloacetate (OAA) through the obligate dicarboxylate exchanger OMT1[30].

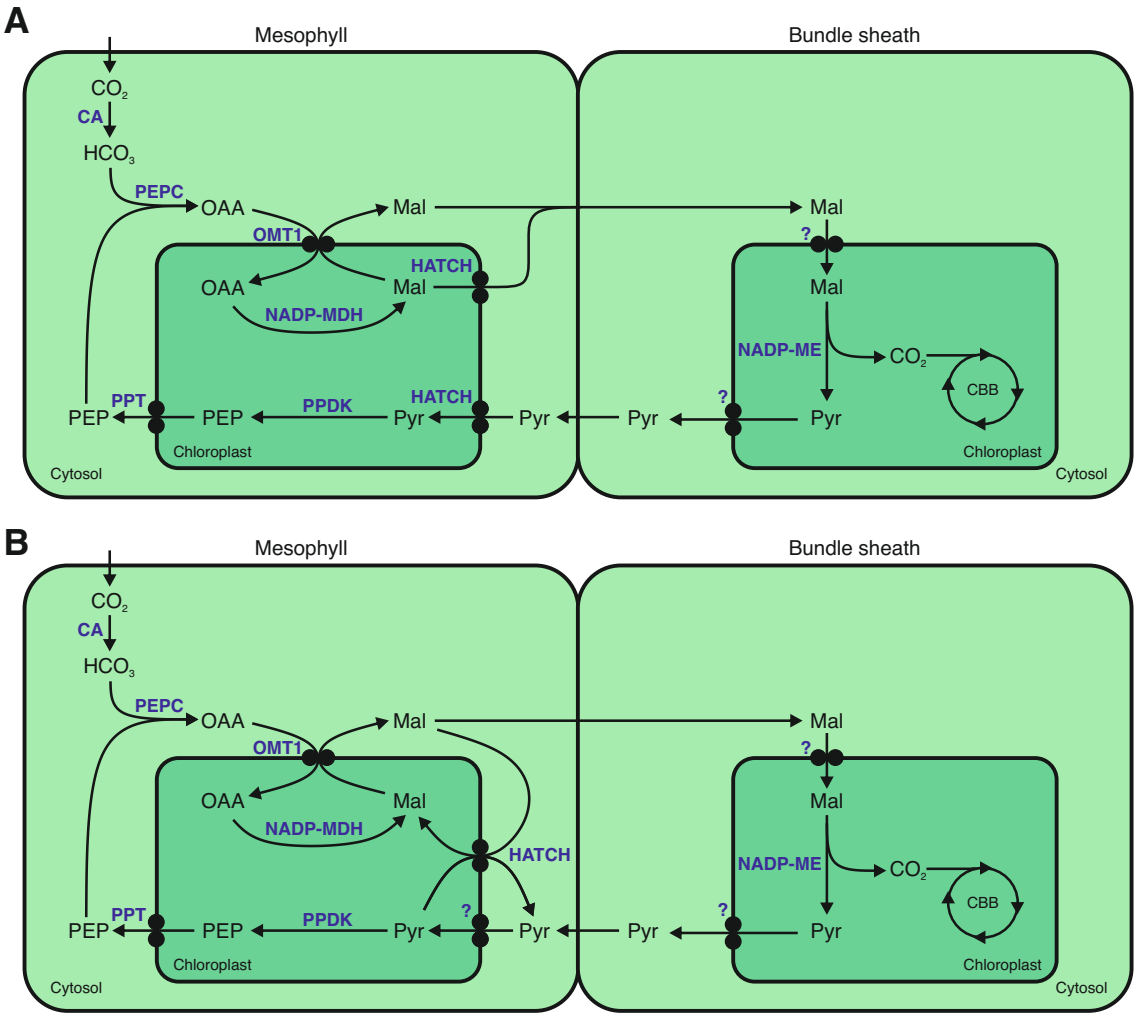


Figure 5.2: Hypotheses for HATCH3.1/HATCH1 functional involvement with the C_4 cycle in *Z. mays*. **A**) HATCH3.1/1 facilitate malate export and pyruvate import into the mesophyll chloroplast as part of the core C_4 cycle. **B**) HATCH3.1/1 act in the opposite direction to the direction of core C_4 cycle flux, to facilitate malate import (and possibly pyruvate export) to kick-start the C_4 cycle. For abbreviations see appendix A.1, table A.1.

5.1.2 Chapter aims and outline

Given the previous work in chapters 2 and 3 investigating the role of *HATCH1* and *HATCH3* in the *Setaria viridis* C_4 cycle, and the identification of *HATCH3.1* and *HATCH1* as candidate C_4 cycle transporters in chapter 4, in this chapter I first wanted to perform

a brief phylogenetic analysis of the *HATCH* family. The second half of this chapter then focusses on a more in-depth investigation into potential *HATCH* gene function in maize, with particular focus on the most highly expressed gene, *HATCH3.1*.

5.2 Results and discussion

5.2.1 The *HATCH* gene family in plants

The *HATCH* gene family is a previously uncharacterised gene family, whose members are defined by the presence of a UPF0114 PFAM domain (PF03350). *HATCH* family genes are found in land plants, algae and bacteria (Figure 5.3). Bacteria and chlorophyte algae generally possess a single *HATCH* gene. In the streptophytes, two gene duplication events have given rise to three distinct clades in angiosperms. The first duplication (D1) appears to have occurred before the divergence of land plants from charophyte algae: charophyte algae *Klebsormidium nitens*, *Interfilum paradoxum HATCH* genes form a single cluster, whilst land plants and sister charophyte algae *Mesotaenium endlicherianum*[292] *HATCH* genes form two clusters (*HATCH1/2* and *HATCH3*). Within the land plants, a second duplication (D2) has occurred, with the most parsimonious explanation being a duplication just prior to the divergence of the angiosperms. Almost all *HATCH* family members encode proteins with four predicted transmembrane domains (Figure 5.3). Most plant and algal *HATCH* family members encode proteins of approximately 300 amino acids in length, with bacterial members encoding shorter proteins, probably due to their lack of a chloroplast transit peptide (Figure 5.3). Predicted protein subcellular localisation varies across the tree (Figure 5.3). It should be

noted that this prediction may be adversely affected by poorer quality gene models in some species. However, it is noted that most HATCH1 and HATCH3 members have predicted chloroplast transit peptides (cTPs). In contrast predicted subcellular localisation amongst HATCH2 members is more variable. This is consistent with on average longer branch lengths in the *HATCH2* clade (Ø: 5.03 substitutions per site ± 0.59 (SD, N = 28)) compared to *HATCH1* (Ø: 3.90 ± 0.22 (SD, N = 33)) and *HATCH3* (Ø: 3.95 ± 0.35 (SD, N = 46)) (Figure 5.4). As greater evolutionary change has occurred in the *HATCH2* clade, this suggests that HATCH2 proteins may have functionally diverged from HATCH1 and HATCH3 (including potentially with respect to subcellular localisation).

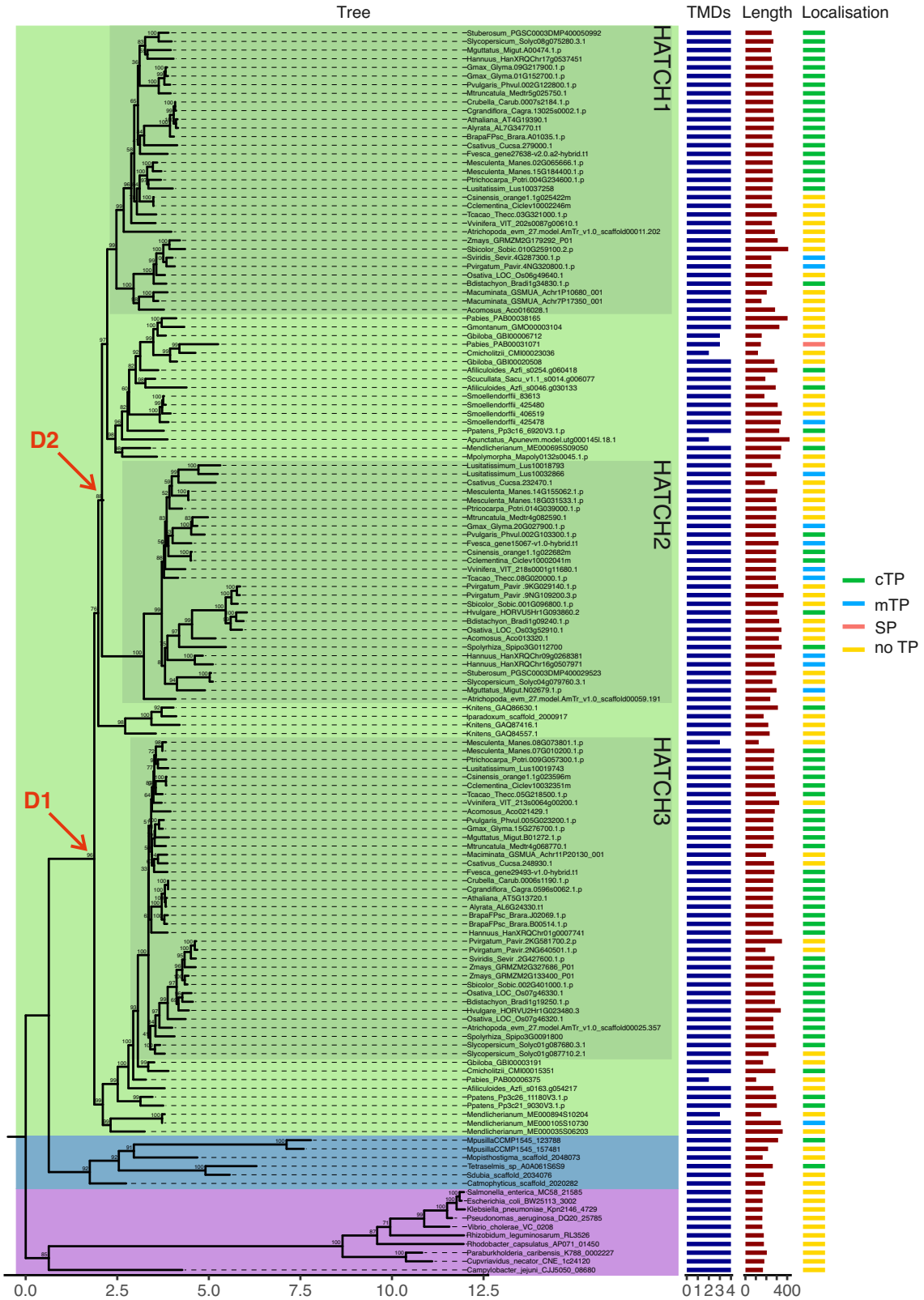

Tree
TMDs
Length
Localisation
HATCH1
HATCH2
HATCH3
D1
D2
cTP
mTP
SP
no TP
0.0
2.5
5.0
7.5
10.0
12.5
0 1 2 3 4
0
400

Figure 5.3: Phylogenetic tree depicting relationships between *HATCH* family members from across bacteria, algae, and land plants. Tree nodes are labelled with bootstrap support values. Purple = bacterial members, Blue = chlorophyte algae members, Green = streptophyte members (including both charophyte algae and embryophytes). Darker green = three *HATCH* clades in angiosperms. Gene duplication events 1 and 2 are labelled in red. On the right are graphical representations of predicted structural features (number of transmembrane domains, protein length, predicted transit peptide) for proteins encoded by each gene on the tree.

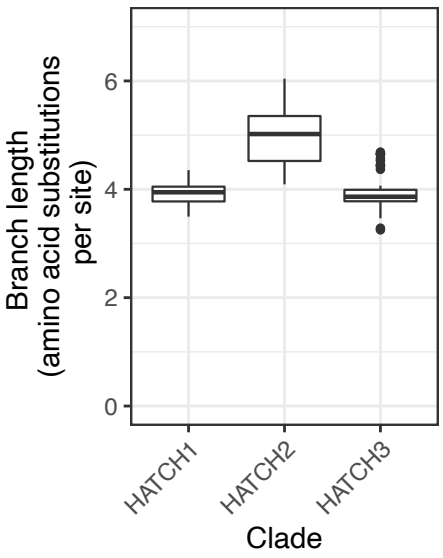


Figure 5.4: Branch length comparison between different angiosperm *HATCH* clades. Branch lengths are calculated tip to root for each species.

5.2.2 Transport properties of representative HATCH family proteins exogenously expressed in *E. coli*

In order to investigate any potential variation in the transport properties of proteins from across the HATCH family, a selection of *HATCH* family genes were over-expressed

in *E. coli*, and the supernatant was collected and analysed by metabolite mass spectrometry. All the *HATCH* genes from the C_4 monocots maize (*ZmHATCH1, ZmHATCH3.1*, *ZmHATCH3.2*) and *Setaria* (*SiHATCH1, SiHATCH3*), the C_3 monocot rice (*OsHATCH1, OsHATCH2, OsHATCH3.1, OsHATCH3.2*), and the C_3 dicot (*AtHATCH1*, *AtHATCH3*) were chosen, as well as *HATCH2* from *S. bicolor* in order to have a C_4 representative gene from the *HATCH2* clade (which is missing in both maize and *Setaria*).

Induction of *HATCH1* from *A. thaliana*, *S. italica*, and *Z. mays* compared to non-induced cultures led to a metabolite export profile broadly similar to expected (Figure 5.5). Pyruvate was exported at the highest rate, with slower but statistically significant export of dicarboxylate succinate, malate, and fumarate, as well as 3-PGA. Interestingly, induction of all three of these genes also led to a slight but still significant increase in the rate of export of PEP: another 3-carbon C_4 cycle metabolite. In contrast, *HATCH1* from rice was comparatively inactive, and did not show significant export of pyruvate or dicarboxylates (Figure 5.5).

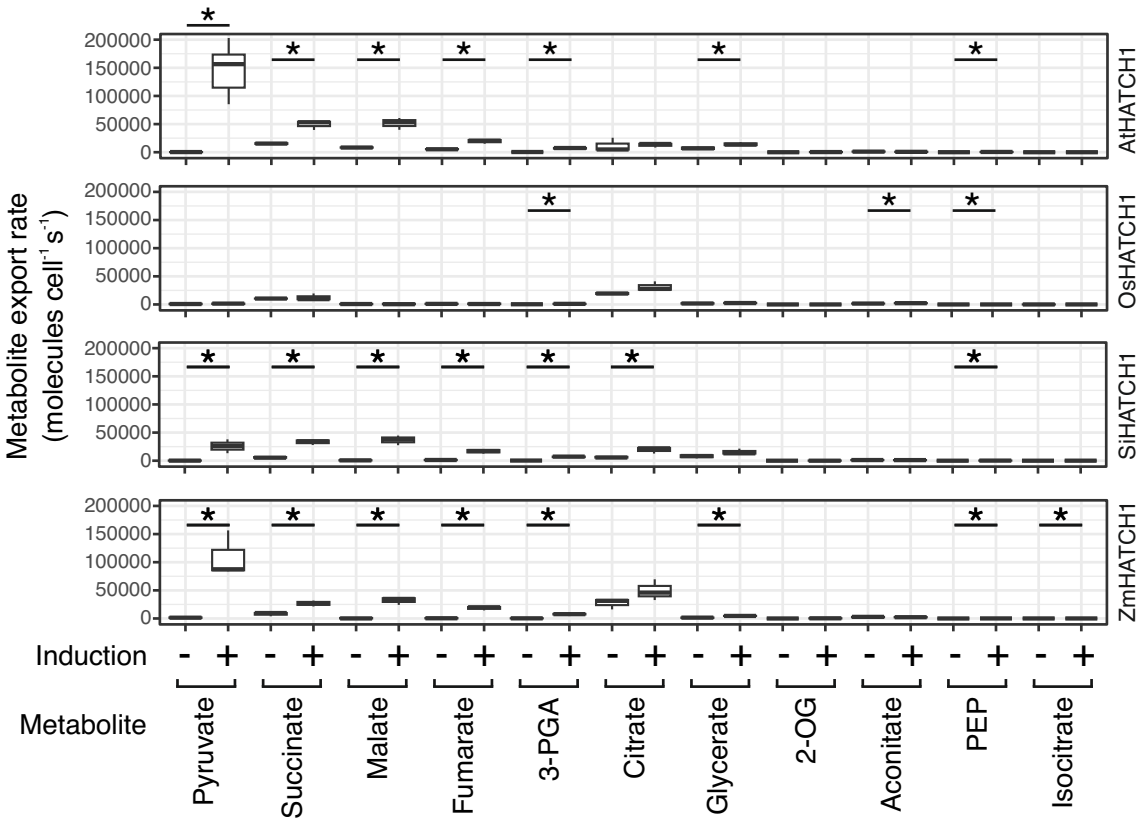


Figure 5.5: Metabolite profiling of cell culture supernatant of BW25113 *E. coli* after induction (+) or not (-) of the expression of *HATCH1* from difference species. Asterisks indicate significant differences ($p < 0.05$) between induced and non-induced samples using t-tests with Benjamini-Hochberg correction for multiple comparisons as appropriate.

Both HATCH2 proteins tested were also comparatively inactive for metabolite transport in *E. coli* (Figure 5.6). Induction of *HATCH2* from *S. bicolor* led to only slight export of succinate, fumarate, 3-PGA, and PEP, compared to non-induced control, whilst induction of *HATCH2* from rice led to no significant change in metabolite export profile. This is consistent with the higher rate of molecular evolution that has occurred within the *HATCH2* clade, suggesting that HATCH2 transporters may have either lost their

function or neofunctionalised to transport a distinct metabolite that was not picked up in this experiment.

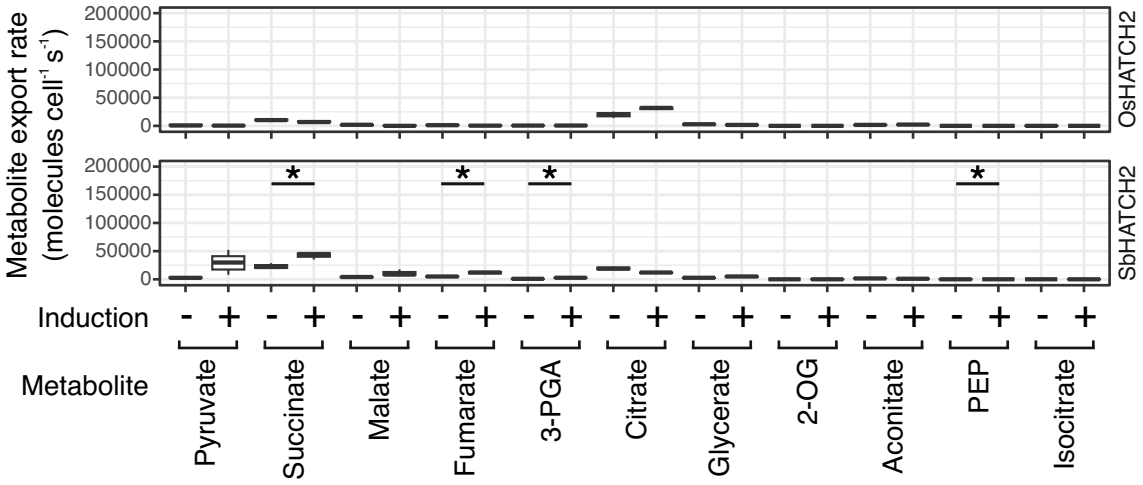


Figure 5.6: Metabolite profiling of cell culture supernatant of BW25113 *E. coli* after induction (+) or not (-) of the expression of *HATCH2* from difference species. Asterisks indicate significant differences ($p < 0.05$) between induced and non-induced samples using t-tests with Benjamini-Hochberg correction for multiple comparisons as appropriate.

The metabolite export profiles of the selected HATCH3 proteins in *E. coli* were relatively similar to those observed for HATCH1. Specifically, whilst both HATCH3.1 and HATCH3.2 from rice appeared to show some pyruvate export activity, this difference was not significant. Overall, this suggests therefore that HATCH proteins from rice have a somewhat different role (if any) compared to those in the other plants tested. *A. thaliana* and *S.italica* HATCH3, and *Z. mays* HATCH3.2 showed significant export activity, showing highest export of pyruvate, followed by four-carbon dicarboxylates (succinate,

malate, and fumarate), 3-PGA, and PEP. Interestingly, *Z. mays* HATCH3.1 showed more specific pyruvate export activity, with no significant export even of dicarboxylates.

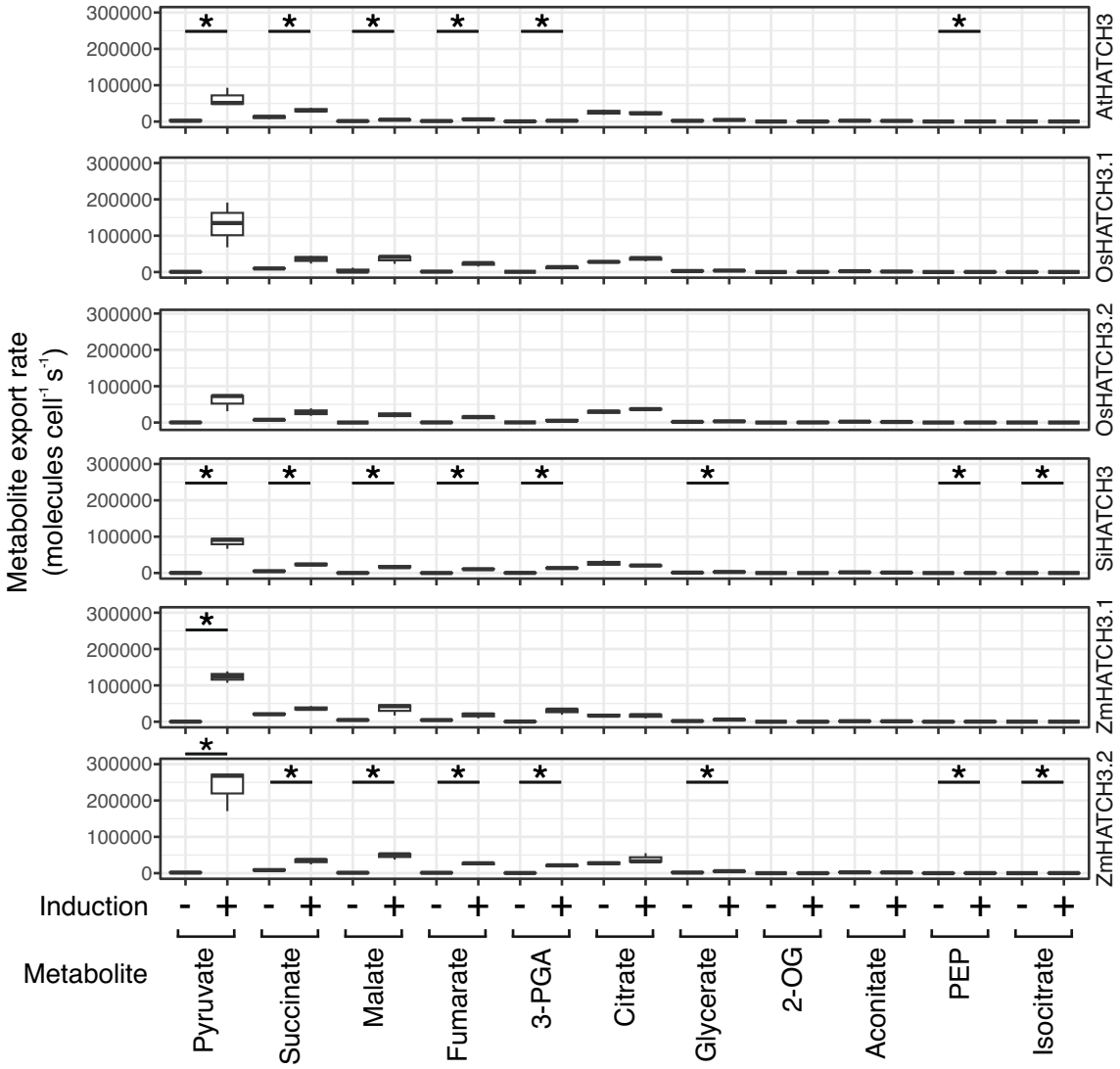


Figure 5.7: Metabolite profiling of cell culture supernatant of BW25113 *E. coli* after induction (+) or not (-) of the expression of *HATCH3* from difference species. Asterisks indicate significant differences ($p < 0.05$) between induced and non-induced samples using t-tests with Benjamini-Hochberg correction for multiple comparisons as appropriate.

5.2.3 HATCH3 is localised to the chloroplast inner envelope

HATCH1 was already showed to be chloroplast-localised in chapter 2 (see figure 2.6). Before further investigating the *in planta* function of HATCH3, we therefore wanted to first also confirm the predicted chloroplast envelope localisation of HATCH3. Thus, in the same way as in chapter 2, *C*-terminal GFP fusion proteins were transiently expressed in *Nicotiana benthamiana* leaves by agroinfiltration, followed by examination by fluorescence confocal microscopy.

Both *Z. mays* HATCH3.1-GFP (Figure 5.8 A) and *A. thaliana* HATCH3-GFP (Figure 5.8 B) showed punctate localisation around the periphery of the chloroplast, indicating localisation to the chloroplast inner envelope.

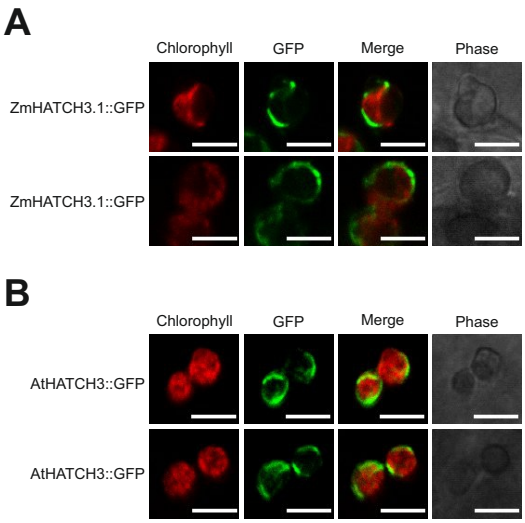


Figure 5.8: Confocal microscopy images showing the subcellular localisation of GFP-tagged *HATCH3* proteins overexpressed in *N. benthamiana* leaves. Representative images of two chloroplasts each are shown for leaves overexpressing GFP-tagged versions of **A**) ZmHATCH3.1, **B**) AtHATCH3. Scale bars represent 5 μm.

5.2.4 Generation of *hatch3.1* knock-out lines in *Z. mays* using CRISPR/Cas9

In order to investigate the role of HATCH family proteins in maize, we decided to generate a single knock-out line for the most highly expressed *HATCH* gene in maize: *HATCH3.1*, as well as a double knock-out for *HATCH3.1* and *HATCH1*. CRISPR-mediated knock-out lines were generated by the Flemish Institute for Biotechnology (VIB). Transformation of the construct targeting both *HATCH3.1* and *HATCH1* yielded no transgenic plants. This could be due to either a high editing efficiency combined with lethal phenotype, or simply a technical error with the transformation. However, transformation of the construct targeting only *HATCH3.1* was successful, and led to the generation of three independent, transgene-free, homozygous mutant lines, along with segregating wild-types. The lines were named 5A, 8B, and 10B, and each contained the same mutation in *HATCH3.1*: a 43 bp deletion between the predicted Cas9 cut-sites for the two guides used. This leads to a frame-shift and premature stop codon (Figure 5.9).

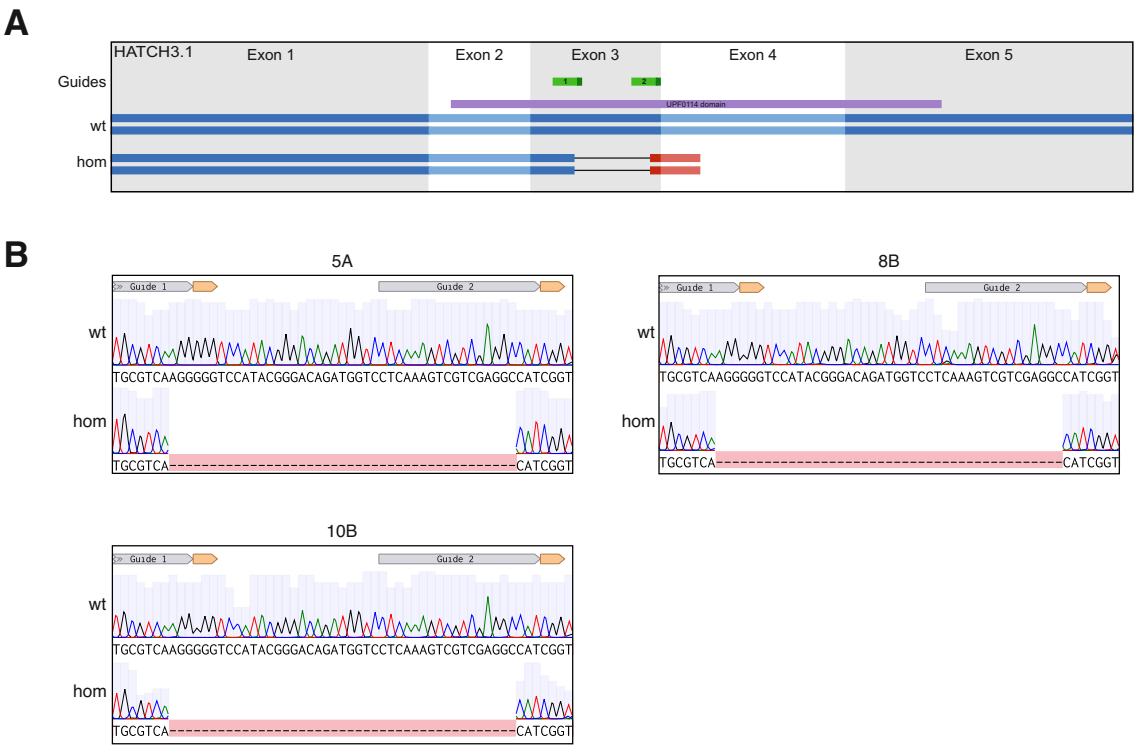


Figure 5.9: Mutation in *HATCH3.1* observed in three independent mutant *Z. mays* lines. **A**) Cartoon representations of *HATCH3.1* coding sequences (CDS) in segregating wild-type (wt) and homozygous mutant (hom) lines. Blue indicates regions encoding wild-type amino acid sequence. Red indicates regions encoding frame-shifted amino acid sequence, stopping where premature stop codon occurs. Green indicates predicted sgRNA binding sites. Purple indicates UPF0114 domain. **B**) Sanger sequencing traces comparing *HATCH3.1* sequences of segregating wild-type (wt) vs homozygous mutant (hom) plants from each mutant line (5A, 8B, 10B) at the predicted binding sites for the CRISPR sgRNAs that have induced the mutation.

5.2.5 *hatch3.1* knock-out *Z. mays* do not show inhibited growth compared to wild-type

T_2 plants from the three independent mutant lines were grown under normal greenhouse conditions, with positional randomisation to ensure a lack of positional effects on plant growth. These plants were genotyped using PCR to identify segregating

wild-type, heterozygous and homozygous mutant plants. Measurements of various physical traits were taken during the early stages of the plants' lifecycle. There was no significant difference in growth rate as determined by plant height measurements between homozygous mutant and segregating wild-type plants from all three lines (Figure 5.10 A). At 28 days after germination, the height of mutant plants was not significantly different to segregating wild-type plants from all three lines (Figure 5.10 B) (t-tests, 5A: p-value=0.662, 8B p-value=0.549, 10B p-value=0.930). Over the same time period, leaf number did not vary significantly between homozygous mutant and segregating wild-type plants from any of the three lines (Figure 5.10 C). After 35 days of growth, the plants were left to dry, and dry weight was not significantly different between homozygous mutant and segregating wild-type plants from all three lines (Figure 5.10 D) (t-tests, 5A p-value=0.363, 8B p-value=0.283, 10B p-value=0.616). Photos of homozygous and segregating wild-type T_3 plants can be found at Appendix A.13.1, Figure A.10. Thus, *hatch3.1* knock-out leads to no visible phenotypic defect in maize, and *HATCH3.1* is therefore not essential for C_4 photosynthesis.

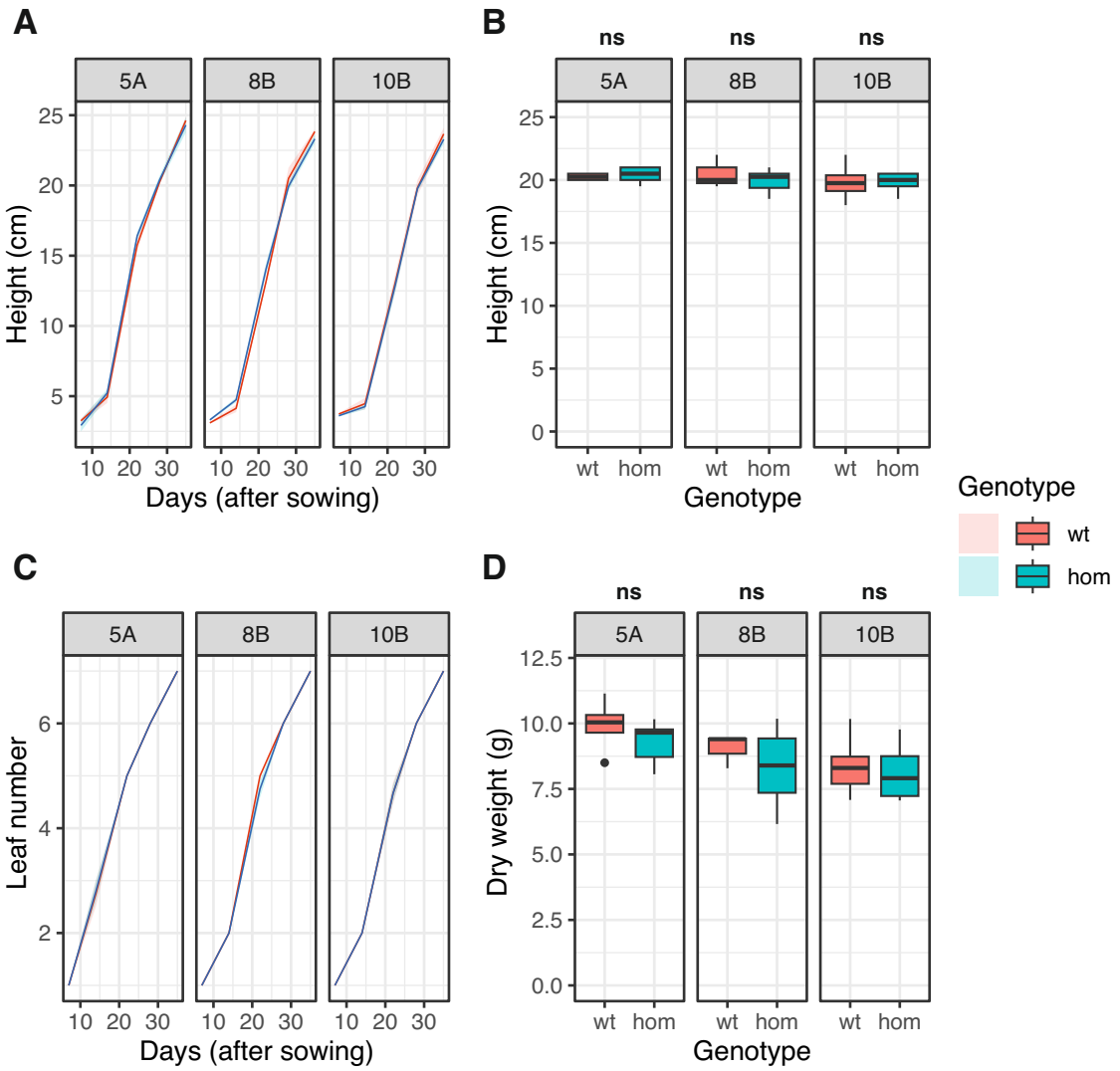


Figure 5.10: Plant growth measurements of *hatch3.1* knock-out *Z. mays* lines. **A**) Plant height measurements (cm) (to base of youngest fully expanded leaf) over time. **B**) Plant height measurements (cm) at 28 days after germination. **C**) Plant leaf number over time. **D**) Dry weight of plants, left to dry 35 days after germination. ns indicates non-significant differences ($p>0.05$) for mutant lines compared to segregating wild-types using t-tests. wt = segregating wild-type. hom = homozygous *HATCH3.1* mutant.

5.2.6 *hatch3.1* knock-out *Z. mays* do not show altered CO_2 assimilation compared to wild-type

In order to investigate whether HATCH3.1 is part of the C_4 cycle in maize, an LI-6800 Portable Photosynthesis System was used to measure photosynthetic gas exchange at varying CO_2 concentrations. The initial slope of an A-C_i curve, which plots intercellular CO_2 concentration (C_i) against net CO_2 assimilation rate (A), is determined by the activity of the two enzymes that facilitate carbon entry into the C_4 cycle: CA and PEPC[293], the latter of which would likely be affected by reduced C_4 cycle metabolite flux across the mesophyll chloroplast envelope. In the interests of maximising limited time, it was therefore decided to generate A-C_i curves to cover only this lower range of CO_2 concentrations. If HATCH3.1 functions as part of the C_4 cycle, facilitating pyruvate import and/or malate export from the mesophyll chloroplast, then the initial slope of the A-C_i curve would be expected to be reduced in the *hatch3.1* knock-out lines compared to wild-type. In contrast, if HATCH3.1 transports metabolites in the opposite direction to the main C_4 cycle flux, in order to kick-start the C_4 cycle during photosynthetic induction, then *hatch3.1* knock-out lines may actually show a steeper initial slope of the A-C_i curve compared to wild-type, as futile metabolite cycling would be reduced, and the efficiency of the C_4 cycle would be increased. The obtained A-C_i curves show no consistent differences between *hatch3.1* knock-out and segregating wild-type lines (Figure 5.11 A). At approximately ambient CO_2 concentration (400 ppm) there was no significant difference in A between *hatch3.1* knock-out and segregating wild-type plants from any of the three lines (Figure 5.11 B) (t-tests, 5A: p-value=0.195, 8B p-value=0.203, 10B p-

value=0.869). Thus, no evidence is found to support a role for HATCH3.1 in the C_4 cycle in maize.

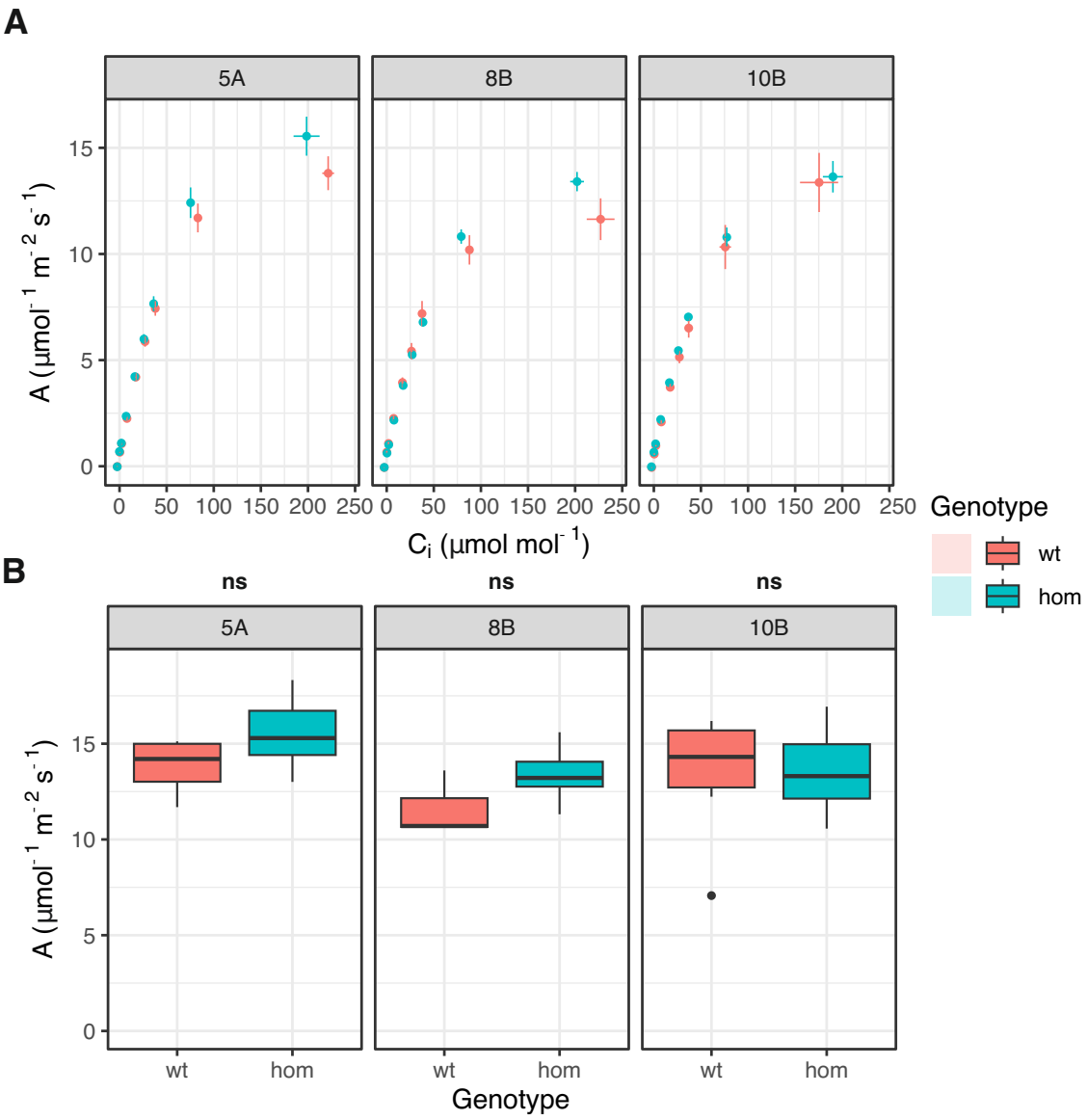


Figure 5.11: Characterisation of gas-exchange of *hatch3.1* knock-out *Z. mays* lines compared to segregating wild-type under varying CO_2 concentrations. **A**) Net CO_2 assimilation rate (A) over a range of intercellular CO_2 concentrations (C_i). Each point represents the average of 3-9 plants per line. Error bars represent standard error of the mean. **B**) Net CO_2 assimilation rate at approximately atmospheric CO_2 concentration (400 ppm). ns indicates non-significant differences ($p>0.05$) for mutant lines compared to segregating wild-types using t-tests. wt = segregating wild-type. hom = homozygous *HATCH3.1* mutant.

5.2.7 *hatch3.1* knock-out *Z. mays* do not show altered photosynthetic induction compared to wild-type

In order to further investigate whether HATCH3.1 functions to 'kick-start' the C_4 cycle during photosynthetic induction, photosynthetic induction experiments were carried out using the LI-6800 Portable Photosynthesis System. *hatch3.1* knock-out and segregating wild-type plants were dark-adapted for 30 minutes, before the light was turned on, and the change in net CO_2 assimilation rate (A) was measured over time. If HATCH3.1 does function to 'kick-start' the C_4 cycle during photosynthetic induction, then *hatch3.1* knock-out lines would be expected to show slower photosynthetic induction compared to wild-type. The data again showed no consistent differences between *hatch3.1* knock-out and segregating wild-type lines (Figure 5.12 A). At 300s after transition from dark to high light, there was no significant difference in A between *hatch3.1* knock-out and segregating wild-type plants from any of the three lines (Figure 5.12 B) (t-tests, 5A: p-value=0.903, 8B p-value=0.405, 10B p-value=0.476). Thus, no evidence is found to support a role for HATCH3.1 in 'kick-starting' the C_4 cycle in maize.

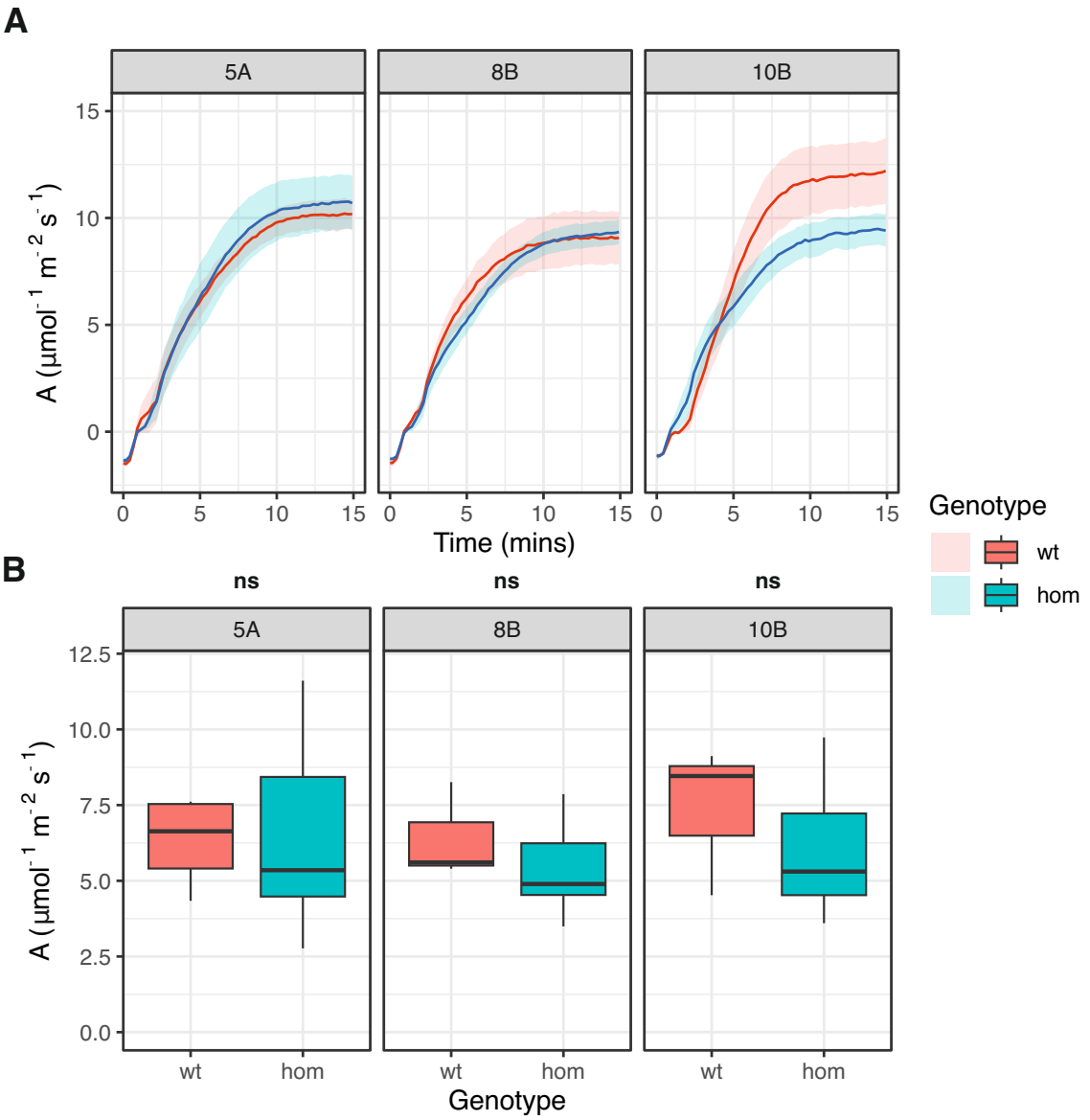


Figure 5.12: Characterisation of gas-exchange of *hatch3.1* knock-out *Z. mays* lines compared to segregating wild-type during photosynthetic induction upon transition from dark to high light. **A**) Change in net CO_2 assimilation rate over time after transition from dark to high light. Line represents the average of 3-7 plants per line. Error ribbons represent standard error of the mean. **B**) Net CO_2 assimilation rate 300s after transition from dark to high light. ns indicates non-significant differences ($p>0.05$) for mutant lines compared to segregating wild-types using t-tests. wt = segregating wild-type. hom = homozygous *HATCH3.1* mutant.

5.3 Chapter conclusion

In this chapter I carried out the first preliminary experimental characterisation of HATCH proteins from additional plant species other than *Setaria*, with particular focus on the C_4 plant maize.

Metabolite export of profiles of *E. coli* exogenously expressing HATCH1 and HATCH3 proteins from *S. italica, Z. mays, S. bicolor,* and *A. thaliana* confirmed a general trend for pyruvate transport, but also with lower levels of transport for other metabolites including four-carbon dicarboxylates (succinate, malate, fumarate), 3-PGA and PEP. This broad metabolite transport profile is consistent with the proposed channel-like mechanism and structure of SiHATCH1 (see chapter 2). Both HATCH2 proteins tested did not cause significant export of these metabolites, which is consistent with the longer branch lengths within the *HATCH2* clade compared to the rest of the *HATCH* family phylogenetic tree. This increased rate of molecular evolution could indicate protein neofunctionalization to transport a different range of metabolites, and/or loss of function, the latter of which is consistent with the loss of *HATCH2* from several species, including *A. thaliana*, *S. italica*, and *Z. mays*.

As *HATCH1* and *HATCH3.1* show relatively high, mesophyll-specific expression in *Z. mays*, and had been identified as potential C_4 cycle transporters in a screen in the previous chapter, in this chapter I further investigated this possibility. First, ZmHATCH3.1 was confirmed to be localised to the chloroplast, as computationally predicted. Next, owing to the slightly higher expression of *ZmHATCH3.1* compared to *ZmHATCH1*, three independent CRISPR/Cas9-based *hatch3.1* knock-out lines were generated. Phenotypic

analysis of these mutant lines revealed no significant disruption to CO_2 assimilation, as would be predicted if *HATCH3.1* played a key role in the C_4 cycle. Similarly, *hatch3.1* mutant plants showed no significant disruption in photosynthetic induction, as would be predicted if *HATCH3.1* played a key role in metabolite transport to 'kick-start' the C_4 cycle upon transition to high light. It is worth noting that the similar expression patterns of maize *HATCH1* and *HATCH3.1*, and similar transport profiles (upon exogenous expression in *E. coli*) of all three HATCH proteins in maize, indicate that knock-out of *HATCH3.1* may be compensated for by *HATCH1* and/or upregulated *HATCH3.2*. Thus *hatch3.1* plants could be phenotypically unaffected even if HATCH1 and HATCH3.1 do together carry out an important role in the C_4 cycle. The failure to recover any transgenic *hatch1,hatch3.1* double mutant plants is consistent with mutations in both of these genes being embryo lethal. However, it is also possible that this transformation merely failed due to technical reasons. Next, it would be useful to test if *HATCH1* and/or *HATCH3.2* are upregulated in response to *hatch3.1* knock-out, which would indicate compensation. Depending on the outcome of this, it could be interesting to a) generate *hatch1* single mutant lines, to confirm if mutation in this gene alone is sufficient to cause a phenotype, b) re-attempt the generation of a *hatch1,hatch3.1* double mutant, to confirm if the resulting plants are viable, or c) attempt to generate a *hatch3.1,hatch3.1,hatch3.2* triple mutant, to investigate the effect of a complete removal of HATCH transport activity on C_4 photosynthesis in maize.

If a role for HATCH proteins in C_4 photosynthesis in maize is not identified, it is possible that HATCH proteins are involved in the supply of pyruvate into the chloroplast for the MEP (methylerythritol 4-phosphate) pathway and terpenoid biosynthesis. In *A. thaliana*

this role is carried out by BASS2[99], but *BASS2* is not expressed in maize. To test this hypothesis, *hatch* mutant maize seedlings could be treated with mevastatin, an inhibitor of the alternate, cytosolic mevalonic acid pathway for terpenoid biosynthesis, as has been done with *bass2* mutant *A. thaliana*[99]. In a similar vein, it could also be worth investigating phenotypes relating to salinity tolerance and fatty acid metabolism, both of which are disrupted in wheat (*Triticum aestivum*) and rapeseed (*Brassica napus*) *bass2* mutants respectively[101,294]. However, non-photosynthetic chloroplast pyruvate import in this case would also occur via whatever transporter is responsible for pyruvate import for the C_4 cycle, and thus it might be expected that *hatch* mutant maize would again not show a phenotype. Overall, it would probably not be worth pursuing these experiments until the C_4 pyruvate transporter had been identified.

5.4 Materials and methods

5.4.1 Phylogenetic tree inference

Where available, whole proteome files were downloaded from Phytozome V13[172], Plaza Gymnosperm (https://bioinformatics.psb.ugent.be/plaza/versions/gymno-plaza/)[295], Fernbase (https://fernbase.org)[296], the University of Zurich Hornwort genome database (https://www.hornworts.uzh.ch/en/download.htmL)[297], and Ensembl bacteria (http://bacteria.ensembl.org)[298]. Proteomes were grouped into Orthogroups using OrthoFinder[275] (v2.3.11). The two orthogroups containing plant HATCH proteins and *E. coli* YqhA, which constitute the *HATCH* family, were combined into a single file. In order to increase phylogenetic resolution, a number of additional algal HATCH protein

sequences were identified using PhycoCosm (https://phycocosm.jgi.doe.gov/phycocosm/home)[299] and added to the orthogroup file. The sequences were aligned using MAFFT (L-INS-1 option)[300] (v7.508). Alignments were subject to manual inspection, and poor-quality gene models were removed. A phylogenetic tree was then generated using IQ-TREE[301] (v2.2.0.3), with 1000 bootstrap replicates.

5.4.2 Phylogenetic tree branch length calculation

Branch lengths, from tip to root for each gene, were calculated using the distRoot function from the adephylo R package[302], and averaged across the angiosperm members of the *HATCH1*, *HATCH2*, and *HATCH3* clades.

5.4.3 Prediction of protein structural features

Transmembrane domains were predicted using DeepTMHMM[143]. Target peptides were predicted using TargetP 2.0[142].

5.4.4 Generation of GFP fusion constructs for expression in *N. benthamiana*

The coding sequence for *Z. mays HATCH3.1* (GRMZM2G327686) and *A. thaliana HATCH3* (AT5G13720) were acquired from Phytozome V13[172]. These sequences were modified by the addition of Golden Gate SC1 module adaptors, domestication to remove BsaI recognition sites, and removal of the stop codon (Appendix A.5.6-7). The sequences were synthesised by Twist Bioscience.

These sequences were assembled into binary vectors for transient expression of *C*-terminally GFP-tagged proteins in *N. benthamiana* (see Appendix A.5.8 Tables A.61 and

A.62) by Golden Gate reactions using the restriction enzyme BsaI-HFv2 (NEB), and the recipe and thermocycler settings at Appendix A.3.1, table A.11 and table A.12 respectively. After *E. coli* transformation and plasmid isolation, correct assembly was confirmed by Sanger sequencing (Source Bioscience) using the primers in Appendix A.5.8, table A.63.

5.4.5 Transient protein over-expression in *N. benthamiana* leaves

See section 2.4.4.

5.4.6 Confocal microscopy analysis of transiently transformed *N. benthamiana* leaves

See section 2.4.5

5.4.7 Generation of knock-out lines in *Z. mays* using CRISPR/Cas9 by VIB

The CRISPR guide design, construct cloning, *Z. mays* transformation, and genotyping of multiple generations of maize plants were carried out by VIB. Briefly, two guides were designed for each gene: both predicted to bind in exon 3 for *HATCH3.1*, and in exon 2 for *HATCH1*. These were incorporated into two constructs: one to target *HATCH3.1* only, and one to target *HATCH3.1* and *HATCH1* simultaneously. The constructs were transformed into inbred maize B104 via *A. tumefaciens*-mediated transformation of immature embryos[290,291]. Transformation of the construct targeting *HATCH3.1* and *HATCH1* yielded no transgenic plants. Transformation of the construct targeting only *HATCH3.1* yielded several transgenic plants. These T_0 plants were genotyped for the presence of mutations at the predicted Cas9 cut-sites by PCR amplification and sanger sequencing. Three independent T_0 plants (755-05, 755-08, 755-10) with promising

mutations were backcrossed with wild-type B104. The resulting T_1 plants were checked for the presence of T-DNA through a hygromycin-resistance assay on detached leaf sections, and mutations were further confirmed using PCR amplification and sanger sequencing. 3-4 T_1 plants from each line that were transgene-free and heterozygous for the mutation were self-crossed. Resulting T_2 seed was sent to me in order to experimentally analyse this generation.

5.4.8 *Z. mays* growth conditions

Maize growth was carried out in a greenhouse facility at 28/20°C (day/night) with long photoperiod (16 hour light, 8 hour dark). Seeds were germinated in a tray of damp, medium vermiculite (Sinclair), covered with a transparent, plastic cover to maintain humidity. When the seedlings were ~10 cm tall they were transferred to pots containing 2/3rd soil (John Innes No.3 Mature Plant Compost) and 1/3rd medium vermiculite. Plants were watered approximately every other day and positionally randomised once per week.

5.4.9 Plant DNA extraction

Plant DNA extraction was attempted as in section 3.4.4, but the quality of DNA was found to be insufficient. Therefore, for maize I instead used a method for isolating higher quality DNA. First ~2 cm^2 of leaf tissue was taken per plant, placed into a 2 mL Eppendorf tube containing a metal bead. 500 µl CTAB buffer (Appendix A.3.6, table A.22) was added and the samples were lysed in a tissue lyser (2 mins, 28 s^{-1}) (Qiagen TissueLyser II). The samples were incubated at 65°C for 30 mins, before 500 µl chloroform:isoamyl alcohol (24:1) was added and vortexed to mix. The samples were centrifuged (13,000

rpm, 10 mins), 350 μl supernatant was removed to a fresh 2 mL tube, and 2.5 volumes 100% ethanol was added and vortexed to mix. After 10 minutes incubation at room temperature, the samples were centrifuged again (13,000 rpm, 10 mins) and the supernatant was discarded. The pellet was washed in 70% ethanol, before the pellet was left to dry at 65°C and finally resuspended in 100 μl ddH_2O.

5.4.10 PCR genotyping of transgenic plants

Plants were genotyped by PCR using Phusion polymerase (NEB) (Appendix A.3.4) with primers designed to amplify the region surrounding the known edit in *HATCH3.1* (Appendix A.12.1, table A.94). Amplification of wild-type sequence gave a 216 bp band, whilst amplification of mutant sequence gave a 173 bp band. PCR products were visualised using gel electrophoresis to determine which plants were segregating wild-types (only 216 bp band), heterozygous (both 216 bp and 173 bp bands) or homozygous (only 173 bp band) for the mutation.

5.4.11 *Z. mays* growth measurements

Plant height (from base of plant to base of youngest fully expanded leaf) and leaf number were taken every week. After 35 days of growth, plant watering was ceased, and the plants were left to dry. After the plants had dried the stems were cut as close to the base as possible and aerial parts of the plant were weighed.

5.4.12 *Z. mays* photosynthetic gas-exchange and fluorescence measurements

Photosynthetic gas-exchange and fluorescence measurements were carried out using a LICOR 6800 (LICOR Bioscience). For all measurements the youngest fully expanded leaf

was used. For A-C_i curves, leaves were first clamped into the LICOR 6800 and equilibrated for 15 minutes to 400 μmol mol^{-1} CO_2, 500 μmol s^{-1} flow rate, 0.2 kPa over pressure, 50% relative humidity, 35°C leaf temperature, and irradiance of 1800 μmol photons m^{-2} s^{-1}. Gas exchange measurements were subsequently taken at 3-minute intervals at the following CO_2 concentrations: 400, 200, 100, 75, 50, 25, 10, 5, 0 μmol mol^{-1}, with all other parameters kept constant. For light induction experiments, leaves were first clamped into the LICOR 6800 and equilibrated for 30 minutes to 400 μmol mol^{-1} CO_2, 500 μmol s^{-1} flow rate, 0.2 kPa over pressure, 50% relative humidity, 35°C leaf temperature, and irradiance of 0 μmol photons m^{-2} s^{-1}. Irradiance was then increased to 1800 μmol photons m^{-2} s^{-1} and gas exchange measurements were taken every 15 s for 15 minutes.

Chapter 6: Discussion

6.1.1 Characterisation of a C_4 cycle malate/pyruvate transporter in *Setaria*

In this thesis I have continued the characterisation of a chloroplast-localised transporter involved in the C_4 cycle in *Setaria viridis*. HATCH1 had previously been identified as a candidate C_4 transporter, based on its predicted transmembrane domains and chloroplast transit peptide, as well as high, bundle-sheath specific expression. This made it a candidate for the missing bundle sheath chloroplast malate importer and/or pyruvate exporter involved in the C_4 cycle, and indeed preliminary characterisation by Michael Niklaus indicated that HATCH1 may carry out both these roles. I showed that HATCH1 protein localises to the chloroplast inner envelope, in agreement with computational predictions. I also functionally characterised HATCH1 through exogenous expression in *Escherichia coli*, and provided evidence that HATCH1 facilitates specific pyruvate export and malate import. Interestingly, and in contrast to other known chloroplast metabolite transporters, the transport activity displayed by HATCH1 upon expression in *E. coli* resembled a channel more than an antiporter, given that pyruvate export could occur even in the absence of external malate. This proposed transport mechanism is in agreement with a preliminary *in silico* prediction of the structure of the *E. coli* ortholog of HATCH1, YqhA. Both HATCH1 and YqhA monomers are predicted to consist of a row of four α-helices, and thus could not facilitate transmembrane metabolite transport in this form. However, a YqhA octamer is predicted to form a pore of approximately the right size to facilitate the transport of malate and pyruvate, which would naturally occur relatively freely in a channel-like manner rather than by strict

antiport. Whilst this computational prediction necessarily remains speculative, a collaboration is now under way with Dr Jani Bolla to confirm the oligomeric number, and hopefully to eventually obtain a crystal structure, of YqhA and HATCH1. Whilst the bacterial system used for characterisation in this thesis has proved useful, it will eventually be necessary to carry out more detailed analysis of transport kinetics through reconstitution of the transporter in liposomes.

In order to confirm that the transport activity observed upon exogenous expression of HATCH1 in *E. coli* is related to a C_4 cycle function *in planta*, Michael Niklaus had previously generated and carried out a preliminary characterisation of RNAi-mediated *HATCH1* knock-down *S. viridis* lines. The rationale behind generating knock-down rather than knock-out lines was that knocking out a putative C_4 cycle transporter was thought likely to be lethal. Indeed, the knock-down plants did show a significant reduction in photosynthetic rate, as determined by A/C_i and light response curves, which I subsequently showed by flux balance analysis modelling to be consistent with reduced flux through a bundle sheath chloroplast malate/pyruvate transporter. Furthermore, I demonstrated that isolated bundle sheath strands from knock-down plants showed significantly slower malate-dependent pyruvate production, indicating a reduction in the expected transport reaction *in planta*. Despite this, it was surprising that the *HATCH1* knock-down lines showed no noticeable disruption to growth, and it was hypothesised that under ordinary growth conditions sufficient C_4 cycle flux could be maintained by either the remaining ~10% *HATCH1* expression in the knock-down lines, or by compensation by its ortholog *HATCH3*, or through the operation of ancillary C_4 cycles (PCK or NAD-ME).

In order to test the above hypotheses, and determine if *HATCH1* is a necessary component of the NADP-ME C_4 cycle in *S. viridis*, I generated three independent *hatch1* knock-out and one *hatch1,hatch3* double knock-out lines using CRISPR/Cas9. I extensively analysed the resulting plants morphologically, photosynthetically, and transcriptomically. As with the *HATCH1* knock-down plants, all the knock-out lines showed significantly reduced CO_2 assimilation rates under a range of CO_2 concentrations and light intensities, which was associated with reduced estimated maximal rates of both carboxylation enzymes, PEPC and rubisco. However, unlike the knock-down lines, the knock-out lines showed a consistent growth defect compared to wild-type. Whilst it is difficult to make direct comparisons between the *HATCH1* knock-down and knock-out lines due to their being grown at different times by different people, it is likely that the complete elimination of HATCH1 function in the knock-out lines contributed to their enhanced phenotypic disruption. Surprisingly however, the reduction in growth rate of the *hatch1* knock-out lines, though consistent, was relatively minor, with knock-out plants still growing to maturity and setting seed. Thus, although HATCH1 enabled high rates of flux through the C_4 cycle, it was not necessary for the C_4 cycle to operate.

A transcriptomic analysis of HATCH1 knock-out plants showed minor disruptions in expression of several NADP-ME subtype C_4 cycle genes in the knock-out lines. However, there was no evidence for upregulated NAD-ME or PCK expression, indicating that C_4 cycle flux is probably not being re-routed through ancillary cycles based on these decarboxylation enzymes. However, owing to the apparently relatively high PCK activity levels even in *Setaria* without high *PCK* expression[204], it was deemed necessary to more directly investigate alterations in metabolic flux in *hatch1* knock-out plants. Thus, a more

complete investigation into disruptions in metabolic flux resulting from *HATCH1* knock-out using pulse-chase labelling is currently under way in collaboration with John Lunn and Regina Feil at the Max Planck Institute for Plant Physiology in Golm. Should the results of this experiment prove inconclusive, it would also be possible to investigate the dependence of *hatch1,hatch3* mutant seedlings on PCK by growing them on a medium containing the PCK inhibitor 3-mercaptopicolinic acid[20,303]. Furthermore, the *hatch1*,*hatch3* double knock-out line showed no significant differences in morphology, photosynthesis or gene expression levels compared to the *hatch1* single knock-out lines, indicating that *HATCH3* does not compensate for *HATCH1*.

Overall, therefore we must conclude that HATCH1, whilst playing a role in the import of malate and export of pyruvate from the bundle sheath chloroplast within the NADP-ME subtype C_4 cycle in *S. viridis*, is likely not necessary for the NADP-ME C_4 cycle to function. Additional transporter(s) presumably exist to support the reduced but still substantial C_4 cycle function in *hatch1,hatch3* knock-out *S. viridis*. For example, it is possible that bundle sheath chloroplast dicarboxylate import involves DCT4 (a highly expressed paralog of maize DCT2), in conjunction with an unidentified malate or aspartate uniporter. Bundle sheath pyruvate export could be facilitated by the same unknown pyruvate exporter that operates in the maize C_4 cycle, given that *HATCH* genes are not highly expressed in the bundle sheath in maize, and thus must rely on another pyruvate transport system.

6.1.2 Identification of further candidate C_4 cycle transporters in maize

Maize is the most widely grown crop plant in the world[8] and thus it is surprising that three C_4 cycle transport steps (bundle sheath chloroplast malate import and pyruvate export, and mesophyll chloroplast pyruvate import) remain unexplained. The bioinformatic and bacterial-based experimental screens in this thesis were an attempt to address this knowledge gap. The bacterial-based screens for malate and pyruvate import were more difficult than expected to set up. A known chloroplast pyruvate importer, BASS2, failed to restore growth of *E. coli* deficient in pyruvate uptake on pyruvate as a sole carbon source, likely due to a failure of the protein to express, fold, or localise correctly. In the absence of more in depth confirmation of exogenous protein expression and localisation, the failure of these screens to return any positive candidates cannot be taken as evidence against any of these proteins carrying out malate or pyruvate import roles *in planta*. Although unlikely to return any positives, it may also be worth carrying out a similar screen for the restoration of *ΔdctA E. coli* on aspartate, given that it has been hypothesised that the missing bundle sheath chloroplast dicarboxylate importer may transport aspartate, rather than malate[52].

In contrast, the screen for metabolite exporters was more successful, and given that many transporters are capable of bidirectional transport, positives from this screen should also be considered as candidate importers. Several candidates should now be taken forward for further study. Candidate 3 (GRMZM2G088138) should be investigated further for pyruvate import activity, whilst candidate 7 (GRMZM2G074053) requires an experimental re-evaluation of its localisation before further investigation into its ability

to transport either malate or pyruvate or (given its relation to amino acid transporters) aspartate or alanine, which are involved in NAD-ME and PCK subtype C_4 cycles.

All three members of the *HATCH* family in maize also gave positive results in the bacterial metabolite exporter screen and, given the previously described role for HATCH1 in the C_4 cycle in *Setaria*, were naturally considered as strong candidates for C_4 cycle roles in maize. I thus carried out a further investigation into HATCH proteins in maize, including experimental validation of HATCH3.1 chloroplast localisation, as well as arranging for the generation of *hatch3.1* knock-out lines in maize. Unfortunately, these lines showed no growth or photosynthetic phenotypic disruption. Further transcriptomic analysis, as well as the generation of double or triple knock-out lines to remove all HATCH function from maize, would be necessary to confirm or deny a role for HATCH proteins in the C_4 cycle in maize.

Despite causing only a small reduction in *E. coli* growth upon exogenous expression on glucose, which was associated with a correspondingly minor (non-statistically significant) increase in pyruvate export rate, the most promising maize C_4 cycle transporter candidate identified in this work is arguably MEP3 (GRMZM2G138258). *MEP3* is expressed to extremely high levels in the maize bundle sheath cell, indicating a possible role in malate import or pyruvate export. The similar reticulated mutant phenotype between *A. thaliana* mutants *rer4* (mutated in the ortholog of *MEP3*)[276] and *ppt* (mutated in *PPT*, the C_4 cycle PEP transporter)[80,81], as well as metabolic analysis of closely related *re* and *rer3* showing disruption in metabolic pathways downstream of pyruvate[276], lend support to the hypothesis that maize MEP3 could transport pyruvate.

Therefore, I ordered the generation of *mep3* mutant maize lines. Preliminary analysis reveals a clear growth stunting and pale-leaf phenotype that is consistent with disruption of a core C_4 cycle transporter, very similar to the mutant phenotype observed in maize *dct2* plants[52]. Further investigation of these lines, both photosynthetically using the LI-6800, as well as metabolically using bundle sheath strand-based transport assays such as those developed in this thesis, as well as pulse-chase labelling, are currently being planned. It will also be necessary to further investigate the transport properties of the MEP3 protein. It is possible that MEP3 is not able to function efficiently in *E. coli* due to low expression or partially incorrect localisation. Experiments are currently underway to confirm if this is the case. Subsequently, the next step may be to try expressing this protein in an alternate expression system, such as yeast, and ideally reconstituting it into liposomes for detailed analysis of transport properties.

6.1.3 Potential biotechnological applications of the expression of HATCH transporter proteins in *E. coli*

The expression of HATCH family proteins from across plants in *E. coli* grown on glucose as a sole carbon source led to the accumulation of a varying array of metabolites in the cell culture medium. In general, highest export rates were observed for the dicarboxylates malate and succinate, and particularly the monocarboxylate pyruvate. Commercial demand exists for all of these compounds, for example as food additives and dietary supplements[304–306]. Pyruvate is also used in the production of a number of pharmaceutical precursors[307–310], whilst succinate and malate have the potential to be

used as bulk chemical precursors that could be produced sustainably by biological means as an alternative to derivation from crude oil[311–313]. There is potential here to exploit the natural diversity in transport profiles of HATCH proteins exogenously expressed in *E. coli* as part of the bio-industrial process to produce these metabolites. The specificity and range of metabolites could be extended through directed evolution on *E. coli* expressing these transporters. Further exploration of the full breadth of the HATCH proteins in plants, algae and bacteria may also reveal novel transport functions: for example it is possible that proteins belonging to the HATCH2 subfamily have evolved to transport quite a different range of metabolites to HATCH1 and HATCH3. Combining *E. coli* metabolic mutants engineered to divert metabolism into the metabolite of interest, with the specific export of that metabolite through an exogenously expressed HATCH protein, could lead to higher product yield, as well as reduced costs caused by a simplified purification process if the metabolite has only to be extracted from the supernatant rather than from within the cells themselves. In the case of pyruvate production, the pyruvate uptake quintuple mutant generated for the first time in this thesis could be a useful chassis for minimising the re-uptake of exported pyruvate. A patent application has been filed for HATCH1 with this goal in mind.

6.1.4 Increasing our understanding of the C_4 cycle and potential biotechnological applications

Ever since its discovery in the 1960s[314], it has been a major goal to engineer C_4 photosynthesis into C_3 crops such as rice and wheat, and thus confer enhanced

photosynthetic efficiency, water-use efficiency, and nitrogen-use efficiency. This would be expected to significantly boost yield and help meet the target of feeding 10 billion people by 2050[315], particularly as temperatures rise and water scarcity becomes more common due to climate change. Improving photosynthetic efficiency is also a particularly attractive line of investigation as previous yield gains in C_3 crops have been achieved in other areas such as efficiency of light capture and proportion of biomass partitioned into grain, and have now become limited by photosynthetic rate[316,317]. Attempts to engineer C_4 photosynthesis into C_3 plants have so far been hampered by several gaps in our understanding of this trait, one of which has been a complete list of the intracellular metabolite transporters used in the C_4 cycle. The work in this thesis characterises the newly discovered HATCH1 protein from *Setaria*, which carries out bundle sheath chloroplast malate import and pyruvate export as part of the C_4 cycle. Whilst it is likely not the sole malate/pyruvate transporter operating in the C_4 cycle in *Setaria*, the high transport activity of this transporter upon exogenous expression in *E. coli* suggests that it would likely be sufficient to carry out this role in an attempted engineered C_4 cycle in a non-C_4 plant. Thus, we now have for the first time a complete parts list for a C_4 cycle that will be invaluable for future C_4 engineering efforts.

An ultimate goal is to engineer fully-fledged, two-cell C_4 photosynthesis, such as that found in maize, into a C_3 crop. However, this will further require a more complete understanding of the development of the characteristic 'Kranz' anatomy that underpins the two-cell trait, and as such is still predicted to be many years away from completion. A potentially quicker route to engineering C_4 photosynthesis would be to take inspiration from the few lineages that have evolved C_4 photosynthesis within the context

of a single cell. Terrestrial single-cell C_4 lineages, such as *Bienertia sinuspersici* and *Suaeda aralocaspica* still retain a spatial separation of PEPC and rubisco through separated cytoplasmic domains containing dimorphic chloroplasts[318], which may be similarly difficult to engineer to Kranz anatomy. However, several aquatic plant lineages, such as *Hydrilla verticillata*, can perform an inducible form of single-cell C_4 photosynthesis under low CO_2 conditions[319], which lacks any obvious morphological changes and thus may represent a simpler route for the engineering of C_4 photosynthesis into a C_3 plant. The C_4 cycle in *H. verticillata* is thought not to be as efficient as that in terrestrial two- and single-cell C_4 species, which is likely caused by CO_2 leakage from the site of secondary re-fixation, due to the lack of significant morphological adaptations to increase diffusive resistance to CO_2[318]. Therefore, the potential yield benefits from engineering such a cycle into C_3 crops would likely not be as large as if a two-cell C_4 cycle was introduced. However, modelling has suggested that, despite the associated inefficiencies, the addition of an auxiliary single-cell C_4 cycle based on the *H. verticillata* system into the mesophyll cells of a C_3 leaf could lead to increased CO_2 assimilation rates under conditions of high light and/or low intercellular CO_2 concentration (e.g. caused by stomatal closure in water-limited conditions)[320]. Whilst engineering two-cell C_4 photosynthesis into a C_3 plant remains unfeasible, any small yield benefit from a single-cell C_4 cycle would be significant, as would simply be the proof-of-concept that a functional C_4 cycle can be engineered into a C_3 plant.

Several attempts have been made to engineer a single-cell C_4 cycle into a C_3 plant, but none have been successful[321–323]. However, all of these previous attempts have focussed solely on the over-expression of soluble C_4 cycle enzymes, and not intracellular

metabolite transporters. The metabolites involved in the C_4 cycle are transported across the chloroplast envelope to varying degrees in C_3 plants[32,79,99], but these transport reactions occur at significantly higher rates in C_4 plants[90,140,324]. Therefore, chloroplast envelope transporters likely must be over-expressed alongside the soluble enzymes in order to sustain sufficient flux through an engineered C_4 cycle. The identification of HATCH1 presents an opportunity to attempt to engineer a functional single-cell C_4 cycle that includes the necessary transport reactions. To this end I designed a single-cell C_4 cycle based on the carboxylation and decarboxylation reactions catalysed by the reversible enzyme NADP-ME (Figure 6.1). Maize NADP-ME3 and *A. thaliana* NADP-ME4 are chloroplast localised, and primarily catalyse the decarboxylation of malate to pyruvate[325–327]. Over-expression of maize NADP-ME3 in *A. thaliana* leads to an accumulation of pyruvate and depletion of malate[328]. In contrast, *A. thaliana* NADP-ME2 is cytosolic, has a lower K_m for pyruvate than malate[325] and over-expression in *A. thaliana* leads to an accumulation of malate and depletion of pyruvate[329], suggesting a preference for the reverse, carboxylation reaction. Thus, either ZmNADP-ME3/AtNADP-ME4 and AtNADP-ME2 could be used in conjunction with HATCH1 to concentrate CO_2 around rubisco in the chloroplast (Figure 6.1). I have designed constructs to express each of the genes encoding these proteins under a mesophyll-specific promoter in *A. thaliana*. Plants have subsequently been transformed and are currently being analysed by another member of the lab.

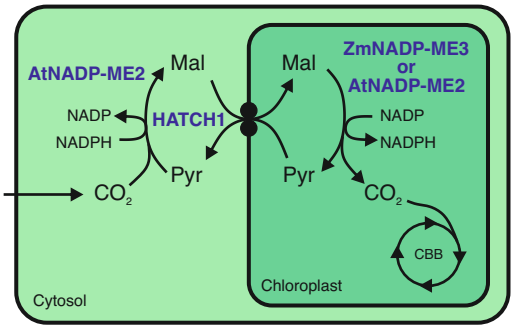


Figure 6.1: Proposed single-cell C_4 cycle for engineering. For abbreviations see appendix A.1, table A.1.

Overall, the work in this thesis has pushed forward our understanding of C_4 photosynthesis in the model species *S. viridis*, as well as providing promising avenues for further research in maize. The biotechnological utilisation of this knowledge has the potential to contribute to our ongoing struggle for food security.

Appendix

A.1 Abbreviations

Table A.1 Abbreviations from figure legends

Name	Abbreviation	Type
Calvin-Benson-Bassham cycle	CBB	Pathway
Glyoxylate cycle	Glyox	Pathway
Mitochondrial electron transport chain	METC	Pathway
Oxidative Pentose Phosphate pathway	OPPP	Pathway
Tricarboxylic acid cycle	TCA	Pathway
1,3-bisphosphoglycerate	1,3-BPG	Metabolite
2-oxoglutarate	2-OG	Metabolite
3-phosphoglycerate	3-PGA	Metabolite
6-phosphogluconate	6-PG	Metabolite
Acetyl Coenzyme A	Acetyl-CoA	Metabolite
Adenosine diphosphate	ADP	Metabolite
Adenosine monophosphate	AMP	Metabolite
Adenosine triphosphate	ATP	Metabolite
ADP-Glucose	ADP-Glc	Metabolite
Alanine	Ala	Metabolite
Amino acid	AA	Metabolite
Aspartate	Asp	Metabolite
Citrate	Cit	Metabolite
Coenzyme A	CoA	Metabolite
Erythrose 4-phosphate	Ery4P	Metabolite
Fructose 1,6-bisphosphate	Fru1,6BP	Metabolite
Fructose 6-phosphate	Fru6P / Fruc6P	Metabolite
Glucose 1-phosphate	Glc1P	Metabolite

Glucose 6-phosphate	Glc6P	Metabolite
Glutamate	Glu	Metabolite
Glutamine	Gln	Metabolite
Glyceraldehyde 3-phosphate	GAP / G3P	Metabolite
Glycine	Gly	Metabolite
Hydroxypyruvate	OH-Pyr	Metabolite
Inorganic phosphate	P_i	Metabolite
Isocitrate	ICit	Metabolite
Malate	Mal	Metabolite
Methylerythritol 4-phosphate	MEP	Metabolite
Mevalonate	MVA	Metabolite
Nicotinamide adenine dinucleotide	NAD	Metabolite
Nicotinamide adenine dinucleotide phosphate	NADP	Metabolite
Nicotinamide adenine dinucleotide (reduced)	NADH	Metabolite
Nicotinamide adenine dinucleotide phosphate (reduced)	NADPH	Metabolite
Oxaloacetate	OAA	Metabolite
Phospho*enol*pyruvate	PEP	Metabolite
Pyrophosphate	PP_i	Metabolite
Pyruvate	Pyr	Metabolite
Ribose 5-phosphate	R5P	Metabolite
Ribulose 5-phosphate	Ru5P	Metabolite
Ribulose 1,5-bisphosphate	RuBP	Metabolite
Sedoheptulose 7-phosphate	Sed7P	Metabolite
Serine	Ser	Metabolite
Succinate	Succ	Metabolite
Sucrose	Suc	Metabolite
Sucrose phosphate	SucP	Metabolite
Triose phosphate	TP	Metabolite

Uridine diphosphate	UDP	Metabolite
Uridine triphosphate	UTP	Metabolite
Xylulose 5-phosphate	Xu5P	Metabolite
1-aminocyclopropane 1-carboxylate synthase 10	ACS10	Enzyme
6-phosphogluconate dehydrogenase	6PGDH	Enzyme
Aconitate hydrolyase / aconitase	AH	Enzyme
ADP-glucose pyrophosphorylase	ADP-Glc Ppase	Enzyme
Alanine aminotransferase	AlaAT	Enzyme
Aspartate aminotransferase	AspAT	Enzyme
Aspartate aminotransferase (chloroplastic)	cpAspAT	Enzyme
Aspartate aminotransferase (cytosolic)	cyAspAT	Enzyme
Aspartate aminotranseferase (mitochondrial)	mtAspAT	Enzyme
ATP-citrate lyase	ACL	Enzyme
ATP synthase	ATPS	Enzyme
Carbonic anhydrase	CA	Enzyme
Citrate synthase	CS	Enzyme
Fumarate hydratase / fumarase	FH	Enzyme
Glucose 6-phosphate dehydrogenase	Glc6PDH	Enzyme
Glyceraldehyde 3-phosphate dehydrogenase	GAPDH	Enzyme
Glutamate oxoglutarate aminotransferase	GOGAT	Enzyme
Glutamine synthetase	GS	Enzyme
Glycine decarboxylase	GDC	Enzyme
Hydroxypyruvate reductase	HPR	Enzyme
Isocitrate dehydrogenase	IDH	Enzyme
NAD(P)-malate dehydrogenase	NAD(P)-MDH	Enzyme
NAD(P)-malic enzyme	NAD(P)-ME	Enzyme
Nitrate reductase	NR	Enzyme
Nitrite reductase	NIR	Enzyme

Phospho*enol*pyruvate carboxylase	PEPC	Enzyme
Phospho*enol*pyruvate carboxykinase	PCK	Enzyme
Phosphoglycerate kinase	PGK	Enzyme
Pyruvate kinase	PyrK	Enzyme
Pyruvate, phosphate dikinase	PPDK	Enzyme
Ribose 5-phosphate isomerase	R5PI	Enzyme
Ribulose 5-phosphate epimerase	Ru5PE	Enzyme
Ribulose 1,5-bisphosphate carboxylase/oxygenase	Rubisco	Enzyme
Succinate dehydrogenase	SDH	Enzyme
Transaldolase	TA	Enzyme
Transketolase	TK	Enzyme
2-oxoglutarate/malate transporter	OMT	Transporter
Bile Acid/Sodium symporter	BASS	Transporter
Dicarboxylate and tricarboxylate carrier	DTC	Transporter
Dicarboxylate carrier	DIC	Transporter
Dicarboxylate transporter	DCT	Transporter
Glucose 6-phosphate / P_i translocator	GPT	Transporter
Mitochondrial pyruvate carrier	MPC	Transporter
Na^+/H^+ antiporter	NHD	Transporter
Nucleotide transporter (plastidic)	NTT	Transporter
Phosphate carrier	PIC	Transporter
Phospho*enol*pyruvate / P_i translocator	PPT	Transporter
Triose phosphate translocator	TPT	Transporter
Uncoupling protein	UCP	Transporter
Xylulose 5-phosphate / P_i translocator	XPT	Transporter
Ferredoxin oxidised/reduced	$Fd_{ox/red}$	Protein

A.2 Media

A.2.1 M9 minimal media

1 L M9 minimal media was prepared by preparing the solution in Table A.2, including the appropriate carbon source as in Table A.3 [330]. For the single-carbon-source media, amount of carbon source in each media is standardised to contain the same number of carbon atoms.

Table A.2: Composition of M9 minimal media, minus C source

Volume	**Stock solution**	**Content**	**Final concentration**
100 mL	M9 salt solution (10x) pH 7.2	NaH_2PO_4	33.7 mM
		KH_2PO_4	22 mM
		NaCl	8.55 mM
		NH_4Cl	9.35 mM
1 mL	$MgSO_4$ (1 M)	$MgSO_4$	1 mM
0.3 mL	$CaCl_2$ (1 M)	$CaCl_2$	0.3 mM
1 mL	Biotin (1mg mL^{-1})	Biotin	1 μg
1 mL	Thamin (1mg mL^{-1})	Thamin	1 μg
10 mL	Trace elements solution (100x) pH 7.5	EDTA	134 μM
		$FeCl_3$-$6H_2O$	31 μM
		$ZnCl_2$	6.2 μM
		$CuCl_2$-$2H_2O$	760 nM
		$CoCl_2$-$2H_2O$	420 nM
		H_3BO_3	1.62 μM
		$MnCl_2$-$4H_2O$	81 nM

Table A.3: Carbon source added to M9 minimal media

Media	Volume	Stock solution	Final concentration
M9 -	-	-	-
M9 glucose	20 mL	Glucose (20% w/v)	22.2 mM (0.4 %)
M9 malate	33.3 mL	L-Malic acid (1 M)	33.3 mM (0.45 %)
M9 pyruvate	22.2 mL	Sodium pyruvate (1 M)	44.4 mM (0.39 %)
M9 fumarate	33.3 mL	Sodium fumarate dibasic (1 M)	33.3 mM (0.44 %)
M9 succinate	33.3 mL	Sodium succinate dibasic (1 M)	33.3 mM (0.39 %)
M9 aspartate	33.3 mL	Aspartic acid (1 M)	33.3 mM (0.39 %)
M9 glucose + malate	20 mL	Glucose (20% w/v)	22.2 mM (0.4 %)
	5 mL	L-Malic acid (1 M)	5 mM
M9 glucose + fumarate	20 mL	Glucose (20% w/v)	22.2 mM (0.4 %)
	5 mL	Sodium fumarate dibasic (1 M)	5 mM
M9 glucose + succinate	20 mL	Glucose (20% w/v)	22.2 mM (0.4 %)
	5 mL	Sodium succinate dibasic (1 M)	5 mM
M9 glucose + aspartate	20 mL	Glucose (20% w/v)	22.2 mM (0.4 %)
	5 mL	Aspartic acid (1 M)	5 mM

A.2.2 *S. viridis* bundle sheath strand isolation and assay solutions

Table A.4: Composition of bundle sheath strand isolation resuspension solution

Adjust to pH 8.

Component	**Final concentration**
Sorbitol	0.3 M
HEPES (pH 8)	25 mM
$MgCl_2$	2 mM
KH_2PO_4	1 mM

Table A.5: Composition of bundle sheath strand isolation enzyme solution

Add all components aside from BSA, $CaCl_2$, and $MgCl_2$. Stir at 50°C until dissolved. Adjust to pH 5.7. Add remaining components.

Component	**Final concentration**
Cellulase R-10 (Duchefa Biochemie)	2 % (w/v)
Macerozyme R-10 (Duchefa Biochemie)	0.3 % (w/v)
Pectolyase Y-23 (Duchefa Biochemie)	0.05 % (w/v)
Sorbitol	0.6 M
MES	100 mM
BSA	0.1 % (w/v)
$CaCl_2$	1 mM
$MgCl_2$	1 mM

Table A.6: Composition of bundle sheath strand assay buffer 1 (3-PGA+ADP)

Adjust to pH8. $NaHCO_3$, L-malate, ADP, and 3-PGA are at 2x desired final concentration, as assay buffer will be mixed in equal volume with resuspension buffer (containing bundle sheath strands)

Component	**Final concentration**
Sorbitol	0.35 M
HEPES (pH 8)	25 mM
$MgCl_2$	2 mM
KH_2PO_4	1 mM
$NaHCO_3$	20 mM
L-malate (pH 8)	20 mM
ADP (sodium salt) (Sigma-Aldrich)	2 mM
3-PGA (di-sodium salt) (Sigma-Aldrich)	10 mM

Table A.7: Composition of bundle sheath strand assay buffer 2 (NADP)

Adjust to pH 8. $NaHCO_3$, L-malate, and NADP are at 2x desired final concentration, as assay buffer will be mixed in equal volume with resuspension buffer (containing bundle sheath strands)

Component	**Final concentration**
Sorbitol	0.35 M
HEPES (pH 8)	25 mM
$MgCl_2$	2 mM
KH_2PO_4	1 mM
$NaHCO_3$	20 mM
L-malate (pH 8)	20 mM
NADP (di-sodium salt) (Roche)	10 mM

A.2.3 *S. viridis* transformation and selection media

Table A.8: Composition of callus initiation medium (CIM) (pH 7.2)

For preparation of 1 L medium.

Component	Amount	Comments
MS salts + vitamins	4.3 g	Add before pH adjustment and before autoclaving
Maltose	40 g	Add before pH adjustment and before autoclaving
$ZnSO_4$-$7H_2O$	35 mg	Add before pH adjustment and before autoclaving
$CuSO_4$-$5H_2O$	0.6 mg	Add before pH adjustment and before autoclaving
Gelzan	4 g	Add after pH adjustment and before autoclaving. Not added during preparation of liquid CIM.
Kinetin	0.5 mg	Add after autoclaving
2,4-Dichorophenoxyacetic acid	2 mg	Add after autoclaving
Timentin	150 mg	Add after autoclaving to CIM selective only
Hygromycin	40 mg	Add after autoclaving to CIM selective only

Table A.9: Composition of plant regeneration medium (PRM) (pH 5.8)

For preparation of 1 L medium.

Component	Amount	Comments
MS salts + vitamins	4.3 g	Add before pH adjustment and before autoclaving
Sucrose	20 g	Add before pH adjustment and before autoclaving
Phytoblend	7 g	Add after pH adjustment and before autoclaving
Kinetin	2 mg	Add after autoclaving
Timentin	150 mg	Add after autoclaving to PRM selective only
Hygromycin	15 mg	Add after autoclaving to PRM selective only

Table A.10: Composition of rooting medium (RM) (pH 5.7)

For preparation of 1 L medium.

Component	**Amount**	**Comments**
MS salts + vitamins	2.15 g	Add before pH adjustment and before autoclaving
Sucrose	30 g	Add before pH adjustment and before autoclaving
Phytoblend	7 g	Add after pH adjustment and before autoclaving
Timentin	150 mg	Add after autoclaving to RM selective only
Hygromycin	20 mg	Add after autoclaving to RM selective only

A.3 Common protocols

A.3.1 Golden Gate assembly

Table A.11: Golden Gate reaction recipe

Volume	Component	Final component concentration
-	Backbone plasmid	1 fmol/μL
-	Part (plasmid)	2 fmol/μL
-	Part (PCR product)	3 fmol/μL
2 uL	T4 DNA ligase buffer (NEB)	10 x
1 uL	Restriction enzyme (BsaI or BbsI)	20,000 U/mL
1 uL	T4 DNA ligase (NEB)	400,000 U/mL
to 20 uL	ddH_2O	-

Table A.12: Golden Gate reaction thermocycler settings

Step #	Temperature	Time	Repeat
1	37°C	5 min	40 x
2	16°C	5 min	
3	80°C	20 min	1 x
6	16°C	∞	1 x

A.3.2 PCR using GoTaq polymerase (Promega)

Table A.13: GoTaq PCR reaction recipe

Component	**Final component concentration**
5x GoTaq reaction buffer (Promega)	1x
$MgCl_2$	2.5 mM
dNTPs	0.2 mM each
Forward primer	0.8 μM
Reverse primer	0.8 μM
GoTaq G2 DNA polymerase (Promega)	0.04 units/μl
Template	NA (1 μl)

Table A.14: GoTaq PCR reaction thermocycler settings

Step #	**Temperature**	**Time**	**Repeat**
1	95°C	5 min	1 x
2	95°C	20 sec	35 x
3	~ 2°C below melting temp	45 sec	
4	72°C	30 sec	
5	72°C	10 min	1x
6	15°C	∞	1x

A.3.3 Colony PCR using GoTaq polymerase (Promega)

Table A.15: GoTaq Colony PCR reaction recipe

Component	**Final component concentration**
5x GoTaq reaction buffer (Promega)	1x
$MgCl_2$	2.5 mM
dNTPs	0.2 mM each
Forward primer	0.8 μM
Reverse primer	0.8 μM
GoTaq G2 DNA polymerase (Promega)	0.04 units/μl
Template	NA (1 μl LB containing bacterial colony)

Table A.16: GoTaq Colony PCR reaction thermocycler settings

Step #	**Temperature**	**Time**	**Repeat**
1	98°C	10 min	1 x
2	98°C	60 sec	35 x
3	~ 2°C below melting temp	1 min	
4	72°C	2 min	
5	72°C	10 min	1x
6	15°C	∞	1x

A.3.4 PCR using Phusion polymerase (NEB)

Table A.17: Phusion PCR reaction recipe

Component	Final component concentration
5x Phusion GC Buffer	1x
dNTPs	0.2 mM each
Forward primer	0.5 μM
Reverse primer	0.5 μM
Phusion DNA polymerase (NEB)	0.02 units/μl
Template	NA (0.5 μl)

Table A.18: Phusion PCR reaction thermocycler settings

Step #	Temperature	Time	Repeat
1	98°C	30 sec	1 x
2	98°C	10 sec	35 x
3	Using NEB Tm calculator	30 sec	
4	72°C	30 sec	
5	72°C	10 min	1x
6	15°C	∞	1x

A.3.5 PCR using Q5 polymerase (NEB)

Table A.19: Q5 PCR reaction recipe

Component	Final component concentration
5x Q5 reaction buffer (NEB)	1x
5x Q5 high GC enhancer	1x
dNTPs	0.2 mM each
Forward primer	0.5 μM
Reverse primer	0.5 μM
Q5 DNA polymerase (NEB)	0.02 units/μl
Template	NA (0.5 μl)

Table A.20: Q5 PCR reaction thermocycler settings

Step #	Temperature	Time	Repeat
1	98°C	30 sec	1 x
2	98°C	20 sec	35 x
3	Using NEB Tm calculator	45 sec	
4	72°C	60 sec	
5	72°C	2 min	1x
6	15°C	∞	1x

A.3.6 Plant DNA extraction

Table A.21: Plant DNA extraction buffer 1 recipe

Component	Concentration
Tris pH 7.5	200 mM
NaCl	250 mM
EDTA	25 mM
SDS	0.5 % (w/v)

Table A.22: Plant DNA extraction buffer 2 (CTAB) recipe

Component	Concentration
CTAB (Cetrimonium bromide)	1.5% (w/v)
NaCl	1.05 M
EDTA (pH 8)	15 mM
Tris-HCl (pH 8)	75 mM

A.4 Golden Gate cloning of constructs for expression in *E. coli*

A.4.1 *S. italica HATCH1* (Seita.4G275500) sequence (5'-3') modified for Golden Gate cloning (SC module) for *E. coli* expression construct

cactctgtggtctcaaATGCGCGAGGTTTTATTCCGTCCAGTAGGGCTGCCGACGGAGACTAAGTTTGGTGCT
GGTTTAGAGGATCGCATCGAGAAAGTCATCTGCGCCTGCCGTTTCATGACGTTTTTGGGAATCGGAGG
GCTGTTGGCAGGGTGTGTACCTTGCTTCTTAAAGGGGTGCGTCTATGTCATGGACGCCTTTGTAGAATA
CTACCTGCACGGAGGCGGTATGTTGATCCTTATGCTTTTAGAGGCGATTGATATGTTCCTTATTGGAACC
GTTATGTTTGTCTTTGGGACTGGCCTGTACGAACTTTTCATCTCGGAAATGGATATGTCTTATGGGTCTA
ATCTTTTTGGACTGTTTTCGTTGCCAGAACGTCCTAAGTGGCTTGTGATCCAGAGTGTTAACGATCTTAA
AACGAAACTGGGGCACGTTATTGTGATGTCCCTTTTGGTAGGTATTTTTGAGAAAAGTTGGCGTGTCAC
GATTACCTCCTGTACGGATTTATTGTGCTTTGCGGCCTCCATTTTTTTGTCGAGCGGGTGTCTGTACCTGT
TATCACGTTTGAGTAACACAAAAGGTGGATCTCACACTTGAgctttgagaccacgaagtg

A.4.2 *A. thaliana HATCH1* (AT4G19390) sequence (5'-3') modified for Golden Gate cloning (SC module) for *E. coli* expression construct

cactctgtggtctcaaATGAGTACCAACCGTTTTGAAGCCTTAGAGGAAGGGATTGAAAAAGTTATTTATTCG
TGTCGTTTTATGACGTTCTTAGGTACACTGGGGTCCTTGTTAGGTAGCGTGCTGTGTTTCATCAAGGGCT
GTATGTATGTTGTAGATTCTTTTCTTCAATATTCTGTCAATCGCGGGAAGGTTATTTTCCTGTTGGTCGA
GGCCATTGATATTTATTTGTTGGGAACCGTTATGTTAGTGTTTGGACTGGGCCTGTAGAGCTGTTCATCT
CGAATCTGGATACTTCTGAGAGCCGCACCCACGACATCGTTTCTAATCGCTCATCCTTGTTTGGTATGTT
CACCTTGAAGGAGCGCCCCCAATGGCTTGAAGTAAAATCGGTGAGCGAGCTGAAAACGAAACTGGGT
CACGTAATTGTTATGTTGTTACTGATCGGGTTATTTGATAAGTCTAAACGTGTTGTTATCACCAGTGTTA
CGGACCTGTTATGCATTAGTGTAAGCATCTTCTTCAGCTCAGCATGTCTGTTCTTGTTAAGCCGTCTTAA
CGGCAGCCACTGAgctttgagaccacgaagtg

A.4.3 *S. italica HATCH3* (Seita.2G415600) sequence (5'-3') modified for Golden Gate cloning (SC module) for *E. coli* expression construct

cactctgtggtctcaaATGGAAACCACCGTCGAACGCGTTATTTTCGATTTTCGTTTCATGGCTCTTTTGGCCA
TTGCAGGTAGCCTGGCGGGGAGCGTACTGTGTTTCCTTAATGGATGCGTCTTTATTAAAGAGGCGTACC
AGGTATATTGGTCCTCCTGTGTAAAAGGCGTTCATACAGGCCAGATGGTGTTGAAAGTCGTGGAAGCT
ATCGATGTATATCTTGCAGGGACGGTCATGTTGATTTTTGGAATGGGGCTGTACGGCTTGTTTATCTCTA
ATGCCAGTACCGATGTACCAAGTGAGTCCGATCCCTTACGTGGATCCAGCTTATTCGGAATGTTTGCGA
TGAAAGAGCGTCCCAAGTGGATGAAGATTACTTCGCTGGATGAGTTGAAAACAAAAGTGGGCCATGTA
ATTGTAATGATCCTGTTAGTCAAGATGTTTGAACGTTCCAAAATGGTAAAGATCACCACGGGGTTGGAC
TTATTGAGTTACTCGGTATGTATTTTCCTTTCGAGTGCAAGCCTTTACATTCTTCATAACCTTCACCGTGG
AGATCATGAAGAAGGAGCGATGCCGAACCTGTGAgctttgagaccacgaagtg

A.4.4 *A. thaliana HATCH3* (AT5G13720) sequence (5'-3') modified for Golden Gate cloning (SC module) for *E. coli* expression construct

cactctgtggtctcaaATGGAAAGCAACGTGGAACGCATAATCTTCGACTTTCGGTTTCTGGCCCTGCTAGCT
GTCGGCGGCAGCCTGGCAGGTAGCCTGCTGTGCTTCCTGAATGGCTGCGTGTATATTGTGGAAGCGTA
CAAAGTATACTGGACCAACTGTAGCAAAGGCATCCATACCGGTCAGATGGTCCTGCGTTTGGTTGAGG
CGATTGATGTTTATTTAGCCGGGACGGTGATGCTGATTTTTTCAATGGGTCTTTATGGGTTGTTTATTTC
ACACTCACCGCATGATGTTCCGCCTGAATCAGATCGCGCATTACGTTCAAGTAGTTTATTCGGAATGTTT
GCCATGAAGGAGCGTCCAAAATGGATGAAGATCAGTTCCTTAGATGAGCTTAAAACGAAAGTAGGCCA
CGTGATCGTGATGATTCTGTTGGTAAAGATGTTCGAGCGCTCGAAGATGGTCACTATCGCGACGGGAC
TTGACCTGTTATCCTATAGTGTTTGTATTTTTTTATCCTCTGCTTCTCTGTACATTCTGCATAATCTTCACA
AAGGTGAGACATGAgctttgagaccacgaagtg

A.4.5 *Z. mays HATCH1* (candidate 10) (GRMZM2G179292) sequence (5'-3') modified for Golden Gate cloning (SC module) for *E. coli* expression construct

cactctgtggtctcaaATGGAAGCCCGCATTGAGAAAGTCATATACGCGTGCCGGTTTATGACCTTTTTTGGT
ATTTGTGGCCTGCTGCTGGGATCGGTTCCATGCTTCCTGAAAGGCTGTGTGTTCGTAATGGATGCATTT
GTGGAGTACTATCGTCATGGTGCAGGTAAAGTGATTCTGCTGCTGGTCGAGGCCATCGAAATGTTCTTG
ATCGCTACTGTCACATTTGTGTTGGGTACGGGCCTGTACGAACTTTTCATCAGCAACATGGATTCCTTTT
ATGGGAGTAACCTTTTTGGGCTTTTCTCCCTGCCGGAACGCCCTAAATGGGTAGAAATCAAATCCGTTA
ATGACTTGAAAACTAAACTTGGTCACGTGATCGTTATGGTTCTGTTAGTGGGAATCTTTGAAAAGTCGA
AGCGTGTCACTATCACGTCCTGCGCGGATTTACTTTGCTTTGCGGGCTCTATCTTCTTGAGCTCAGTATG
TCTGTATTTGCTTAGCAAGTTACATACAACTAAAGGAGGCAGTCAGGCTTGATGAgctttgagaccacgaagt
g

A.4.6 *Z. mays HATCH3.1* (candidate 1) (GRMZM2G327686) sequence (5'-3') modified for Golden Gate cloning (SC module) for *E. coli* expression construct

cactctgtggtctcaaATGACGAAAAGTACAGTCGTCGAAACGACGGTTGAGCGTGTTATTTTTGACTTCCGC
TTTTTAGCCCTGTTAGCTGTCGCTGGTTCCCTTGCAGGGTCCCTGCTTTGTTTTTTGAATGGGTGTGTCTT
TATCAAAGAGGCGTACCAAGTGTATTGGTCGTCATGCGTAAAAGGGGTACATACTGGCCAGATGGTCT
TGAAGGTAGTCGAGGCAATTGATGTTTATCTTGCCGGAACCGTAATGCTTATCTTCGGAATGGGTTTGT
ACGGGTTGTTTGTAAGTAACGCTAGTGCAGGGGTCGGTAGCGAATCGGATCGCGCGCTTAGCGGAAG
TTCTCTTTTCGGGATGTTTGCCCTTAAAGAACGCCCGAAGTGGATGAAAATCACCTCACTGGACGAGTT
AAAGACGATTGTTGGTCATGTGATCGTTATGATTCTTTTGGTGAAGATGTTTGAACGTAGTAAAATGGT
AACTATTGCGACCGGATTGGACTTACTTAGCTATTCGATTTGCATCTTTTTAAGCAGTGCAAGCCTGTAT
ATCCTGCACAACCTGCATAAGGGCGACGATCACGAGGAATAAgctttgagaccacgaagtg

A.4.7 *Z. mays HATCH3.2* (candidate 11) (GRMZM2G133400) sequence (5'-3') modified for Golden Gate cloning (SC module) for *E. coli* expression construct

cactctgtggtctcaaATGGAAACGACCGTAGAACGCGTCATTTTCGATTTTCGGTTCCTGGCCCTGCTGGCG
GTTGCTGGCAGCCTGGCGGGTTCTGTCCTGTGCTTTCTGAATGGTTGTGTGTTCATAAAAGAAGCCTAT
CAGGTTTACTGGAGCTCATGCGTGAAAGGCGTCCATACGGGTCAAATGGTGCTGAAGGTAGTAGAAGC
AATCGATGTTTACTTAGCGGGGACTGTGATGCTTATTTTTGGGATGGGCTTGTATGGCCTGTTCATCTCG
AACGCGCCAGCCTCGGTCGCGCCAGAATCCGACCGCGCCCTGAGCGGGAGTTCCCTGTTTGGGATGTT
CGCATTAAAGGAGCGTCCAAAGTGGATGAACATCACATCTCTTGACGAGCTTAAAACCAAGGTGGGCC
ACGTTATTGTTATGATCTTATTAGTGAAAATGTTTGAGAAATCGAAGATGGTGACTATCGCTACCGGAC
TGGATCTGCTTAGCTATTCAATCTGTATCTTTTTGAGTTCCGCATCGCTTTACATCCTTCACAATTTACATA
AAGGTGATCACGAAGAGTAAgctttgagaccacgaagtg

A.4.8 *O. sativa HATCH1* (LOC_Os06g49640) sequence (5'-3') modified for Golden Gate cloning (SC module) for *E. coli* expression construct

cactctgtggtctcaaATGGAAGCCCGGATTGAAAAAGTTATATATGCGTGCCGTTTCATGACCTTCCTGGCT
ATCGCCGGCAGCCTGATCGGCAGCGTTCCGTGTTTCCTGAAAGGCTGCGTGTACGTGATGGATGCGTT
CATTGAATACTATTTACATGGAGGTGGCAAAGTTACCCTGATGCTGGTTGAGGCAATCGACATGTTTCT
GGTGGGCACGGTAATGTTTGTGTTTGGTACGGGGTTGTACGAGTTATTTATTTCAAATATGGATATTGC
AAAGAGCTCTTCCTATGGTTCCAACTTGTTTGGACTTTTCCGCTTGCCTGAACGCCCCGAGTGGCTTGAA
ATCCAATCAGTCAACGATCTTAAAACCAAGTTGGGGCACGTCATTGTAATGGTCCTGCTGGTCGGTATC
TTTGAGAAGAGCAAGCGTGTGACGATTACTTCTTGCACAGACTTGTTCTGTTTTGCAGCCTCCATTTTTC
TTAGTAGTGCGTGTCTGTATCTTCTTTCGCGCCTTAGTTCTAAATGAgctttgagaccacgaagtg

A.4.9 *O. sativa HATCH2* (LOC_Os03g52910) sequence (5'-3') modified for Golden Gate cloning (SC module) for *E. coli* expression construct

cactctgtggtctcaaATGGAAGCGACGATCCATCAGGGCGTGGTGCGATGTCAGTCGTTTACCCTGATAGG
CGTAGCGGGGTCCCTGGTCGGCTCTGTTCCGTGCTTTCTGGAAGGTTGCGGTGCCGTAGTCCGCAGTTT
TTTTGTTCAGTTTCGGGCCCTGACCCAGACCATTGATCAGGCGGAGATCATTAAGCTGCTGATTGAGGC
GATTGATATGTTCTTGATTGGGACAGCATTGCTGACGTTCGGCATGGGAATGTACATTATGTTCTACGG
GAGCCGTTCAATCCAAAATCCGGGTATGCAAGGCGACAATTCACACCTTGGGAGCTTTAATTTGAAAAA
ATTAAAAGAAGGCGCACGCATCCAGAGCATTACTCAAGCTAAAACACGTATTGGACATGCCATCTTATT
ACTTTTACAGGCAGGAGTGCTTGAAAAATTCAAGTCCGTGCCATTGGTCACGGGGATCGATATGGCTTG
TTTCGCAGGGGCGGTGCTGGCAAGCAGTGCTGGCGTTTTCTTATTATCGAAATTAAGCACTACAGCGGC
CCAAGCCCAGCGCCAACCTCGTAAACGTACCGCATTTGCTTGAgctttgagaccacgaagtg

A.4.10 *O. sativa HATCH3.1* (LOC_Os07g46320) sequence (5'-3') modified for Golden Gate cloning (SC module) for *E. coli* expression construct

cactctgtggtctcaaATGGAGTCCACTGTGGAAAAACTGATTTTTGATTTCCGCTTTCTGGCGCTACTCGCG
GTGGCCGGGTCGCTGCTGTGTTTTCTCAGCGGTTGCGTCTATATAAAAGAAGCGTATCACGTATATTGG
ACCAGCTGCGTGCGCGGCGTACATACCGGCCAGATGGTGCTGCGGCTGGTGGAAGCGATCGACGTTTA
TTTAGCAGGTACGGTTATGCTTATTTTCGGCATGGGTCTGTACGGGTTATTCATCTCAAACGACTTTCCT
GACGTCCCCGCTTCTGATCGCGCACTTAAGGGCTCGTCACTGTTTGGAATGTTCGCCTTGAAGGAGCGT
CCACGTTGGATGCGTATTAGCTCACTGGATGAACTTAAAACGAAGTTAGGACATATCATTGTTATGATT
TTACTGGTCAAGATGTTTGAACGTTCTAAGATGGTTCATATCACAACAGGTATGGACCTGCTTAGCTAC
GCCGTGTGTATCTTCCTGTCCTCTGCATCTTTATACATTCTTCATAACCTGCACCGTCCAGAGCACGAATG
Agctttgagaccacgaagtg

A.4.11 *O. sativa HATCH3.2* (LOC_Os07g46330) sequence (5'-3') modified for Golden Gate cloning (SC module) for *E. coli* expression construct

cactctgtggtctcaaATGGAAACCACCGTGGAGCGGGTCATTTTTGATTTCCGATTTCTTGCTCTGCTGGCTG
TGGCAGGCAGTCTGGCCGGCAGCCTGCTGTGCTTCCTAAACGGTTGTGTGTACATTAAAGAAGCCTACA
GCGTTTATTGGAGTGGCTGTCTGAAAGGAGTGCACACCGGCCAGATGGTGCTGAAAGTGGTAGAGGC
CATCGACGTCTACTTAGCGGGGACAGTTATGTTGATCTTCGGGATGGGGTTGTATGGGCTGTTTATTTC
CAATACTTCGACAGACGTTCCGTCTGAATCGGATCGTGCGCTTCAGGGTTCATCCTTGTTCGGCATGTTT
GCACTTAAAGAACGTCCCAAGTGGATGAAAATCACTAGCCTGGATGAATTGAAAACGAAGGTTGGTCA
TGTGATTGTTATGATCCTGTTGGTAAAGATGTTTGAGCGCTCAAAAATGGTAAAAATCACTACGGGTCT
GGATCTGCTGTCTTATTCAGTCTGCATCTTTTTGTCGAGCGCGTCCTTATATATCTTACATAACTTGCATC
GCCCCGAACAAGACGACTCTGTCATGCCAAATTTATGAgctttgagaccacgaagtg

A.4.12 *S. bicolor HATCH2* (Sobic.001G096800) sequence (5'-3') modified for Golden Gate cloning (SC module) for *E. coli* expression construct

cactctgtggtctcaaATGGAACGTATGATTCATCAGGTGATAATCAACTGCCAGTCTTTTACGCTGATTGGC
GTCGCCGGCTCTCTGGTGGGCAGCGTGCCACTATTTGCAGAGGGCTGTGCGGTGGTCATGAAAAGTTT
TTTTATGCGGTTTCACGCCATGAGCCACACCGTTGATCAGGGCGAAACCATTCGTCTGCTGATTGAAGC
CTTAGATATGTTCTTAATTGGAACGGCTTTATTGACCTTTGGGATGGGTATGTACACTATGTTTTACGGT
AGCCAAAGTATTCAGCAACAAGCTCGCCATGTAGACACATCCCATCTTGGGGCGTTCAACTTAAAGAAG
TTAAAGGAAGGTGCTCGTATCCGCTCGGTCACTCAAGCAAAAACACGCATCAGCCACGCAATCTTACTT
CTTTTGCAGGTAGGGGTATTGGAGAAATTCAAGTCGGTTCCGCTGCTTACCGGTCTTGACATGGCCTGC
TTCGGCGGGGCTGTACTGGCTTCCTCCGCTTCCGTTTTCCTTCTTTCCAAGTTGTACATGGGTCAACAGG
GACTGAAGCAGACGGTTACCTGAgctttgagaccacgaagtg

A.4.13 *Z. mays* candidate 1 (GRMZM2G138258) sequence (5'-3') modified for Golden Gate cloning (SC module) for *E. coli* expression construct

cactctgtggtctcaaATGGCGTCTGCAGACGCTGCAGCAAGCGATGTGGGGGGTGGTGGAGCGGGTTTTG
GTGGCCACGATGACGATGGACATAACCACCATGGCGGTGAGGGCGGAGGAGATGACGGAGGGCATG
GAGATGATGCAGGTCACGGAGATGACGCGCCCGGAGGTGGAGATGCGCGTGGTGAAGCCTTATTCGT
GCTGGCTCAACTTGGGCGTAAGTTAGACTCGCTGCCAAGCGACCTTGCTGCGGCTGTGGATTCAGGCC
GTATTGGAGCGGACATTGTTCGTCGCTTCACTGAGCTGGAGGCTAATGGGTTCTTCCGTTGGTTATTAC
AATTTCAGGGATTCCGCGAGCGCTTGCTTGCTGATGAACTGTTCCTGACTAAGCTGGGGATTGAGTGTG
GGATTGGGTTAGTAGCCAAAACTGTGGCGGAGTTCCAAAAACGTGGGGATAATTTTTTTAAGGAAATT
GAGGTCGTGATTTCGGACGTAGTCATGGCGATCGTTGCCGACGTAATGCTTGTTTATCTTCCGGCACCA
ACCATCGGGCTGCAACCCCCGCTTGCTCGTAATGCCGGGGCTATCGCGAATTTTTTCTACAATTGTCCCG
ATAACACTTTTCAGATCGCAATGGCCGGCCGTAGCTTCTCGTTGCTGCAACGTATTGGCGCTTTCGTGC
GTAATGGGATTAAATTATTAGCAGTGGGTACCACAGCGAGCCTTATCGGTACGAGCGTCACGAATGCT
GCCCTTAAAGCGAAACGTGCCGTAGATAAGGACTTGGAGGATGAAGTTATGGAGATTCCCGTCGTATC
AACTTCCGTCGCGTACGGAGTCTACATGAGCATCAGTAGTAATCTGCGCTACCAATTACTTGCGGGCGT
TATTGAACAGCGCATGTTGGAACCGCTGCTTCATAATCAGAAATTGTTACTGTCGGCAATGTGTTTCATC
GTCCGCACTGGGAACACCTTTTTGGGTTCATTGTTATGGGTAGACTACGCCCGTTGGATCGGAGTTCAG
AAGTCTCATGAAGAAGCATAAgctttgagaccacgaagtg

A.4.14 *Z. mays* candidate 2 (GRMZM2G099367) sequence (5'-3') modified for Golden Gate cloning (SC module) for *E. coli* expression construct

cactctgtggtctcaaATGTCCCTTGCCAAGGTCAGTTTTGGAGTGATTGGCTTAGGATTGGGTCTTTCCCTTT
TGAGTTACGGGTTCGGTAGCTACTTTAATCTTTTGCCTGGCTCCGAATGGAGCGCTCTGCTGCTGACCTA
CGGATTTCCTCTGACTATCATTGGTATGGCGCTTAAGTACGCTGAACTTAAGCCCGTCCCTTGTATCACA
TACGCCGACGCGTTTGCTTTGCGTGAGAAATGCGCCACCCCGATCCTGAAGCAGGTCCGTTCGGATGTT
ACCCGCTTCCGCTATGGCGACGAGCAACATTTAGATGAGGCGTTACAGCGCATTTTTCAGTATGGGTTG
GGCGGGGGGATCCCCCGTCGCAATGCACCAATCTTACAAAAAATCCGCGAGGAAGTCACGGATGACG
GAAAATACTCGCTGGTGTTGGTATTTGAAGCCAAGGCTTTAGAGTTAAGTGACTTTGAAAAGCGTCAA
GCAAAGTTCACGTCTTTCTTTGGACCAGGTATTAAAGCAGAAATCGACGCGGGCAAGGGAGGAGGCG
ACTTGTACGAGGTTCGCTTGATCTCGGACACTGTGTAAgctttgagaccacgaagtg

A.4.15 *Z. mays* candidate 3 (GRMZM2G088138) sequence (5'-3') modified for Golden Gate cloning (SC module) for *E. coli* expression construct

cactctgtggtctcaaATGACCGTAGTGCCAGACCCCGATTATCGTTTACCTATCGCTATTCTTGGAATTGCGT
CAGCATTCGCTTATGCAGATAACCTTTTAGCAGCGGCACCGGTTGGACTTCTTGGGTTATTGTTGCTGTT
TCAAACGACGCGCGTTCGTTTCGTATTCGATAATGAAGCACTGGAGGTTAAAGTCGGCGAGCAATTAC
AAGAATCAGGTGAAAACGTATTCGTCGGGGGTAAGAACCGTTGGAAGTACAGCACATTCGTTAACTGG
GAGCTTTGGTGGCCTCAATTTCCAGTTTTAGTTTACTTCAAGGAAACGCAGACAAAGCCCGAGGGGCA
GATCCACTTTTTTCCCGTAATCTTCGACGGTCGTCAGTTGTATGACGTCATGGTCGAACGCGCGGGGCC
TAGTAAGACTAGTGGTCCGAAATAAgctttgagaccacgaagtg

A.4.16 *Z. mays* candidate 5 (GRMZM2G145500) sequence (5'-3') modified for Golden Gate cloning (SC module) for *E. coli* expression construct

cactctgtggtctcaaATGTCGTCGTCGCCATCTTCGGGCCCCTTATACCCGACCCCTCCACCTACTGAACAAG
AAGTAGAGCGTGTTAAACTTGAACAAGTAATGCGCCGTCTTGAGAAAACAGCTCGTTACTTCAAAAATC
TGGGTACATTAGGCTTCTGGAGCCAACTTGTATGTACCATTATCTCCGCGGGGATCTTGGCGTTTTCGA
CCGTTGTTACCGGGAAAGTGACGTCCCCCTTTACGTTTTATGCAACTGCCGCTGGGATTGCGGCAGCGT
TCATCTCGGTGTTTTGGTCCTTCGGTTATATTCGTCTTTCTGAGCGTTTGCGTAAGACAAGTAAGGAGCC
GGCAAAGGCGCCTCCCCGCGCGGATGTTGTGAAGTCGTTGAAAAATGGTATCGTGGTAAACATTTTAG
GAATGGGAGCTGCGGTCTTGGGAATGCAAGCGACCGTTGGAGCTTTGGTCGCGAAAGCGTTGACAAC
ATCAGCCGTTCCCTATTATCAGGCTGCCGGGCAGTCCCCGGTCCTGGCGTTAGATGTGTTCCTTGTGCA
AGCTTCCGCAAACACTATCTTATCGCACTTCCTGGGCGTCGCTTCTTCATTGGAACTGCTTCGCTCAGTA
AGTCTGCCACCCGCAGAGGCTGCGGTGCCTGCACCTGCTCCAGCACGTGCGTAAgctttgagaccacgaagtg

A.4.17 *Z. mays* candidate 6 (GRMZM2G004352) sequence (5'-3') modified for Golden Gate cloning (SC module) for *E. coli* expression construct

cactctgtggtctcaaATGGCCGCCATTGACAAGAAGTCCTCAGACGATGTTCGCCAGGCTTTGGACCAATTA
CGCGAATCGGGGTGGGCTAAGCGCTGGGGGTCGCAGCCATACGTGTCACGTCGCACAACCTCATTACG
CGAATTAACAACACTGGGAATTAAGAACGCCGAGAACCTTGCAATTCCTTCGGTACGCAACGACGCGG
CCTTTCTTTTTACTGTTGTTGGAACAACCGGATTCCTGGCAGTTCTTGCTGGGCAACTTCCGGGGGACTG
GGGCTTCTTTGTACCATATCTGATCGGTAGCATCTCACTTATTGTACTGGCCGTCGGTTCTGTCGCCCCT
GGGCTTTTACAAGCGGCAATCGGGGCCTTCTCTACAGTCTTTCCTGACTACCAAGAGCGCATTGCACGC
CATGAGGCGGCTCACTTTTTAGTGGCCTACTTAATTGGCTTACCTATCTTAGGTTATAGTCTTGATATTG
GTAAGGAACACGTAAATCTGATTGATGAACAATTACAGAAACTTATTTACTCCGGACAGCTGGATGGG
AAAGAATTAGACCGCCTTGCTGTGGTTTCCATGGCAGGGTTAGCAGCAGAAGGTCTTGAGTATGATAA
AGTGGTAGGACAAAGTGCAGATCTTTTCACCCTTCAACGCTTCATCAATCGCACTAAACCTCAGTTAAGT
AAGGACCAACAGCAGAACTTAACGCGTTGGGCCGTGCTGTTCGCTGCCTCACTGCTGAAGAATAATAA
GGCCGCGCATGAGGCATTAATGTCCGCAATGTCGCAAAAAGCATCAGTATTGGGTTGTATTGAAGCGA
TTGAGAACGCCTCGTGAgctttgagaccacgaagtg

A.4.18 *Z. mays* candidate 7 (GRMZM2G074053) sequence (5'-3') modified for Golden Gate cloning (SC module) for *E. coli* expression construct

cactctgtggtctcaaATGTCATCACAACCTAAAACATTCGCGAATGTATTCATCGCCGTTGTGGGTGCTGGA
GTACTGGGTCTGCCATACACCTTCTCACGCACAGGCTGGGCTGCTGGGAGCTTATTGTTATTTAGCGTC
GCCGCTCTTACTTTTTACTGTATGATGCTTCTGGTGGCCTGTCGTCGTCGTTTAGCAGACGAACATCCAA
AAATCGCTAGTTTTGGAGACTTGGGTGACGCAGTGTTTGGGGCCCACGGGCGTTTTGCAGTAGATGTA
ATGCTGGTCTTGTCTCAGTTCAGTTTTTGTGTGGGATACCTGATTTTCATTTCAAACACAATGGCTCACTT
GTATCCTATTACAGCGGCTGCTTCTTCTAGCTCAGCATTGCTGAGCCCTAAAGCCCTGGTGATCTGGGCT
ATGTTGCCCTTCCAGTTGGGTTTGAATAGTATTAAGACGTTGACATTATTAGCGCCTCTTTCAATCTTCG
CGGATGTTGTAGATCTTGGAGCCATGGGGGTCGTCTTAGGCCAAGACGTGGCCGCATGGCTTGCAAAG
CCAGTACCAGTAGTAGCGTTCGGTGGTGCGGGGGCATTGTTGTACGGCTTGGGCGTTTCAGTCTACGC
GTTCGAAGGAATTGGTATGGTACTTCCCTTGGAGGCAGAGGCTGCGAATAAATCCAAATTTGGAGTCA
CTTTAGGCCTGTCAATGGCATTCATTGCTGTAATGTATGGATTATTCGGTGTGATGGGCTATGTGGCTTT
TGGGGATGCAACCCGCGACATTATTACTACAAATCTGGGCGCGGGCTGGCTGAGCGCCGCTGTCCAGC
TTGGATTGTGCATCAACCTTTTTTTCACGATGCCGGTTATGATGAACCCTGTCTACGAAGTTGCGGAGC
GCTTGCTGCACGGCAAGCGTTACTGTTGGTGGTTACGTTGGCTTCTTGTCGTCGTGGTTGGTCTTGCGG
CTATGTACGTACCAAATTTTACCGACTTCTTGGCCTTGGTTGGTTCTAGTGTCTGTGTTCTTTTGGGGTTT
GTCTTACCAGCGTCCTTTCACTTGAAAGTCTTTGGCGCCGAAATGGAGTGGCCGGGAGTCTTATCGGAT
GTTCTGCTTGTCGTCATCGGTTTAGCGCTGGCAGTGTTTGGCACCTATACATCCCTGCTGCAAATTTTTC
AGTCATCAAGTGCATGAgctttgagaccacgaagtg

A.4.19 *Z. mays* candidate 8 (GRMZM2G015033) sequence (5'-3') modified for Golden Gate cloning (SC module) for *E. coli* expression construct

cactctgtggtctcaaATGGTTAGTGCTGCAGAGTCCGAACCTCCCGATGCGGGGGCTGTGGAGGCGGAGG
GTCAGGGACCCGTAGAGTTGCGCGCACCGACTTTGTTTAGCACCGATGATAATCCAACTCCTCTGCAGA
CGGCTACGTCGTTGCTTCTGACTGGGGCTATCTCCGTTTTCCTGTTCCGTGCCTTACGCCGTCGTGCTCG
TCGTGCGAAGGAGTTGCGCGTTCGCTCGTCGGGTCTGAAAAAACCCAACAACTTAACAGAGGAAGCGT
TGGAACGTCTGCGTTTAATGAGCGCCTCCCCCATCGAAACTGAAAAAGCGACTTCACCAATCCAGGCCC
TGTTAGGTGGTATTGCAGCCGGAGTGATCGCTTTAATTCTTTACAAATTTTCAACCACAGTTGAAGCTGC
GTTAAATCGCCAGACGATTTCAGACTCATTTTCCGTCCGCCAAATTACGATTACGATCCGTACTATTGTG
ACGGGTCTGTGCTATTTAGCCACCTCCGTTTTTGGCATTAATGCGGTTGGGCTGGTGCTGTATAGCCTG
CAACTGACACTTCAAAGCATTATGGACGATGACTCTAGCTCGTCCACCGGCAAAATTTCCGAACAGAGT
AACACTCTTGCTTCCTCTGATAGCTCTACGTCGTCGCGCGAATCAGACTCCTCAGACATGCAGGTCACAG
ATAAATCCAAGAACAGCGCGGAAAAGCAAAGTTGAgctttgagaccacgaagtg

A.4.20 *Z. mays* candidate 9 (GRMZM2G078118) sequence (5'-3') modified for Golden Gate cloning (SC module) for *E. coli* expression construct

cactctgtggtctcaaATGGTGTCTAGCGATTCGCCGAAACCCGTAGCGAGCACCTCCTCCCCGACGGGCGG
GGATTCACCGGACGAAGAGCCTCCTGTACTTCCGCTGTTACAGGAGTTAGCGGACTGCTTAATTTTACC
CCCGAAGCTTCTTGCGCAGTTACCTTCAGATCTTCGCTTGGATTTGAACGATGCTGCTTTTGATTTGTCTA
ATGGGCCAGTCCTTGATGAATGTGGACAAGAAGTGGGGGACCTTTTACTGAACTTGGCTAAGGCATGG
GAGCTTGCCGACACAAGTACCAGCAATTCAATTGCGAAGCAACTTCCACTGATGGAGCCGTACCTGAC
GCGTAGTGCAAAGTCCGCTTTCGGGAAGCGCCTGGTCAGTGCCGGACGCCGCTTCCAATCCATGGGAC
AATATAATCAGGGGGAATTGAAAATTGCTGAGACTATGATTAAAAATGGCAAATTGTTGTCGAAGCAC
CCGGTTGTGCAGTCTGATGTTCAGGCCATGAAAGAAAAACGTCGTCTGAAATTTGGAGAGTTAGAGTT
TGGCGTCACAGCCCAGTCAGCCACCATTGGATCAGCAGTGGCAGCGATCTTCGGCCTTATCTCTTGGCA
GCTTGCGCGTGGCGTGCAGAGTATTCCTGATAACACCACACAATATACAAATGACAATGCTTTACAACT
GGCAAAATCATTAAAGACTGCACTGTTAGTATTAGGCTATACATCGACCGGACTTTCGCTGTTTGCTGC
ATTTGGGCTTCTGCTTTTGGCTCAGCAAATCAATTCCGAGAATAAGTCCGAGTGAgctttgagaccacgaagt
g

A.4.21 *Z. mays DCT2* (GRMZM2G086258) sequence (5'-3') modified for Golden Gate cloning (SC module) for *E. coli* expression construct

cactctgtggtctcaaATGTCTGGCGGGGCGAAGCCCTTGCCTTTATTTTTATCTCTTGCAGCTGGGCTGGCC
GTTCGCTTCTTGGTGCCCCGCCCGGCTGAAGTAACCCCCGAGGCATGGCAGTTACTGTCTATTTTTCTGT
CCACTATCGCAGGATTGGTTTTGGGCCCGTTGCCCGTAGGTGCGTGGGCGTTCTTGGGGTTGACGGCG
ACTGTTGCTACCCGCACGCTTCCTTTTACAGCGGCTTTTGGAGCGTTCACTAATGAAGTCATTTGGCTGA
TCGTTATTTCTTTCTTCTTCGCACGTGGATTTGTGAAAACTGGTCTTGGCGATCGTGTCGCTACATATTTT
GTGAAGTGGCTTGGCCGCTCCACCCTGGGTTTATCTTATGGGCTTGCAATTTCGGAAGCCTTGATCTCCC
CGGCCATGCCTAGCACGACAGCGCGTGCGGGGGGTGTGTTTCTTCCTATCGTCAAATCCCTTTCGTTGT
CATCGGGATCCAAACCAAACGATCCATCTGCTAAAAAGTTAGGATCTTACTTAGTGCAGTCTCAGTTAC
AAGCGGCCGCTAATTCTAGCGCACTGTTCTTGACGGCGGCCGCCCAGAATTTACTTTGCCTTAAATTAG
CGGAAGAGATCGGAGTTTCTATCGGGAATCCGTGGATCACATGGTTTAAAGTTAGCTCCGTTCCGGCTG
TCCTGGGTCTTTTGGTGACTCCGTATTTGATCTATAAGATTTTCCCACCGGAAATTAAGGACACCCCAGA
GGCACCCGCTTTGGCGGCTGAAAAACTGAAAAACATGGGTCCTGTGACTAAAAATGAGTGGGTAATGA
TTGGTACTATGCTTTTGGCGGTTTCATTGTGGATTTTTGGGGAAACCATTGGAGTCTCCAGCGTCGTAG
CTGCGATGATTGGTTTAAGTATTTTGCTGGTATTGGGTGTGCTGAACTGGGAGGACTGTTTGAACGAG
AAGTCAGCCTGGGACACATTGGCTTGGTTTGCAATCTTGGTAGGTTTGGCGGGGCAGCTTACGAATCT
GGGGATCGTAAGTTGGATGTCGAATTGCGTAGCCAAGGTACTGCAGTCTTTCAGCCTGAGCTGGCCCG
CAGCATTTGTTGTGTTGCAGGCATCATATTTTTTTATCCATTATATTTTCGCGTCACAAACGGCTCATGTG
GGAGCACTTTATAGTGCATTTTTGGCTATGCATTTAGCAGCAGGAGTCCCAGCATTAATGTCAGCCTTA
GCTTTGGCCTACAACGCAAATTTGTTCGGCGCTCTTACTCACTATAGCAGTGGCCAAAGTGCTGTGTACT
TCGGCGCCGGGTATGTAGATTTGCCAGATGTTTTTAAACTGGGCTTTATCACGGCCGCAATTAATGCAG
TGATCTGGGGAGTAGCAGGAGCTTTGTGGTGGAAGTTTCTTGGGTTATACTAAgctttgagaccacgaagtg

A.4.22 *E. coli CstA* sequence (5'-3') modified for Golden Gate cloning (SC module) for *E. coli* expression construct

cactctgtggtctcaaATGAACAAATCAGGGAAATACCTCGTCTGGACAGTGCTCTCTGTAATGGGAGCATTT
GCTCTGGGATACATTGCTTTAAATCGTGGGGAACAGATCAACGCGCTGTGGATTGTGGTGGCGTCGGT
CTGTATCTATCTGATCGCTTACCGTTTTTATGGGCTGTATATCGCCAAAAATGTGCTGGCGGTTGACCCG
ACGCGTATGACGCCAGCGGTGCGCCATAACGACGGGCTGGACTATGTGCCGACGGACAAGAAAGTGC
TGTTCGGTCACCATTTTGCGGCCATTGCCGGAGCAGGTCCGCTGGTGGGGCCGGTACTGGCGGCGCAA
ATGGGCTACCTGCCGGGGATGATCTGGCTGCTCGCTGGGGTGGTTCTCGCCGGTGCGGTGCAGGATTT
CATGGTGCTGTTTGTTTCTACGCGCCGTGACGGTCGCTCGCTGGGTGAGCTGGTCAAAGAAGAGATGG
GGCCAACCGCCGGGGTGATTGCGCTGGTGGCCTGCTTTATGATCATGGTCATTATCCTTGCAGTGCTGG
CGATGATCGTGGTGAAAGCCCTGACTCATAGCCCGTGGGGAACATATACCGTTGCGTTCACCATTCCGC
TGGCGCTGTTTATGGGGATCTACCTGCGCTATCTGCGTCCGGGGCGTATTGGCGAAGTGTCGGTCATC
GGTCTGGTATTCCTGATTTTCGCCATTATCTCTGGCGGCTGGGTGGCAGAAAGTCCGACCTGGGCACCG
TACTTTGACTTTACCGGCGTGCAGTTGACCTGGATGCTGGTGGGTTACGGTTTTGTGGCGGCGGTGCTG
CCGGTGTGGTTACTGCTGGCCCCGCGTGACTACCTCTCTACCTTCCTGAAAATCGGGACTATCGTCGGT
CTGGCGGTAGGCATTTTGATTATGCGCCCGACGCTGACCATGCCTGCGCTGACCAAATTTGTCGATGGC
ACTGGCCCGGTATGGACCGGTAACCTGTTCCCGTTCCTGTTTATCACCATCGCCTGTGGCGCGGTGTCT
GGCTTCCATGCGCTGATCTCTTCTGGGACCACGCCGAAGATGCTGGCGAACGAAGGGCAGGCGTGCTT
TATCGGCTACGGTGGGATGTTAATGGAATCCTTCGTGGCGATTATGGCGCTGGTTTCCGCCTGTATCAT
CGATCCGGGCGTGTATTTTGCCATGAACAGCCCGATGGCGGTGCTGGCTCCGGCAGGGACGGCGGAT
GTGGTCGCTTCTGCCGCGCAGGTGGTGAGTAGCTGGGGCTTTAGCATTACACCAGATACGCTAAACCA
GATTGCCAGCGAAGTGGGTGAACAGTCGATCATTTCCCGTGCGGGCGGTGCGCCGACGCTGGCGGTG
GGGATGGCCTACATTCTGCACGGCGCGCTGGGCGGCATGATGGATGTGGCGTTCTGGTATCACTTCGC
CATTTTGTTTGAAGCACTGTTTATTCTGACGGCGGTGGATGCAGGTACGCGTGCTGCGCGCTTTATGTT
GCAGGATCTGCTGGGCGTGGTGTCTCCTGGCCTGAAGCGGACCGATTCACTGCCTGCTAACCTGCTGG
CAACAGCGCTGTGCGTACTGGCGTGGGGCTACTTCCTCCATCAGGGCGTGGTCGATCCGCTGGGCGGC
ATTAACACTCTTTGGCCGCTGTTTGGTATTGCCAACCAGATGCTGGCAGGGATGGCGCTGATGCTCTGT
GCCGTGGTGTTGTTCAAGATGAAACGTCAACGTTACGCCTGGGTGGCGCTGGTACCAACGGCCTGGCT
GCTGATTTGTACCCTGACCGCAGGCTGGCAGAAAGCGTTTAGCCCGGATGCGAAAGTCGGCTTCCTGG
CCATTGCTAATAAGTTCCAGGCAATGATCGACAGCGGCAATATTCCGTCGCAGTATACTGAGTCACAGC
TGGCGCAACTGGTGTTCAACAACCGTCTGGATGCCGGGTTAACCATCTTCTTTATGGTGGTCGTGGTGG
TTCTGGCACTGTTCTCGATTAAGACGGCACTTGCGGCATTGAAAGATCCGAAGCCAACGGCGAAAGAA
ACGCCGTATGAACCAATGCCGGAAAATGTCGAGGAGATCGTGGCGCAGGCAAAAGGCGCACACTAAg
ctttgagaccacgaagtg

A.4.23 *E. coli BtsT* sequence (5'-3') modified for Golden Gate cloning (SC module) for *E. coli* expression construct

cactctgtggtctcaaATGGATACTAAAAAGATATTCAAGCACATACCCTGGGTGATTCTCGGAATCATCGGT
GCATTCTGCCTCGCGGTAGTTGCATTACGTCGGGGGGAGCACATCAGCGCCCTGTGGATCGTGGTCGC
CTCTGTATCGGTGTATCTGGTGGCGTATCGCTACTACAGTCTGTACATCGCCCAGAAGGTGATGAAACT
CGACCCCACGCGCGCGACGCCTGCGGTTATTAACAACGACGGTCTGAACTACGTTCCGACCAACCGTTA
CGTGTTGTTTGGTCACCACTTCGCCGCTATCGCCGGTGCTGGTCCGCTGGTGGGTCCGGTTCTCGCCGC
GCAGATGGGCTACCTGCCTGGCACGCTGTGGCTGCTGGCGGGGGTCGTGCTGGCCGGTGCGGTTCAG
GACTTTATGGTGCTGTTTATCTCCTCTCGCCGTAATGGCGCATCTCTTGGTGAGATGATCAAAGAAGAG
ATGGGACCAGTACCGGGGACTATCGCGCTGTTTGGCTGTTTCTTAATCATGATCATCATCCTCGCCGTCC
TGGCGCTGATTGTGGTTAAAGCCCTGGCCGAAAGTCCGTGGGGTGTCTTCACCGTTTGCTCAACCGTAC
CGATTGCGCTGTTTATGGGTATCTACATGCGCTTTATTCGTCCGGGGCGTGTGGGTGAAGTCTCTGTCA
TTGGTATCGTGCTGCTGGTTGCCTCTATCTACTTCGGTGGCGTGATTGCTCACGATCCGTACTGGGGTCC
GGCACTGACCTTTAAAGACACCACCATTACCTTCGCGCTGATTGGCTATGCGTTTGTTTCCGCACTGCTG
CCAGTGTGGCTGATCCTCGCACCGCGCGACTATCTGGCAACCTTCCTGAAAATCGGCGTTATCGTCGGC
CTGGCGCTGGGTATCGTGGTGCTGAACCCGGAACTGAAAATGCCTGCCATGACCCAGTACATTGACGG
TACTGGCCCGCTGTGGAAAGGCGCTCTGTTCCCGTTCCTGTTCATCACCATCGCCTGTGGTGCGGTATCT
GGCTTCCACGCGCTGATCTCTTCCGGTACGACGCCAAAACTGCTGGCTAATGAAACCGACGCGCGTTTC
ATCGGCTACGGCGCAATGCTGATGGAGTCCTTCGTGGCGATTATGGCGCTGGTTGCTGCGTCCATCATC
GAACCGGGTCTTTACTTCGCGATGAACACCCCGCCTGCTGGCCTTGGCATCACCATGCCTAACCTGCAT
GAAATGGGTGGCGAGAACGCGCCGATCATCATGGCGCAGCTGAAAGACGTTACCGCACACGCGGCAG
CGACCGTCAGCTCCTGGGGCTTCGTGATTTCGCCAGAGCAGATCCTGCAAACCGCGAAAGACATTGGT
GAGCCTTCTGTCCTGAACCGTGCAGGTGGCGCGCCGACGCTGGCGGTAGGTATCGCGCACGTGTTCCA
CAAAGTGCTGCCGATGGCTGACATGGGCTTCTGGTATCACTTCGGTATTCTGTTCGAAGCCCTGTTCATC
CTGACCGCGCTGGATGCGGGTACCCGTTCTGGCCGCTTTATGCTGCAAGACCTGCTGGGTAACTTCATC
CCGTTCCTGAAAAAAACCGATTCTCTGGTTGCCGGTATCATCGGTACTGCGGGCTGTGTGGGTCTGTGG
GGCTACCTGCTGTATCAGGGCGTGGTTGATCCGCTGGGCGGCGTTAAGAGCCTGTGGCCGCTGTTCGG
TATCTCCAACCAGATGCTGGCAGCCGTAGCGCTGGTACTGGGCACCGTTGTGCTGATTAAGATGAAGC
GCACCCAATACATCTGGGTAACTGTTGTTCCGGCTGTATGGCTGCTTATCTGCACCACCTGGGCGCTGG
GCCTGAAACTGTTCAGCACCAACCCGCAGATGGAAGGCTTCTTCTACATGGCAAGCCAGTACAAAGAG
AAGATTGCTAACGGTACTGACCTGACGGCGCAGCAGATTGCCAATATGAACCACATCGTTGTGAACAA
CTACACCAACGCAGGTCTGAGTATTCTGTTCCTGATTGTGGTGTACAGCATCATCTTCTACGGTTTCAAA
ACCTGGCTTGCGGTGCGTAACAGCGACAAACGTACTGACAAAGAAACACCGTACGTTCCAATCCCGGA
AGGCGGCGTGAAGATCTCTTCGCACCACtaagctttgagaccacgaagtg

A.4.24 *E. coli YhjX* sequence (5'-3') modified for Golden Gate cloning (SC module) for *E. coli* expression construct

cactctgtggtctcaaATGACACCTTCAAATTATCAGCGTACCCGCTGGCTGACACTCATCGGTACTATCATTA
CCCAGTTTGCGCTGGGGTCGGTTTATACCTGGAGCCTGTTTAATGGCGCGCTTTCCGCCAAGCTGGATG
CGCCGGTAAGCCAGGTCGCTTTCTCTTTCGGCTTGTTAAGTCTGGGGCTGGCAATTTCGTCTTCTGTTGC
GGGCAAATTACAGGAACGTTTTGGCGTTAAACGCGTCACCATGGCTTCCGGCATTTTGCTGGGATTAGG
CTTCTTCCTGACAGCGCATTCTGACAACCTGATGATGCTGTGGTTAAGCGCCGGTGTGCTGGTGGGACT
GGCAGATGGCGCGGGTTATCTGCTGACGCTCTCTAACTGTGTGAAGTGGTTCCCGGAGCGTAAAGGTC
TGATCTCCGCGTTCGCTATCGGTTCTTATGGTCTGGGTAGCCTGGGTTTCAAATTTATCGACACGCAGCT
GCTGGAAACGGTCGGTCTGGAAAAAACCTTTGTGATTTGGGGAGCGATTGCGCTGTTGATGATTGTTTT
CGGCGCAACGTTAATGAAAGACGCACCAAAACAGGAAGTGAAAACCAGCAATGGTGTGGTGGAGAAA
GATTACACGCTGGCAGAGTCGATGCGTAAACCGCAGTACTGGATGTTAGCGGTAATGTTCCTGACCGC
CTGCATGAGCGGCCTGTACGTGATTGGGGTAGCGAAAGATATCGCCCAAAGTCTGGCACACCTTGATG
TGGTTTCCGCAGCCAATGCAGTCACTGTTATTTCCATCGCCAACCTTTCAGGTCGTCTGGTGCTGGGTAT
TCTGTCTGACAAAATCGCCCGTATCCGTGTTATTACCATTGGTCAGGTGATATCGCTGGTGGGTATGGC
GGCCCTGCTGTTTGCACCATTGAATGCAGTGACGTTCTTTGCAGCGATTGCCTGCGTGGCATTTAACTTT
GGCGGCACTATTACCGTCTTTCCGTCACTGGTCAGTGAGTTCTTTGGCCTCAATAACCTGGCGAAAAACT
ACGGTGTGATTTATCTCGGTTTCGGTATCGGTAGCATTTGTGGTTCGATTATCGCCTCACTGTTTGGCGG
CTTCTATGTGACCTTCTACGTGATTTTCGCCCTGCTGATTCTGTCATTGGCGCTTTCTACGACGATTCGTC
AGCCAGAGCAGAAAATGTTGCGTGAGGCGCATGGCTCCCTTTAAgctttgagaccacgaagtg

A.4.25 *E. coli YnfM* sequence (5'-3') modified for Golden Gate cloning (SC module) for *E. coli* expression construct

cactctgtggtctcaaATGAGCCGTACTACAACTGTTGATGGCGCTCCGGCAAGCGACACTGACAAGCAAAG
CATTTCTCAGCCAAATCAATTTATTAAACGCGGTACGCCGCAATTTATGCGCGTCACCCTGGCGCTGTTC
TCTGCCGGACTGGCAACATTTGCACTTCTCTATTGTGTGCAGCCTATCCTTCCGGTGCTTTCGCAGGAGT
TTGGCTTAACCCCCGCGAACAGTAGTATTTCACTGTCCATTTCCACGGCGATGTTGGCTATTGGTTTGCT
GTTTACTGGCCCGCTATCCGATGCCATTGGTCGCAAACCAGTGATGGTCACGGCGCTACTGTTGGCCTC
CATTTGTACGTTACTTTCGACAATGATGACCAGCTGGCACGGCATTTTGATTATGCGCGCCTTGATTGG
GCTTTCGTTAAGTGGCGTGGCAGCTGTTGGCATGACTTATCTTAGCGAGGAAATCCATCCCAGTTTCGT
GGCCTTTTCAATGGGGTTGTATATCAGCGGCAACTCAATTGGCGGCATGAGCGGACGCTTAATTAGCG
GTGTCTTCACGGACTTTTTCAACTGGCGAATTGCTCTGGCGGCAATCGGTTGTTTCGCGCTGGCCTCGG
CGTTGATGTTCTGGAAAATCCTCCCTGAATCACGCCATTTTCGCCCGACTTCGCTGCGCCCTAAGACGTT
GTTTATCAACTTTCGTCTGCACTGGCGTGACCGGGGATTACCGTTATTGTTCGCAGAAGGCTTTTTGCTG
ATGGGGTCGTTCGTCACGCTGTTTAATTACATCGGCTATCGGTTGATGCTCTCCCCCTGGCATGTCAGTC
AGGCCGTGGTTGGCTTATTATCGCTGGCTTATTTGACCGGTACATGGAGCTCACCCAAAGCCGGAACCA
TGACCACCCGCTATGGGCGTGGTCCAGTGATGTTGTTTTCGACGGGGGTTATGCTGTTTGGTTTACTGA
TGACCTTATTCAGCTCGCTGTGGCTGATCTTTGCCGGAATGTTACTCTTCTCAGCAGGATTCTTCGCAGC
CCACTCAGTAGCCAGCAGCTGGATCGGCCCCCGCGCAAAACGCGCTAAAGGCCAGGCCTCCTCGCTGT
ATCTGTTCAGTTACTATCTGGGGTCGAGTATTGCCGGGACGCTGGGTGGTGTTTTCTGGCATAACTATG
GCTGGAACGGCGTCGGCGCATTTATTGCTCTGATGCTGGTCATTGCTCTGCTGGTCGGGACGCGTTTGC
ATCGTCGTCTGCACGCCTGAgctttgagaccacgaagtg

A.4.26 Generation of *YfeH* level 0 SC module using PCR

Table A.23: PCR primers used to amplify *E. coli YfeH* and append level 0 SC module adaptors for Golden Gate cloning

Primer ID	Primer sequence (5'-3')
YfeH_GGextensions_for	cactctgtggtctcaaATGAAACTTTTTCGTATCCTCGATCC
YfeH_GGextensions_rev	cacttcgtggtctcaaagcTTAAGCTTTATCGGCGCTGC

Table A.24: Parts list for *Yfeh* level 0 SC module vector

Part ID	Part description	Selection	Reference
pICH41308	Backbone	SpecR	Weber et al. (2011)[331]
NA	Lv0 SC module YfeH CDS. PCR product. Digested with BsaI.	NA	This study

A.4.27 Generation of *DctA* level 0 SC module using PCR

Table A.25: PCR primers used to amplify *E. coli DctA* and append level 0 SC module adaptors for Golden Gate cloning

Primer ID	Primer sequence (5'-3')
DctA_GGextensions_for	cactctgtggtctcaaATGAAAACCTCTCTGTTTAAAAGC
DctA_GGextensions_rev	cacttcgtggtctcaaagcTTAAGAGGATAATTCGTGCGTT

Table A.26: Parts list for *DctA* level 0 SC module vector

Part ID	Part description	Selection	Reference
pICH41308	Backbone	SpecR	Weber et al. (2011)[331]
NA	Lv0 SC module DctA CDS. PCR product. Digested with BsaI.	NA	This study

A.4.28 Level 1 vectors for protein expression in *E. coli*

Table A.27: Parts list for lv1 vector pOM032, for expression of *S. italica HATCH1* in *E. coli*

Part ID	Part description	Selection	Reference
pOGG024	Backbone	GentR	Geddes et al. (2019)[332]
pOGG137	Lv0 PU module DAPG-inducible promoter	SpecR	Grant (2019)[333]
pUC57_Seita.4G275500	Lv0 SC module SiHATCH1 CDS	AmpR	From M. Niklaus
pOGG156	Lv0 T module DT15 terminator	SpecR	Grant (2019)[333]

Table A.28: Parts list for lv1 vector pOM033, for expression of *A. thaliana HATCH1* in *E. coli*

Part ID	Part description	Selection	Reference
pOGG024	Backbone	GentR	Geddes et al. (2019)[332]
pOGG137	Lv0 PU module DAPG-inducible promoter	SpecR	Grant (2019)[333]
pUC57_AT4G19390	Lv0 SC module AtHATCH1 CDS	AmpR	From M. Niklaus
pOGG156	Lv0 T module DT15 terminator	SpecR	Grant (2019)[333]

Table A.29: Parts list for lv1 vector pOM004, for expression of *Z. mays HATCH3.1* (candidate 4) in *E. coli*

Part ID	Part description	Selection	Reference
pOGG024	Backbone	GentR	Geddes et al. (2019)[332]
pOGG137	Lv0 PU module DAPG-inducible promoter	SpecR	Grant (2019)[333]
pTwist_GRMZM2G327686	Lv0 SC module ZmHATCH3.1 CDS	AmpR	This study
pOGG156	Lv0 T module DT15 terminator	SpecR	Grant (2019)[333]

Table A.30: Parts list for lv1 vector pOM014, for expression of *Z. mays HATCH1* (candidate 10) in *E. coli*

Part ID	Part description	Selection	Reference
pOGG024	Backbone	GentR	Geddes et al. (2019)[332]
pOGG137	Lv0 PU module DAPG-inducible promoter	SpecR	Grant (2019)[333]
pTwist_GRMZM2G179292	Lv0 SC module ZmHATCH1 CDS	AmpR	This study
pOGG156	Lv0 T module DT15 terminator	SpecR	Grant (2019)[333]

Table A.31: Parts list for lv1 vector pOM013, for expression of *Z. mays HATCH3.2* (candidate 11) in *E. coli*

Part ID	**Part description**	**Selection**	**Reference**
pOGG024	Backbone	GentR	Geddes et al. (2019)[332]
pOGG137	Lv0 PU module DAPG-inducible promoter	SpecR	Grant (2019)[333]
pTwist_GRMZM2G133400	Lv0 SC module ZmHATCH3.2 CDS	AmpR	This study
pOGG156	Lv0 T module DT15 terminator	SpecR	Grant (2019)[333]

Table A.32: Parts list for lv1 vector pOM015, for expression of *A. thaliana HATCH3* in *E. coli*

Part ID	**Part description**	**Selection**	**Reference**
pOGG024	Backbone	GentR	Geddes et al. (2019)[332]
pOGG137	Lv0 PU module DAPG-inducible promoter	SpecR	Grant (2019)[333]
pTwist_AT5G13720	Lv0 SC module AtHATCH3 CDS	AmpR	This study
pOGG156	Lv0 T module DT15 terminator	SpecR	Grant (2019)[333]

Table A.33: Parts list for lv1 vector pOM016, for expression of *S. bicolor HATCH2* in *E. coli*

Part ID	Part description	Selection	Reference
pOGG024	Backbone	GentR	Geddes et al. (2019)[332]
pOGG137	Lv0 PU module DAPG-inducible promoter	SpecR	Grant (2019)[333]
pTwist_ Sobic.001G096800	Lv0 SC module SbHATCH2 CDS	AmpR	This study
pOGG156	Lv0 T module DT15 terminator	SpecR	Grant (2019)[333]

Table A.34: Parts list for lv1 vector pOM017, for expression of *O. sativa HATCH1* in *E. coli*

Part ID	Part description	Selection	Reference
pOGG024	Backbone	GentR	Geddes et al. (2019)[332]
pOGG137	Lv0 PU module DAPG-inducible promoter	SpecR	Grant (2019)[333]
pTwist_ LOC_Os06g49640	Lv0 SC module OsHATCH1 CDS	AmpR	This study
pOGG156	Lv0 T module DT15 terminator	SpecR	Grant (2019)[333]

Table A.35: Parts list for lv1 vector pOM018, for expression of *O. sativa HATCH2* in *E. coli*

Part ID	Part description	Selection	Reference
pOGG024	Backbone	GentR	Geddes et al. (2019)[332]
pOGG137	Lv0 PU module DAPG-inducible promoter	SpecR	Grant (2019)[333]
pTwist_ LOC_Os03g52910	Lv0 SC module OsHATCH2 CDS	AmpR	This study
pOGG156	Lv0 T module DT15 terminator	SpecR	Grant (2019)[333]

Table A.36: Parts list for lv1 vector pOM019, for expression of *O. sativa HATCH3.1* in *E. coli*

Part ID	Part description	Selection	Reference
pOGG024	Backbone	GentR	Geddes et al. (2019)[332]
pOGG137	Lv0 PU module DAPG-inducible promoter	SpecR	Grant (2019)[333]
pTwist_ LOC_Os07g46320	Lv0 SC module OsHATCH3.1 CDS	AmpR	This study
pOGG156	Lv0 T module DT15 terminator	SpecR	Grant (2019)[333]

Table A.37: Parts list for lv1 vector pOM020, for expression of *O. sativa HATCH3.2* in *E. coli*

Part ID	Part description	Selection	Reference
pOGG024	Backbone	GentR	Geddes et al. (2019)[332]
pOGG137	Lv0 PU module DAPG-inducible promoter	SpecR	Grant (2019)[333]
pTwist_ LOC_Os07g46330	Lv0 SC module OsHATCH3.2 CDS	AmpR	This study
pOGG156	Lv0 T module DT15 terminator	SpecR	Grant (2019)[333]

Table A.38: Parts list for lv1 vector pOM001, for expression of *Z. mays* candidate 1 (*MEP3*) in *E. coli*

Part ID	Part description	Selection	Reference
pOGG024	Backbone	GentR	Geddes et al. (2019)[332]
pOGG137	Lv0 PU module DAPG-inducible promoter	SpecR	Grant (2019)[333]
pTwist_ GRMZM2G138258	Lv0 SC module Zm candidate 1 CDS	AmpR	This study
pOGG156	Lv0 T module DT15 terminator	SpecR	Grant (2019)[333]

Table A.39: Parts list for lv1 vector pOM002, for expression of *Z. mays* candidate 2 (*APE1*) in *E. coli*

Part ID	Part description	Selection	Reference
pOGG024	Backbone	GentR	Geddes et al. (2019)[332]
pOGG137	Lv0 PU module DAPG-inducible promoter	SpecR	Grant (2019)[333]
pTwist_ GRMZM2G099367	Lv0 SC module Zm candidate 2 CDS	AmpR	This study
pOGG156	Lv0 T module DT15 terminator	SpecR	Grant (2019)[333]

Table A.40: Parts list for lv1 vector pOM003, for expression of *Z. mays* candidate 3 in *E. coli*

Part ID	Part description	Selection	Reference
pOGG024	Backbone	GentR	Geddes et al. (2019)[332]
pOGG137	Lv0 PU module DAPG-inducible promoter	SpecR	Grant (2019)[333]
pTwist_ GRMZM2G088138	Lv0 SC module Zm candidate 3 CDS	AmpR	This study
pOGG156	Lv0 T module DT15 terminator	SpecR	Grant (2019)[333]

Table A.41: Parts list for lv1 vector pOM008, for expression of *Z. mays* candidate 5 (*PIC1*) in *E. coli*

Part ID	Part description	Selection	Reference
pOGG024	Backbone	GentR	Geddes et al. (2019)[332]
pOGG137	Lv0 PU module DAPG-inducible promoter	SpecR	Grant (2019)[333]
pTwist_ GRMZM2G145500	Lv0 SC module candidate 5 CDS	AmpR	This study
pOGG156	Lv0 T module DT15 terminator	SpecR	Grant (2019)[333]

Table A.42: Parts list for lv1 vector pOM009, for expression of *Z. mays* candidate 6 in *E. coli*

Part ID	Part description	Selection	Reference
pOGG024	Backbone	GentR	Geddes et al. (2019)[332]
pOGG137	Lv0 PU module DAPG-inducible promoter	SpecR	Grant (2019)[333]
pTwist_ GRMZM2G004352	Lv0 SC module candidate 6 CDS	AmpR	This study
pOGG156	Lv0 T module DT15 terminator	SpecR	Grant (2019)[333]

Table A.43: Parts list for lv1 vector pOM010, for expression of *Z. mays* candidate 7 (*AVT3*) in *E. coli*

Part ID	Part description	Selection	Reference
pOGG024	Backbone	GentR	Geddes et al. (2019)[332]
pOGG137	Lv0 PU module DAPG-inducible promoter	SpecR	Grant (2019)[333]
pTwist_ GRMZM2G074053	Lv0 SC module candidate 7 CDS	AmpR	This study
pOGG156	Lv0 T module DT15 terminator	SpecR	Grant (2019)[333]

Table A.44: Parts list for lv1 vector pOM011, for expression of *Z. mays* candidate 8 in *E. coli*

Part ID	Part description	Selection	Reference
pOGG024	Backbone	GentR	Geddes et al. (2019)[332]
pOGG137	Lv0 PU module DAPG-inducible promoter	SpecR	Grant (2019)[333]
pTwist_GRMZM2G015033	Lv0 SC module candidate 8 CDS	AmpR	This study
pOGG156	Lv0 T module DT15 terminator	SpecR	Grant (2019)[333]

Table A.45: Parts list for lv1 vector pOM012, for expression of *Z. mays* candidate 9 in *E. coli*

Part ID	Part description	Selection	Reference
pOGG024	Backbone	GentR	Geddes et al. (2019)[332]
pOGG137	Lv0 PU module DAPG-inducible promoter	SpecR	Grant (2019)[333]
pTwist_ GRMZM2G078118	Lv0 SC module candidate 9 CDS	AmpR	This study
pOGG156	Lv0 T module DT15 terminator	SpecR	Grant (2019)[333]

Table A.46: Parts list for lv1 vector pOM005, for expression of *Z. mays DCT2* in *E. coli*

Part ID	Part description	Selection	Reference
pOGG024	Backbone	GentR	Geddes et al. (2019)[332]
pOGG137	Lv0 PU module DAPG-inducible promoter	SpecR	Grant (2019)[333]
pTwist_ GRMZM2G086258	Lv0 SC module ZmDCT2 CDS	AmpR	This study
pOGG156	Lv0 T module DT15 terminator	SpecR	Grant (2019)[333]

Table A.47: Parts list for lv1 vector pOM026, for expression of *S. viridis BASS2* in *E. coli*

Part ID	Part description	Selection	Reference
pOGG024	Backbone	GentR	Geddes et al. (2019)[332]
pOGG137	Lv0 PU module DAPG-inducible promoter	SpecR	Grant (2019)[333]
pUC57_Sevir.5G250900	Lv0 SC module SvBASS2 CDS	AmpR	From M. Niklaus
pOGG156	Lv0 T module DT15 terminator	SpecR	Grant (2019)[333]

Table A.48: Parts list for lv1 vector pOM006, for expression of *E. coli DctA* in *E. coli*

Part ID	Part description	Selection	Reference
pOGG024	Backbone	GentR	Geddes et al. (2019)[332]
pOGG137	Lv0 PU module DAPG-inducible promoter	SpecR	Grant (2019)[333]
DctA_lv0	Lv0 SC module DctA CDS	SpecR	This study
pOGG156	Lv0 T module DT15 terminator	SpecR	Grant (2019)[333]

Table A.49: Parts list for lv1 vector pOM007, for expression of *E. coli CstA* in *E. coli*

Part ID	Part description	Selection	Reference
pOGG024	Backbone	GentR	Geddes et al. (2019)[332]
pOGG137	Lv0 PU module DAPG-inducible promoter	SpecR	Grant (2019)[333]
CstA_pTwist	Lv0 SC module CstA CDS	AmpR	This study
pOGG156	Lv0 T module DT15 terminator	SpecR	Grant (2019)[333]

Table A.50: Parts list for lv1 vector pOM022, for expression of *E. coli BtsT* in *E. coli*

Part ID	Part description	Selection	Reference
pOGG024	Backbone	GentR	Geddes et al. (2019)[332]
pOGG137	Lv0 PU module DAPG-inducible promoter	SpecR	Grant (2019)[333]
BtsT_pTwist	Lv0 SC module BtsT CDS	AmpR	This study
pOGG156	Lv0 T module DT15 terminator	SpecR	Grant (2019)[333]

Table A.51: Parts list for lv1 vector pOM023, for expression of *E. coli YfeH* in *E. coli*

Part ID	Part description	Selection	Reference
pOGG024	Backbone	GentR	Geddes et al. (2019)[332]
pOGG137	Lv0 PU module DAPG-inducible promoter	SpecR	Grant (2019)[333]
YfeH_lv0	Lv0 SC module DctA CDS	SpecR	This study
pOGG156	Lv0 T module DT15 terminator	SpecR	Grant (2019)[333]

Table A.52: Parts list for lv1 vector pOM024, for expression of *E. coli YnfM* in *E. coli*

Part ID	Part description	Selection	Reference
pOGG024	Backbone	GentR	Geddes et al. (2019)[332]
pOGG137	Lv0 PU module DAPG-inducible promoter	SpecR	Grant (2019)[333]
YnfM_pTwist	Lv0 SC module YnfM CDS	AmpR	This study
pOGG156	Lv0 T module DT15 terminator	SpecR	Grant (2019)[333]

Table A.53: Parts list for lv1 vector pOM025, for expression of *E. coli YhjX* in *E. coli*

Part ID	**Part description**	**Selection**	**Reference**
pOGG024	Backbone	GentR	Geddes et al. (2019)[332]
pOGG137	Lv0 PU module DAPG-inducible promoter	SpecR	Grant (2019)[333]
YhjX_pTwist	Lv0 SC module YhjX CDS	AmpR	This study
pOGG156	Lv0 T module DT15 terminator	SpecR	Grant (2019)[333]

Table A.54: Level 1 *E. coli* expression vector sequencing primers

Primer ID	**Primer sequence (5'-3')**
GG_seq_for_1	ACATTTCCCCGAAAAGTGCCAC
GG_seq_rev_1	ACAGTTTACGAACCGAACAGGC
GG_seq_rev_2	CTTCTTGGTCGAAGGCAGCAAG

A.5 Golden Gate cloning of GFP fusion constructs for expression in *Nicotiana benthamiana*

A.5.1 *S. italica TPT* (Seita.5G002900) sequence (5'-3') modified for Golden Gate cloning as a coding sequence with a *C*-terminal tag (SC1 module)

cactctgtggtctcaaATGTCGGCGCTCGGCGCCCTCTCCGGCGGCGCCGCCGGTGTCGCTGGCCTCCTCAGC
CTCCGCCGCCGTGCAGCCGCGCCACCACCACCCTCTGCGCTCGCCTCCTCGTCGCACCTCCCGCCGCTTA
AGTGCGCCGCCGTTCCCGATGCCGGCCACCTCGTCTGGGGCCGCCAGCTCCGCCCGGCGCTCGCCCTCC
TCCCGTTCCCGTCGCAGGCCGCCAGGAAGCAGACCCCACGCCCGCCCGCCGCCGCCTCCGCAGGGGAG
GCCAAGCCCGCTGGTTTCCTCAGCAAGTACCCTGCCCTGGTCACCGGCTTCTTCTTCTTCATGTGGTACT
TCCTCAACGTCATATTCAACATCCTCAACAAGAAGATCTACAACTACTTCCCCTACCCATACTTTGTGTCG
GCGATCCATCTTTTAGTGGGTGTCGTATACTGCCTCATCAGCTGGAGTCTCGGACTCCCTAAGCGTGCG
CCTGTCAATGCGAATCTCCTAAAGCTGCTCTTTCCAGTGGCGCTGTGCCATGCTCTTGGTCATGTCACAA
GCAACGTCTCCTTCGCTGCTGTTGCAGTCTCATTTGCCCACACTATTAAAGCTCTGGAGCCCTTCTTCAAC
GCAGCCGCTACTCAGTTTATCCTTGGACAGCAGGTTCCCTTGTCTCTGTGGCTGTCCCTTGCCCCTGTTG
TCATTGGTGTTTCAATGGCATCCCTCACTGAGCTCTCATTCAACTGGACTGGCTTCATCAATGCCATGAT
ATCTAACATCTCCTTCACTTACAGGAGCATTTACTCCAAGAAAGCTATGACTGACATGGATAGCACCAAC
GTGTATGCTTACATCTCAATAATTTCTCTCATTGTTTGCATTCCACCTGCACTGATGTTTGAAGGGCCTAA
ACTGATGCAGCACGGATTTAATGATGCTATTGCAAAAGTAGGTCTGCAAAAGTTTGTTACGGACCTTTT
CTTAGTTGGACTGTTCTACCATCTGTATAACCAGGTTGCTACAAATACTTTGGAGCGAGTGGCTCCTCTG
ACACACGCTGTTGGCAATGTGTTGAAACGTGTGTTTGTCATTGGTTTTTCAATCATTGTTTTTGGCAACA
GAATCACGACACAAACTGGTATTGGCACATCCATTGCCATAGCTGGTGTGGCCATGTACTCTTACATCA
AGGCTAAGATTGAAGAGGAGAAAAGGAAAAAGAAGAGCGCGgtgtgagaccacgaagtg

A.5.2 *S. italica BASS2* (Seita.5G242900) sequence (5'-3') modified for Golden Gate cloning as a coding sequence with a *C*-terminal tag (SC1 module)

cactctgtggtctcaaATGGCACCATCAGCTACATCAAGTATGGCTTCAGTTAGCCGTGCGCTTCGACCAAGA
CCAAGAGCAGCGGCTTGTTCCGCCCCCAGGTTAGGATGTGGCCTGGGTATAGCTTGTTCCATGCCTTCC
CTGGGTCTTGCGGTCGTTACCGCCCCAAGCGCCACTGTGACACCCGCACTCAGACGTAGGCAGATACTT
TGTCAAGCCGAGGCAAATATTAGCAACAACTTGCCCGAGAAACTACCCGCCAAAGTCTCTCAACCCGCA
AAGGTGTCTCAACCAGCAGAGGTTTCTCAACCAGCAGAGGTGTCACAATATGAGAAGATCGTTGAACT
CTTAACAACATTGTTTCCCGTTTGGGTGATTATTGGCACAGTGATCGGCATATATAAACCTGCGATGGT
GACTTGGCTCGATACAGACTTATTTACAATAGGTCTGGGCCTATTGATGCTCAGCATGGGACTTACTCTT
ACCTTCGAGGACTTCCGTCGTTGTCTTAGAAACCCTTGGACCGTGGGCATTGGTTTCCTGGCGCAATAC
TGTGTGAAGCCACTTTTGGGGCTAGCAATAGCTACCACACTCAAACTCCCAGCGCCACTGGCTACCGGG
TTGATATTAGTGTCCTGCTGTCCAGGCGGCCAGGCTTCCAACGTGGCTACATATATCTCTAAAGGTAAC
GTGGCTCTCTCTGTCTTAATGACAACTTGCTCTACCATCGGAGCCATTATCATGACCCCGCTATTGACAA
AGTTACTCGCAGGACAGTTGGTGCCAGTCGACGCCGCCGGCCTGGCAATTTCTACCTTTCAAGTAGTGC
TTCTACCAACCGTGTTGGGTGTTCTTGCACATGAGTACTTTCCTAAATTCACCGAAAGAATAATAACAGT
GGCGCCACTATTTGGGGTTCTATTGACTACACTTCTGTGCGCAAGTCCAATAGGACAGGTGGCTGAGGT
CCTTAAGACGCAGGGCGCACAGTTAATAATACCGGTAGCATTGCTTCACGTAGCGGCTTTCGCGTTGGG
TTATTGGCTGAGTCGGTTTTCATCTTTCGGAGAGTCTACAAGTCGTACAATAAGTATAGAGTGCGGAAT
GCAAAGCAGTGCCCTAGGTTTCCTACTGGCCCAGAAACATTTTACCAACCCCTTAGTCGCCGTACCAAG
TGCAGTTTCAGTCGTTGCTATGGCACTAGGTGGCTCCGCATTGGCCGTGTTTTGGCGCTCAATTGGACT
CCCTGCTAACGATAAAGATGATTTTAAAGAAggtgtgagaccacgaagtg

A.5.3 *A. thaliana TPT* (AT5G46110) sequence (5'-3') modified for Golden Gate cloning as a coding sequence with a *C*-terminal tag (SC1 module)

cactctgtggtctcaaATGGAGTCACGCGTGCTGTTACGCGCCACCGCGAATGTCGTTGGAATTCCGAAATTG
AGACGACCAATCGGAGCGATCCACCGTCAATTCAGCACTGCATCGTCATCCTCGTTCTCGGTTAAACCA
ATCGGAGGAATCGGAGAGGGAGCGAATCTGATCTCCGGTCGTCAGCTTCGTCCAATTCTTCTTCTCGAT
TCGTCGGCGATCAACGGAGGAGAGAAAAGAGAAATTCTCAAACCGGTTAAAGCCGCCGCTGCTGAAG
GTGGAGATACCGCTGGGGATGCTAAAGTTGGATTCCTCGCCAAGTATCCATGGCTAGTCACTGGATTCT
TCTTCTTCATGTGGTACTTCTTGAATGTGATTTTCAACATCCTTAACAAGAAGATCTATAATTACTTCCCT
TATCCCTATTTTGTATCGGTGATACACTTGTTCGTGGGAGTTGTTTACTGCTTGATCAGCTGGTCCGTGG
GTCTACCTAAACGTGCCCCGATTGACTCGAACCTCCTCAAGGTATTGATACCAGTCGCAGTCTGTCACGC
CTTAGGCCATGTCACTAGCAATGTCTCTTTCGCTGCGGTTGCTGTCTCCTTCACTCACACCATCAAAGCA
CTTGAGCCATTCTTCAATGCGGCTGCTTCTCAATTCATTATGGGACAATCCATCCCCATAACACTATGGTT
GTCTCTAGCTCCTGTTGTTCTTGGTGTTGCAATGGCTTCACTAACTGAGCTATCATTCAACTGGCTCGGT
TTCATCAGTGCTATGATATCAAACATTTCTTTCACTTACCGAAGCATCTTCTCCAAGAAAGCCATGACTG
ATATGGACAGTACAAATGTCTACGCTTACATCTCCATCATCGCACTCTTCGTCTGCATTCCTCCTGCCATC
ATCGTTGAAGGTCCTAAACTACTTAACCATGGTTTCGCCGACGCGATTGCTAAAGTTGGAATGACTAAA
TTCATCTCTGATCTCTTCTGGGTTGGAATGTTTTACCATCTCTACAATCAGCTGGCTACCAATACCTTGGA
GAGGGTTGCACCGCTGACTCACGCGGTTGGAAACGTTCTGAAACGTGTGTTCGTGATCGGTTTCTCCAT
CGTTATCTTCGGAAACAAGATATCGACACAGACAGGTATAGGAACAGGAATAGCCATTGCTGGAGTTG
CAATGTACTCTATCATTAAGGCCAAGATCGAAGAAGAGAAACGGCAAGGAAAGAAAGCAggtgtgagacc
acgaagtg

A.5.4 *S. italica HATCH1* (Seita.4G275500) sequence (5'-3') modified for Golden Gate coding sequence with a *C*-terminal tag (SC1 module)

cactctgtggtctcaaATGAAGCTCAGGCCTCTCACTTGCGTGGCGGCGGGGTGCGCCGGGTGGGCGTGGA
GGCCGAGGTCGCGCGTGCGGTCAGAGGCGGTGTCACCCAAGCGTTCCCACGCGGCAGCGGCGGCGGC
GGGCGCGGTTCATTCGGAGGAGCACCGCCGCGGCGGCATGCGCGAGGTGCTCTTCCGCCCGGTGGGG
CTGCCCACCGAGACGAAGTTCGGGGCGGGGCTGGAGGATCGGATCGAGAAGGTCATCTGCGCCTGCC
GCTTCATGACCTTCCTCGGCATCGGCGGCTTGCTCGCCGGCTGCGTCCCCTGCTTCCTCAAGGGATGCG
TTTATGTGATGGACGCCTTCGTCGAGTACTACCTGCACGGCGGTGGAATGCTCATCCTAATGTTGCTTG
AAGCCATTGACATGTTTCTCATTGGAACGGTCATGTTTGTATTCGGGACGGGCTTGTATGAGCTGTTCA
TCAGTGAAATGGACATGTCTTATGGCTCCAACTTGTTTGGCTTGTTCAGTCTTCCGGAACGACCCAAGT
GGCTGGTAATCCAGTCGGTGAATGATCTTAAGACAAAGCTGGGCCATGTCATTGTCATGAGTCTACTGG
TTGGCATCTTTGAGAAGAGCTGGAGAGTGACCATTACATCCTGTACTGACCTCCTTTGCTTCGCTGCATC
AATCTTCCTCTCCTCAGGTTGCCTCTACCTACTTTCCAGGCTCAGTAACACCAAAGGAGGGAGCCATACC
ggtgtgagaccacgaagtg

A.5.5 *A. thaliana HATCH1* (AT4G19390) sequence (5'-3') modified for Golden Gate cloning as a coding sequence with a *C*-terminal tag (SC1 module)

cactctgtggtctcaaATGACGACGCCGTGCCGGACCATTAACGCTAATGCTATCGCCGCACCGTCGCCTTCG
GGATTGATCTTTAACGGGTTTCGTGATTTCGTCCCGATTGAAAAGAGGCTTGTTATATCAAGCTTTCGTG
GCTTGAAACTACCGTCGAGGACGACGAAAACTATTACGAGCTCTGACTGGAGTTGGAGTTATAGATCT
CCGGGAAGATTGGCGTCTGCGTCTACGTCTACGTCTGCGTCTACGTCTACGTCTGCGGCTGTAACATCG
AATTCTACGAATAGATTTGAAGCTTTAGAAGAGGGGATTGAGAAGGTCATTTATAGTTGTCGATTTATG
ACGTTTCTTGGAACATTAGGATCTTTGCTCGGATCGGTTCTCTGTTTCATCAAGGGATGTATGTATGTTG
TAGACTCGTTTTTGCAGTATTCGGTGAATCGTGGGAAAGTGATATTCCTTTTGGTTGAGGCCATAGATA
TATATCTCCTAGGAACTGTGATGTTAGTCTTTGGATTGGGACTTTACGAGTTGTTCATCAGCAATCTCGA
TACTTCAGAGTCGCGCACACACGATATTGTTTCCAATAGGTCGAGCCTCTTTGGCATGTTCACCTTAAAG
GAGCGACCTCAATGGTTAGAGGTGAAATCGGTAAGCGAGCTGAAAACAAAGTTGGGACATGTAATAG
TGATGCTACTATTGATTGGTTTGTTTGACAAGAGCAAGAGAGTAGTCATAACTTCTGTCACGGATTTGCT
ATGTATCTCTGTTTCTATCTTTTTTTCTTCTGCTTGTCTCTTCTTGCTCTCTAGGCTCAATGGTTCACACggt
gtgagaccacgaagtg

A.5.6 *Z. mays HATCH3.1* (GRMZM2G327686) sequence (5'-3') modified for Golden Gate cloning as a coding sequence with a *C*-terminal tag (SC1 module)

cactctgtggtctcaaATGGCTCTTTTGGTCCTGCGAGCTCCCGCTGCCGTACATGCAGCCTCTCGACTTCTTC
CTCCACAGCCAAGGCGTAGGCGACGTTTGGTGGCAGTAGCTTCAGCAGCCAGCTCTGCCCCCTCCGGA
GAGGTATCAAGTCAACACGGTGGCGGAGGAGGCGGTGGTTATGGGATAGTCGGTGGGCCAAATGGA
AACGCCGTCGTCCCGGCTACTAAATCAACAGTCGTTGAAACCACTGTGGAAAGGGTGATCTTTGATTTT
AGATTTCTCGCCTTGCTCGCCGTCGCAGGATCCCTTGCAGGGTCCCTCCTTTGCTTCTTGAATGGCTGCG
TTTTTATCAAGGAGGCATACCAAGTTTATTGGAGTTCATGTGTTAAAGGTGTCCACACTGGGCAAATGG
TCTTAAAGGTAGTGGAAGCAATAGATGTGTATTTAGCCGGGACAGTCATGCTCATATTTGGGATGGGA
TTGTATGGGCTCTTTGTCTCAAACGCTAGCGCCGGGGTAGGCAGTGAATCAGACAGGGCTTTGAGCGG
GTCATCTCTCTTCGGCATGTTCGCCCTCAAAGAAAGACCCAAGTGGATGAAGATAACCTCATTAGATGA
GTTAAAGACCATTGTGGGGCATGTAATTGTGATGATCTTACTCGTAAAAATGTTTGAAAGATCTAAAAT
GGTTACTATAGCAACTGGATTAGACTTATTGTCTTATTCCATATGTATCTTTTTGAGCTCTGCATCTCTCT
ACATTTTACATAATCTTCATAAAGGGGACGACCACGAAGAGggtgtgagaccacgaagtg

A.5.7 *A. thaliana HATCH3* (AT5G13720) sequence (5'-3') modified for Golden Gate cloning as a coding sequence with a *C*-terminal tag (SC1 module)

cactctgtggtctcaaATGGCTCTCTCCTCGTTGATCTCAGCCACTCCGCTCTCCCTTTCGGTGCCGAGATATCT
CGTGTTACCCACGCGCCGCCGTTTCCACTTGCCGTTAGCAACTCTTGACTCATCGCCTCCAGAGTCATCC
GCATCATCATCGATTCCTACCTCGATTCCCGTCAACGGAAACACGTTACCTAGTTCTTACGGAACTCGCA
AAGACGACAGCCCGTTTGCTCAGTTCTTTCGCTCCACCGAATCCAACGTTGAGAGGATAATATTTGATTT
CCGGTTCCTAGCGCTTTTGGCAGTAGGAGGTTCGCTGGCTGGTTCGCTACTCTGCTTTCTCAATGGGTG
TGTCTACATAGTGGAGGCATATAAAGTCTACTGGACTAACTGTTCAAAAGGCATCCATACCGGCCAAAT
GGTTTTACGCCTAGTCGAAGCTATCGATGTTTATCTAGCTGGGACTGTGATGTTAATATTTAGTATGGGT
TTGTATGGACTCTTCATCAGTCACTCGCCTCATGATGTTCCACCGGAATCCGATCGTGCCCTTAGATCCT
CTTCCCTCTTTGGTATGTTTGCAATGAAGGAGAGGCCAAAATGGATGAAGATCAGCTCACTTGATGAGC
TTAAAACCAAAGTGGGACATGTCATTGTTATGATTCTGCTAGTGAAGATGTTCGAGAGAAGCAAGATG
GTTACTATCGCCACCGGTCTAGATTTGCTTAGTTATTCCGTTTGCATCTTCTTGTCCTCTGCTTCTCTGTAT
ATCCTCCATAATCTCCACAAAGGAGAGACAggtgtgagaccacgaagtg

A.5.8 Binary vectors for transient expression of GFP or GFP-tagged proteins in *N. benthamiana*

Table A.55: Parts list for lv1 binary vector pOM038, for expression of *S. italica HATCH1* with *C*-terminal GFP tag in *N. benthamiana*

Part ID	Part description	Selection	Reference
pAGM37443	Backbone	KanR	From J. Kourelis
pICH51288	Lv0 PU module 2x35s promoter	SpecR	Engler et al. (2014)[334]
pUC57_Seita.4G275500	Lv0 SC1 module SiHATCH1 CDS	AmpR	From M. Niklaus
EC15095	Lv0 C2 module eGFP C-tag	SpecR	From G. Oldroyd
pICH41414	Lv0 T module 35s terminator	SpecR	Engler et al. (2014)[334]

Table A.56: Parts list for lv1 binary vector pOM039, for expression of *A. thaliana HATCH1* with *C*-terminal GFP tag in *N. benthamiana*

Part ID	Part description	Selection	Reference
pAGM37443	Backbone	KanR	From J. Kourelis
pICH51288	Lv0 PU module 2x35s promoter	SpecR	Engler et al. (2014)[334]
pUC57_AT4G19390	Lv0 SC1 module AtHATCH1 CDS	AmpR	From M. Niklaus
EC15095	Lv0 C2 module eGFP C-tag	SpecR	From G. Oldroyd
pICH41414	Lv0 T module 35s terminator	SpecR	Engler et al. (2014)[334]

Table A.57: Parts list for lv1 binary vector pOM040, for expression of free GFP in *N. benthamiana*

Part ID	Part description	Selection	Reference
pAGM37443	Backbone	KanR	From J. Kourelis
pICH51288	Lv0 PU module 2x35s promoter	SpecR	Engler et al. (2014)[334]
EC15112	Lv0 SC module eGFP CDS	SpecR	From G. Oldroyd
pICH41414	Lv0 T module 35s terminator	SpecR	Engler et al. (2014)[334]

Table A.58: Parts list for lv1 binary vector pOM041, for expression of *S. italica BASS2* with *C*-terminal GFP tag in *N. benthamiana*

Part ID	Part description	Selection	Reference
pAGM37443	Backbone	KanR	From J. Kourelis
pICH51288	Lv0 PU module 2x35s promoter	SpecR	Engler et al. (2014)[334]
pTwist_Seita.5G242900	Lv0 SC1 module SiBASS2 CDS	AmpR	This study
EC15095	Lv0 C2 module eGFP C-tag	SpecR	From G. Oldroyd
pICH41414	Lv0 T module 35s terminator	SpecR	Engler et al. (2014)[334]

Table A.59: Parts list for lv1 binary vector pOM042, for expression of *S. italica TPT* with *C*-terminal GFP tag in *N. benthamiana*

Part ID	Part description	Selection	Reference
pAGM37443	Backbone	KanR	From J. Kourelis
pICH51288	Lv0 PU module 2x35s promoter	SpecR	Engler et al. (2014)[334]
pTwist_Seita.5G002900	Lv0 SC1 module SiTPT CDS	AmpR	This study
EC15095	Lv0 C2 module eGFP C-tag	SpecR	From G. Oldroyd
pICH41414	Lv0 T module 35s terminator	SpecR	Engler et al. (2014)[334]

Table A.60: Parts list for lv1 binary vector pOM043, for expression of *A. thaliana TPT* with *C*-terminal GFP tag in *N. benthamiana*

Part ID	Part description	Selection	Reference
pAGM37443	Backbone	KanR	From J. Kourelis
pICH51288	Lv0 PU module 2x35s promoter	SpecR	Engler et al. (2014)[334]
pTwist_AT5G46110	Lv0 SC1 module AtTPT CDS	AmpR	This study
EC15095	Lv0 C2 module eGFP C-tag	SpecR	From G. Oldroyd
pICH41414	Lv0 T module 35s terminator	SpecR	Engler et al. (2014)[334]

Table A.61: Parts list for lv1 binary vector pOM080, for expression of *Z. mays HATCH3.1* with *C*-terminal GFP tag in *N. benthamiana*

Part ID	Part description	Selection	Reference
pAGM37443	Backbone	KanR	From J. Kourelis
pICH51288	Lv0 PU module 2x35s promoter	SpecR	Engler et al. (2014)[334]
pTwist_GRMZM2G327686	Lv0 SC1 module ZmHATCH3.1 CDS	AmpR	This study
EC15095	Lv0 C2 module eGFP C-tag	SpecR	From G. Oldroyd
pICH41414	Lv0 T module 35s terminator	SpecR	Engler et al. (2014)[334]

Table A.62: Parts list for lv1 binary vector pOM081, for expression of *A. thaliana HATCH3* with *C*-terminal GFP tag in *N. benthamiana*

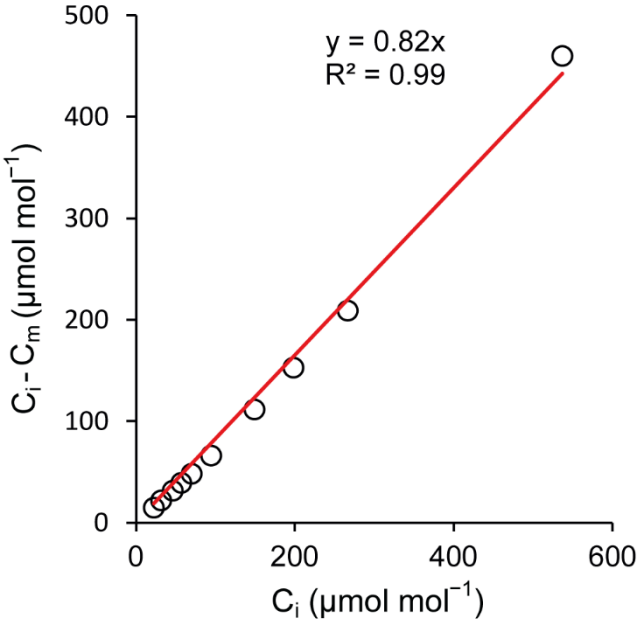
Part ID	**Part description**	**Selection**	**Reference**
pAGM37443	Backbone	KanR	From J. Kourelis
pICH51288	Lv0 PU module 2x35s promoter	SpecR	Engler et al. (2014)[334]
pTwist_AT5G13720	Lv0 SC1 module AtHATCH3 CDS	AmpR	This study
EC15095	Lv0 C2 module eGFP C-tag	SpecR	From G. Oldroyd
pICH41414	Lv0 T module 35s terminator	SpecR	Engler et al. (2014)[334]

Table A.63: Binary vector sequencing primers

Primer ID	**Primer sequence (5'-3')**
pJK268_seq_for_2	ATACCCGCGAAATTCAGGCCCG
pJK268_seq_rev	CGTGTCCCTCGGTCACATGTGC
pJK268_seq_rev	GCTTCATGTGGTCGGGGTAGCG

A.6 *Setaria viridis* flux balance analysis modelling

Figure A.1 Fitted linear regression between C_i-C_m and C_i. Made by Steven Kelly.



A.7 Golden Gate cloning of CRISPR/Cas9 constructs

A.7.1 Guide sequences

Table A.64: List of guides

Guide ID	Guide sequence (5'-3')
HATCH1_g2	CAGGCCTCTCACTTGCGTGG
HATCH1_g12	CCCCGCCCCGAACTTCGTCT
HATCH1_g11	CATTCGGAGGAGCACCGCCG
HATCH1_g5	GTCGAGTACTACCTGCACGG
HATCH3_g14	GCGCGACGCAGCTGCTCAGC
HATCH3_g5	TGGCCCTCCTCGCGATCGCC
HATCH3_g6	TCCCACGACAACGGGGTCGT
HATCH3_g12	TGTGGCTGCTCCCGCCGCGC

A.7.2 Level 0 cloning vectors containing single guide and tRNA-sgRNA backbone

Table A.65: Parts lists for level 0 vectors

Plasmid ID	Guide	Level 0 acceptor plasmid	Level 0 acceptor plasmid reference
pOM057	HATCH1_g2	pFH51	Hahn et al. (2020)[235]
pOM058	HATCH1_g12	pAK002	Hahn et al. (2020)[235]
pOM059	HATCH1_g11	pAK003	Hahn et al. (2020)[235]
pOM060	HATCH1_g5	pAK004	Hahn et al. (2020)[235]
pOM061	HATCH3_g14	pFH51	Hahn et al. (2020)[235]
pOM062	HATCH3_g5	pAK002	Hahn et al. (2020)[235]
pOM063	HATCH3_g6	pAK003	Hahn et al. (2020)[235]
pOM064	HATCH3_g12	pAK004	Hahn et al. (2020)[235]

Table A.66: Golden Gate reaction recipe

Volume	Component	Component concentration
1 uL	Level 0 backbone	100 ng/uL
1 uL	Annealed guide oligonucleotides	10 μM
1.5 uL	T4 DNA ligase buffer (NEB)	10 x
1.5 uL	BSA (NEB)	1 mg/mL
1 uL	BbsI-HF (NEB)	20,000 U/mL
1 uL	T4 DNA ligase (NEB)	400,000 U/mL
to 15 uL	ddH_2O	-

Table A.67: Golden Gate reaction thermocycler settings

Step #	Temperature	Time	Repeat
1	37°C	20 s	1 x
2	37°C	3 min	50 x
3	16°C	4 min	
4	50°C	5 min	1 x
5	80°C	5 min	1 x
6	16°C	∞	1 x

Table A.68: Level 0 vector sequencing primer

Primer ID	Primer sequence (5'-3')
FH32	GCAATGTAACATCAGAGATTTTGA

A.7.3 Level 1 cloning vectors containing PolIII promoter and tRNA-sgRNA array of multiple guides

Table A.69: Parts list for level 1 vector pOM065

Part ID	Part description	Selection	Reference
pICH47822	Lv1 backbone	AmpR	Weber et al. (2011)[331]
pFH38	Lv0 promoter module	SpecR	Hahn et al. (2020)[235]
pOM057	Lv0 guide module	KanR	This study
pOM058	Lv0 guide module	KanR	This study
pOM059	Lv0 guide module	KanR	This study
pOM060	Lv0 guide module	KanR	This study
pAK-EL04	Lv0 end-linker module	KanR	Hahn et al. (2020)[235]

Table A.70: Parts list for level 1 vector pOM066

Part ID	Part description	Selection	Reference
pICH47822	Lv1 backbone	AmpR	Weber et al. (2011)[331]
pFH38	Lv0 promoter module	SpecR	Hahn et al. (2020)[235]
pOM061	Lv0 guide module	KanR	This study
pOM062	Lv0 guide module	KanR	This study
pOM063	Lv0 guide module	KanR	This study
pOM064	Lv0 guide module	KanR	This study
pAK-EL04	Lv0 end-linker module	KanR	Hahn et al. (2020)[235]

Table A.71: Parts list for level 1 vector pOM067

Part ID	Part description	Selection	Reference
pICH47831	Lv1 backbone	AmpR	Weber et al. (2011)[331]
pFH38	Lv0 promoter module	SpecR	Hahn et al. (2020)[235]
pOM061	Lv0 guide module	KanR	This study
pOM062	Lv0 guide module	KanR	This study
pOM063	Lv0 guide module	KanR	This study
pOM064	Lv0 guide module	KanR	This study
pAK-EL04	Lv0 end-linker module	KanR	Hahn et al. (2020)[235]

Table A.72: Level 1 vector sequencing primers

Primer ID	Primer sequence (5'-3')
Lvl1_F(0229)	GAACCCTGTGGTTGGCATGCACATAC
Lvl1_R(0230)	CTGGTGGCAGGATATATTGTGGTG

A.7.4 Level 2 transformation vectors containing tRNA-sgRNA array expression cassette(s), SpCas9 expression cassette and Hygromycin expression cassette

Table A.73: Parts list for level 2 vector pOM068

Part ID	Part description	Selection	Reference
pAGM37443	Lv2 backbone	KanR	Grützner et al. (2011)[335]
EC17100	Lv1 Hyg expression cassette	AmpR	From Jane Langdale lab
pFH66	Lv1 SpCas9 expression cassette	AmpR	Hahn et al. (2020)[235]
pOM065	Lv1 *HATCH1*-targeting tRNA-sgRNA array expression cassette	AmpR	This study
pICH49277	Lv1 end-linker module	AmpR	Weber et al. (2011)[331]

Table A.74: Parts list for level 2 vector pOM069

Part ID	Part description	Selection	Reference
pAGM37443	Lv2 backbone	KanR	Grützner et al. (2011)[335]
EC17100	Lv1 Hyg expression cassette	AmpR	From Jane Langdale lab
pFH66	Lv1 SpCas9 expression cassette	AmpR	Hahn et al. (2020)[235]
pOM066	Lv1 *HATCH3*-targeting tRNA-sgRNA array expression cassette	AmpR	This study
pICH49277	Lv1 end-linker module	AmpR	Weber et al. (2011)[331]

Table A.75: Parts list for level 2 vector pOM070

Part ID	Part description	Selection	Reference
pAGM37443	Lv2 backbone	KanR	Grützner et al. (2011)[335]
EC17100	Lv1 Hyg expression cassette	AmpR	From Jane Langdale lab
pFH66	Lv1 SpCas9 expression cassette	AmpR	Hahn et al. (2020)[235]
pOM065	Lv1 *HATCH1*-targeting tRNA-sgRNA array expression cassette	AmpR	This study
pOM067	Lv1 *HATCH3*-targeting tRNA-sgRNA array expression cassette	AmpR	This study
pICH49283	Lv1 end-linker module	AmpR	Weber et al. (2011)[331]

Table A.76: Level 2 vector sequencing primers

Primer ID	Primer sequence (5'-3')
Lvl1_F(0229)	GAACCCTGTGGTTGGCATGCACATAC
Lvl1_R(0230)	CTGGTGGCAGGATATATTGTGGTG
Sv_CRISPR_lv2_seq_primer_1	GCTCATTATCTCTAGAGAGGG
Sv_CRISPR_lv2_seq_primer_2	TTCTCTAAACGCCACCAGCG
Sv_CRISPR_lv2_seq_primer_3	ACGCAAGTGAGAGGCCTGTG

A.8 *Setaria viridis* genotyping and sequencing

A.8.1 PCR primers used to confirm presence/absence of construct

Table A.77: Primers used to amplify Sevir.2G209800 (positive control)

Primer ID	Primer sequence (5'-3')
Sevir_for	CAGCAAGCCGCCTATATGGAG
Sevir_rev	TCGTCTCAGGAGTGGCCAAGT

Table A.78: Primers used to amplify Hygromycin

Primer ID	Primer sequence (5'-3')
Hyg_for	CAACCAAGCTCTGATAGAGT
Hyg_rev	GAAGAATCTCGTGCTTTCA

Table A.79: Primers used to amplify Cas9

Primer ID	Primer sequence (5'-3')
spCas9_for	CGAGCTGGAGAACGGCAG
spCas9_rev	GCTGGTGGCGTTTAGAGAACT

A.8.2 PCR/sequencing primers used to confirm exact mutations

Table A.80: Primers used to amplify/sequence *HATCH1*

Primer ID	Primer sequence (5'-3')
SvHATCH1_for_3	GCGCCGGCTACTTAACCATCCC
SvHATCH1_rev_3	ACCAGCCACTTGGGTCGTTCCT

Table A.81: Primers used to amplify/sequence *HATCH3*

Primer ID	**Primer sequence (5'-3')**
SvHATCH3_for_2	TGTCATCTGTGTCCGCGGGACT
SvHATCH3_rev_4	GCCCCGTATGAACTCCCTTGACAC

A.9 *Setaria viridis* transcriptomics

A.9.1 Accession numbers of relevant genes

Table A.82: Accession numbers of C_4 cycle and candidate C_4 cycle genes

Encoded protein	**Gene accession(s)**
Carbonic anhydrase (CA)	Sevir.5G247800
Phospho*enol*pyruvate carboxylase (PEPC)	Sevir.4G143500
Oxoglutarate/malate translocator 1 (OMT1)	Sevir.3G366500
NADP-malate dehydrogenase (NADP-MDH)	Sevir.6G256400
NADP-malic enzyme (NADP-ME)	Sevir.5G132500
Bile acid:sodium symporter 2 (BASS2)	Sevir.5G250900
Pyruvate, phosphate dikinase (PPDK)	Sevir.3G253900
Phospho*enol*pyruvate transporter (PPT)	Sevir.6G128500
PEPC kinase (PPCK)	Sevir.1G244600
Dicarboxylate transporter (DCT)	Sevir.9G380500, Sevir.2G247400, Sevir.6G194000
Chloroplast localised aspartate aminotransferase (cpAspAT)	Sevir.1G359000
Mitochondrial localised aspartate aminotransferase (mtAspAT)	Sevir.1G023700, Sevir.4G145900
Cytosol localised aspartate aminotransferase (cyAspAT)	Sevir.5G329700
Uncoupling protein (UCP)	Sevir.8G262500, Sevir.7G036800, Sevir.7G124600
Dicarboxylate and tricarboxylate carrier (DTC)	Sevir.3G076000, Sevir.5G144100
NAD-malate dehydrogenase (NAD-MDH)	Sevir.9G221600, Sevir.5G253100, Sevir.2G412900, Sevir.3G419600, Sevir.3G283600, Sevir.6G165400,

	Sevir.3G139600, Sevir.9G072000
Dicarboxylate carrier (DIC)	Sevir.6G191700
NAD-malic enzyme (NAD-ME)	Sevir.9G199800, Sevir.2G333400
Mitochondrial pyruvate carrier (MPC)	Sevir.2G359200, Sevir.6G021600, Sevir.6G148500, Sevir.2G198100, Sevir.6G148600
Alanine aminotransferase (AlaAT)	Sevir.9G253900, Sevir.9G254000, Sevir.2G401100, Sevir.6G209500
Phospho*enol*pyruvate carboxykinase (PCK)	Sevir.9G469000, Sevir.7G231100, Sevir.1G365400
ACC synthase 10 (ACS10)	Sevir.4G020000

A.9.2 Differentially expressed genes between *HATCH1* knock-out and wild-type plants

Table A.83: Significantly upregulated genes in *HATCH1* knock-out compared to wild-type plants

Gene	log2Fold Change	p_{adj}	Arabidopsis accession	Arabidopsis gene function	Localisation	Cell type
Sevir.3G419300	-0.6503419	0.01273519	AT4G39260.3	cold, circadian rhythm, and RNA binding 1	noTP	
Sevir.8G260100	-0.3959411	0.03730611	AT5G02380.1	metallothionein 2B	noTP	BS
Sevir.9G410500	-0.5421875	0.01167906	AT5G03380.1	Heavy metal transport/detoxification superfamily protein	noTP	M
Sevir.4G231800	-0.8801764	0.03842914	AT5G20190.1	Tetratricopeptide repeat (TPR)-like superfamily protein	cTP	BS
Sevir.7G275600	-1.0566162	0.02374412	AT1G47128.1	Granulin repeat cysteine protease family protein	SP	BS
Sevir.9G268900	-0.5870742	0.00823024	AT1G78230.1	Outer arm dynein light chain 1 protein	noTP	BS
Sevir.5G019900	-0.7185172	0.02512457	AT4G02380.1	senescence-associated gene 21	noTP	BS
Sevir.1G078126	-0.7438493	0.04236222			noTP	M
Sevir.3G093500	-0.5770852	0.00122562	AT5G50920.1	CLPC homologue 1	cTP	BS
Sevir.3G129400	-0.654746	0.04390438	AT5G24120.1	sigma factor E	cTP	
Sevir.7G223900	-0.6678061	0.00154292	AT5G13770.1	Pentatricopeptide repeat (PPR-like) superfamily protein	noTP	M
Sevir.5G020000	-0.8212473	0.0082858	AT4G02380.1	senescence-associated gene 21	mTP	BS
Sevir.4G260300	-0.4558147	0.04167221	AT5G57345.1		luTP	
Sevir.9G142200	-1.2628771	0.00079141	AT4G39260.3	cold, circadian rhythm, and RNA binding 1	noTP	M
Sevir.2G115600	-0.6691908	0.00012279	AT4G22890.5	PGR5-LIKE A	cTP	M
Sevir.4G299800	-0.4428272	0.04390438	AT1G20693.2	high mobility group B2	noTP	

Sevir.2G208800	-1.1099475	0.00449174	AT1G71010.1	FORMS APLOID AND BINUCLEATE CELLS 1C	noTP	BS
Sevir.2G136000	-0.574899	0.00152029	AT3G26580.1	Tetratricopeptide repeat (TPR)-like superfamily protein	cTP	
Sevir.9G041200	-0.4329104	0.02679249	AT1G52340.1	NAD(P)-binding Rossmann-fold superfamily protein	noTP	M
Sevir.1G159600	-0.9576175	0.00038018	AT4G39090.1	Papain family cysteine protease	SP	BS
Sevir.5G034900	-1.2246109	0.00074921	AT3G62410.1	CP12 domain-containing protein 2	cTP	BS
Sevir.8G039700	-0.7444351	0.03440491	AT3G11670.1	UDP-Glycosyltransferase superfamily protein	noTP	BS
Sevir.1G372900	-0.9749168	0.0089491	AT2G33150.1	peroxisomal 3-ketoacyl-CoA thiolase 3	noTP	
Sevir.5G308500	-0.5908879	0.04044911	AT2G42750.1	DNAJ heat shock N-terminal domain-containing protein	cTP	BS
Sevir.9G004800	-2.000857	1.78E-07	AT5G24490.1	30S ribosomal protein, putative	cTP	M
Sevir.3G102200	-0.8029713	0.02926549	AT1G06110.1	SKP1/ASK-interacting protein 16	cTP	M
Sevir.4G299100	-0.5334613	0.01164943	AT4G35090.1	catalase 2	SP	
Sevir.1G284600	-0.4921201	0.01273519	AT4G17830.1	Peptidase M20/M25/M40 family protein	noTP	BS
Sevir.2G323400	-0.6520702	0.01264779	AT1G76080.1	chloroplastic drought-induced stress protein of 32 kD	cTP	BS
Sevir.7G273000	-0.9111231	0.00205496	AT5G13800.1	pheophytinase	cTP	BS
Sevir.5G037100	-0.7214378	0.02814104	AT2G29650.1	phosphate transporter 4;1	cTP	M
Sevir.6G183100	-0.9670307	0.00030206	AT2G45550.1	cytochrome P450, family 76, subfamily C, polypeptide 4	SP	
Sevir.1G056600	-0.9793679	0.0450756	AT4G15480.1	UDP-Glycosyltransferase superfamily protein	cTP	
Sevir.4G020000	-0.6686145	4.13E-05	AT1G62960.1	ACC synthase 10	noTP	BS
Sevir.2G227800	-0.6759539	0.04404903	AT1G54100.1	aldehyde dehydrogenase 7B4	noTP	
Sevir.9G461000	-1.1338074	0.00195571	AT1G07350.1	RNA-binding (RRM/RBD/RNP motifs) family protein	noTP	
Sevir.9G027300	-0.5946765	0.01946956	AT1G07230.1	non-specific phospholipase C1	SP	
Sevir.1G244600	-1.8185148	0.00187825	AT1G08650.1	phosphoenolpyruvate carboxylase kinase 1	noTP	M
Sevir.7G082000	-1.0826357	0.00641196	AT3G27020.1	YELLOW STRIPE like 6	noTP	M
Sevir.5G004600	-0.5866185	0.02569461	AT1G64500.1	Glutaredoxin family protein	noTP	BS
Sevir.5G018200	-0.4330692	0.03940739	AT2G46420.1	Plant protein 1589 of unknown function	noTP	BS
Sevir.2G135500	-0.3764421	0.04382525	AT4G24770.1	31-kDa RNA binding protein	cTP	M
Sevir.1G132200	-1.0345692	3.70E-06	AT4G12290.1	Copper amine oxidase family protein	SP	BS
Sevir.5G063500	-0.6003711	0.00010479	AT2G43750.1	O-acetylserine (thiol) lyase B	cTP	
Sevir.5G042300	-0.6078754	0.04432427	AT1G48910.1	Flavin-containing monooxygenase family protein	noTP	M
Sevir.1G057600	-0.3837223	0.00278284	AT4G35220.1	Cyclase family protein	SP	BS
Sevir.1G369200	-0.504491	0.03721339	AT5G40810.1	Cytochrome C1 family	noTP	
Sevir.2G230100	-1.0137251	0.00573073	AT5G60360.1	aleurain-like protease	SP	
Sevir.5G010400	-0.6173071	0.01019939	AT3G48280.1	cytochrome P450, family 71, subfamily A, polypeptide 25	SP	M
Sevir.3G333700	-1.1285896	0.00136063	AT5G13180.1	NAC domain containing protein 83	noTP	BS
Sevir.5G314900	-1.1642056	0.00010192	AT5G19740.1	Peptidase M28 family protein	noTP	
Sevir.9G541400	-0.8855493	0.0332219	AT4G05160.1	AMP-dependent synthetase and ligase family protein	noTP	M
Sevir.9G442500	-0.5832241	0.01413029	AT1G07080.1	Thioredoxin superfamily protein	SP	M
Sevir.3G359500	-0.6320048	0.01917109	AT2G47870.1	Thioredoxin superfamily protein	noTP	BS
Sevir.3G146000	-1.3459249	0.03842914	AT2G34420.1	photosystem II light harvesting complex gene B1B2	cTP	
Sevir.5G392700	-0.542291	0.03223306	AT5G01260.2	Carbohydrate-binding-like fold	cTP	BS
Sevir.4G083100	-0.3563022	0.00544716	AT1G77800.2	PHD finger family protein	noTP	BS
Sevir.4G154500	-0.5139269	0.02107793	AT1G30900.1	VACUOLAR SORTING RECEPTOR 6	SP	BS
Sevir.6G039600	-0.5078763	0.04388092	AT5G54540.1	Uncharacterised conserved protein (UCP012943)	noTP	
Sevir.1G316900	-0.551449	0.01471991	AT1G70670.1	Caleosin-related family protein	SP	M

Sevir.2G296500	-1.2129015	3.03E-05	AT5G24470.1	pseudo-response regulator 5	noTP	M
Sevir.9G525600	-0.5062195	0.00618353	AT5G12040.1	Nitrilase/cyanide hydratase and apolipoprotein N-acyltransferase family protein	cTP	
Sevir.5G391700	-0.796186	0.02814104	AT2G29670.1	Tetratricopeptide repeat (TPR)-like superfamily protein	cTP	
Sevir.6G229700	-0.8800077	0.00062673	AT5G50400.1	purple acid phosphatase 27	SP	M
Sevir.2G424100	-0.4855547	0.03142645	AT2G17930.1	Phosphatidylinositol 3- and 4-kinase family protein with FAT domain	noTP	BS
Sevir.4G033000	-0.3630391	0.02713254	AT1G10500.1	chloroplast-localized ISCA-like protein	cTP	
Sevir.9G301100	-0.5455357	0.02561778	AT3G50670.1	U1 small nuclear ribonucleoprotein-70K	noTP	
Sevir.1G350400	-0.9102306	0.00832801	AT4G33150.2	lysine-ketoglutarate reductase/saccharopine dehydrogenase bifunctional enzyme	noTP	
Sevir.4G152200	-0.3524249	0.0211856	AT4G32850.6	nuclear poly(a) polymerase	noTP	
Sevir.1G379400	-0.9756062	0.03851102	AT3G10470.1	C2H2-type zinc finger family protein	noTP	BS
Sevir.8G214600	-0.3903588	0.04941374	AT4G36690.4	U2 snRNP auxilliary factor, large subunit, splicing factor	noTP	BS
Sevir.2G394400	-0.4897841	0.01471991	AT3G32930.1		cTP	M
Sevir.6G059200	-1.180506	4.57E-08	AT1G64940.1	cytochrome P450, family 87, subfamily A, polypeptide 6	SP	
Sevir.9G241200	-0.5817891	0.00380961	AT5G09880.1	Splicing factor, CC1-like	noTP	BS
Sevir.8G209800	-0.7689159	0.00492384	AT4G37900.1	Protein of unknown function (duplicated DUF1399)	noTP	M
Sevir.1G183600	-1.4826865	0.00013394	AT3G52740.1		noTP	
Sevir.3G258200	-1.2936159	2.19E-05			noTP	M
Sevir.5G474000	-0.78885	0.03672069	AT5G40610.1	NAD-dependent glycerol-3-phosphate dehydrogenase family protein	noTP	M
Sevir.9G177200	-0.8162384	0.04151993	AT3G15840.1	post-illumination chlorophyll fluorescence increase	cTP	BS
Sevir.6G205100	-0.7310954	0.04382525	AT4G34050.1	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein	cTP	BS
Sevir.2G050400	-1.9271627	7.70E-13	AT2G23840.1	HNH endonuclease	cTP	
Sevir.9G037100	-0.9044871	0.00248356	AT5G58490.1	NAD(P)-binding Rossmann-fold superfamily protein	noTP	M
Sevir.7G054600	-0.4416907	0.04563702	AT4G13160.1	Protein of unknown function, DUF593	noTP	BS
Sevir.8G090100	-0.6166367	0.00153352	AT3G07040.1	NB-ARC domain-containing disease resistance protein	noTP	BS
Sevir.8G063000	-0.6782039	0.03016598	AT1G44170.1	aldehyde dehydrogenase 3H1	noTP	M
Sevir.7G073600	-0.6172518	0.01839015	AT2G43710.1	Plant stearoyl-acyl-carrier-protein desaturase family protein	cTP	BS
Sevir.2G079500	-1.813813	5.03E-07	AT5G18130.1		noTP	
Sevir.2G050300	-1.1001503	0.02499886	AT1G55510.1	branched-chain alpha-keto acid decarboxylase E1 beta subunit	mTP	
Sevir.3G125100	-0.6219925	0.02358282	AT5G62740.1	SPFH/Band 7/PHB domain-containing membrane-associated protein family	noTP	M
Sevir.9G112900	-0.4370346	0.0254844	AT3G17770.1	Dihydroxyacetone kinase	noTP	
Sevir.7G195100	-0.9388286	0.0025196	AT5G53550.2	YELLOW STRIPE like 3	noTP	BS
Sevir.4G066700	-2.2445717	0.01264779	AT5G45910.1	GDSL-like Lipase/Acylhydrolase superfamily protein	SP	M
Sevir.9G551500	-0.4268015	0.02687283	AT2G33440.1	RNA-binding (RRM/RBD/RNP motifs) family protein	noTP	BS
Sevir.9G518400	-0.464077	0.00106834	AT2G22540.1	K-box region and MADS-box transcription factor family protein	noTP	
Sevir.9G309000	-0.616355	0.03004842	AT3G10450.1	serine carboxypeptidase-like 7	noTP	BS

Sevir.4G001400	-0.6054306	0.00183289	AT2G30600.2	BTB/POZ domain-containing protein	noTP	
Sevir.7G049500	-0.3478712	0.03397225	AT1G51610.1	Cation efflux family protein	mTP	BS
Sevir.1G066532	-1.2594976	0.01361554	AT3G21150.1	B-box 32	noTP	M
Sevir.9G029000	-0.3316067	0.04256751	AT1G74850.1	plastid transcriptionally active 2	cTP	M
Sevir.8G130300	-0.7149851	3.02E-05	AT1G55280.1	Lipase/lipooxygenase, PLAT/LH2 family protein	cTP	
Sevir.9G495200	-0.6069954	0.01753092	AT3G29390.1	RS2-interacting KH protein	noTP	BS
Sevir.4G143000	-0.3424753	0.01162045	AT1G53650.2	CTC-interacting domain 8	noTP	BS
Sevir.6G015900	-0.5425773	0.03994616	AT1G51550.1	Kelch repeat-containing F-box family protein	noTP	BS
Sevir.6G009500	-0.4444289	0.00044898	AT4G10070.1	KH domain-containing protein	noTP	BS
Sevir.9G387532	-0.4390917	0.01413029	AT4G39680.1	SAP domain-containing protein	noTP	BS
Sevir.6G097100	-0.4857787	0.008517	AT4G35785.2	RNA-binding (RRM/RBD/RNP motifs) family protein	noTP	
Sevir.9G475000	-1.1454374	3.01E-05	AT2G17880.1	Chaperone DnaJ-domain superfamily protein	noTP	BS
Sevir.5G433400	-0.2947693	0.04236222	AT1G24706.1	THO2	noTP	BS
Sevir.5G196800	-0.5929558	0.0235678	AT2G23560.1	methyl esterase 7	noTP	BS
Sevir.5G177866	-2.3191135	3.28E-52	AT2G24960.1		noTP	M
Sevir.3G313500	-0.393743	0.03672069	AT3G46610.1	Pentatricopeptide repeat (PPR-like) superfamily protein	cTP	
Sevir.7G115500	-0.8030492	1.81E-06	AT5G08500.1	Transmembrane CLPTM1 family protein	noTP	BS
Sevir.5G441100	-0.6742744	0.00290011	AT5G43280.1	delta(3,5),delta(2,4)-dienoyl-CoA isomerase 1	noTP	BS
Sevir.9G371200	-0.9626239	0.00090425	AT5G13110.1	glucose-6-phosphate dehydrogenase 2	cTP	
Sevir.7G246500	-0.7933302	5.78E-05	AT5G52230.2	methyl-CPG-binding domain protein 13	noTP	M
Sevir.4G246900	-1.0024707	0.01140669	AT5G56860.1	GATA type zinc finger transcription factor family protein	noTP	
Sevir.3G252200	-0.5168625	0.00363377	AT5G27320.1	alpha/beta-Hydrolases superfamily protein	noTP	BS
Sevir.3G066800	-0.6567911	0.00641196	AT5G59400.1		cTP	
Sevir.2G443300	-1.5198593	0.00205357	AT3G06810.1	acyl-CoA dehydrogenase-related	noTP	M
Sevir.9G277700	-0.5486125	0.04430823	AT4G34260.1	1,2-alpha-L-fucosidases	noTP	
Sevir.2G342300	-0.8541377	0.01273519	AT1G11700.1	Protein of unknown function, DUF584	noTP	M
Sevir.1G185900	-0.8730942	0.00251017	AT5G46930.1	Plant invertase/pectin methylesterase inhibitor superfamily protein	SP	BS
Sevir.9G301900	-1.0703761	0.0071525	AT1G59960.1	NAD(P)-linked oxidoreductase superfamily protein	noTP	BS
Sevir.9G156200	-0.891415	0.0416847	AT5G46180.1	ornithine-delta-aminotransferase	mTP	M
Sevir.5G390300	-0.8738979	0.00015995	AT3G09050.1		cTP	
Sevir.3G133100	-0.8698578	0.0025328	AT5G66010.1	RNA-binding (RRM/RBD/RNP motifs) family protein	noTP	
Sevir.1G040400	-1.1939503	3.02E-05	AT1G80300.1	nucleotide transporter 1	cTP	M
Sevir.9G448700	-0.8513808	0.03356364	AT3G53470.2		noTP	BS
Sevir.5G010800	-0.6308331	0.03828761	AT4G13250.1	NAD(P)-binding Rossmann-fold superfamily protein	cTP	
Sevir.2G305700	-0.6496819	0.03871733	AT1G76150.1	enoyl-CoA hydratase 2	noTP	
Sevir.4G108100	-0.5642665	0.00111619	AT5G14950.1	golgi alpha-mannosidase II	noTP	
Sevir.5G101900	-0.8165377	0.01141347	AT4G29210.1	gamma-glutamyl transpeptidase 4	noTP	
Sevir.7G173300	-0.9547216	0.0154138	AT4G21760.1	beta-glucosidase 47	SP	
Sevir.2G303700	-0.8032954	0.00211628	AT1G20225.1	Thioredoxin superfamily protein	SP	
Sevir.7G153200	-2.5559959	0.03074048	AT5G03860.2	malate synthase	noTP	
Sevir.3G250200	-3.3099062	0.0035401	AT3G47340.1	glutamine-dependent asparagine synthase 1	noTP	
Sevir.7G080200	-1.1610715	8.11E-07	AT4G23130.2	cysteine-rich RLK (RECEPTOR-like protein kinase) 5	noTP	BS

Sevir.6G040300	-0.8679317	0.00848058	AT4G27310.1	B-box type zinc finger family protein	noTP	
Sevir.7G018800	-0.6754239	0.03871733	AT5G43430.1	electron transfer flavoprotein beta	noTP	BS
Sevir.9G500000	-0.2516517	0.00698661	AT1G04080.1	Tetratricopeptide repeat (TPR)-like superfamily protein	noTP	
Sevir.2G446500	-1.1500367	0.02713254	AT5G24910.1	cytochrome P450, family 714, subfamily A, polypeptide 1	noTP	M
Sevir.2G393000	-0.6483507	0.03035634	AT2G37410.1	translocase inner membrane subunit 17-2	noTP	M
Sevir.7G078400	-0.6237101	0.00037893	AT1G27385.1		cTP	M
Sevir.7G080201	-0.8880753	0.00015995	AT3G14460.1	LRR and NB-ARC domains-containing disease resistance protein	noTP	M
Sevir.5G411100	-0.5783181	0.03563695	AT4G10080.1		noTP	M
Sevir.8G026807	-1.5933424	0.00221221	AT3G24190.1	Protein kinase superfamily protein	noTP	
Sevir.3G427300	-0.5325964	0.0009875	AT3G25840.1	Protein kinase superfamily protein	noTP	BS
Sevir.7G092100	-0.5855723	0.00848058			noTP	M
Sevir.9G387466	-0.545108	0.00823763	AT4G39680.2	SAP domain-containing protein	noTP	BS
Sevir.4G232900	-0.9597467	0.00208489	AT3G15730.1	phospholipase D alpha 1	noTP	
Sevir.2G311800	-0.6187396	0.04236222	AT4G35500.1	Protein kinase superfamily protein	noTP	BS
Sevir.6G239025	-0.6403427	0.00270219	AT3G14460.1	LRR and NB-ARC domains-containing disease resistance protein	noTP	
Sevir.2G185900	-0.413152	0.01668165	AT1G44910.1	pre-mRNA-processing protein 40A	noTP	
Sevir.5G034300	-0.4476824	0.01188554	AT3G17611.1	RHOMBOID-like protein 14	noTP	
Sevir.2G100500	-0.5024954	0.03672069	AT1G56130.1	Leucine-rich repeat transmembrane protein kinase	SP	BS
Sevir.1G358000	-1.2718811	0.00590506	AT2G32070.1	Polynucleotidyl transferase, ribonuclease H-like superfamily protein	noTP	
Sevir.9G014000	-0.5431459	0.02561778	AT2G43120.1	RmLC-like cupins superfamily protein	noTP	M
Sevir.2G296000	-1.463323	1.93E-05	AT4G22920.1	non-yellowing 1	cTP	BS
Sevir.7G248700	-0.609473	0.04941374	AT1G56130.1	Leucine-rich repeat transmembrane protein kinase	SP	M
Sevir.4G232601	-1.131306	1.31E-05			noTP	BS
Sevir.7G199400	-0.7697016	0.04940707	AT5G47240.1	nudix hydrolase homolog 8	cTP	M
Sevir.5G236500	-0.3013499	0.04390438	AT5G53180.1	polypyrimidine tract-binding protein 2	noTP	BS
Sevir.6G239100	-0.5121785	0.00048323	AT2G40070.1		noTP	M
Sevir.7G327500	-0.8626304	0.00010479	AT1G74040.1	2-isopropylmalate synthase 1	cTP	M
Sevir.1G377600	-0.952248	0.02814104	AT2G45980.1		noTP	
Sevir.2G279200	-0.7716351	0.0001475	AT5G50850.1	Transketolase family protein	mTP	M
Sevir.9G016200	-0.2910717	0.02813579	AT3G03340.1	LUC7 related protein	noTP	
Sevir.7G209900	-1.0488268	0.00543047	AT1G76140.1	Prolyl oligopeptidase family protein	noTP	BS
Sevir.1G144466	-0.6582019	0.00897833			noTP	BS
Sevir.9G449600	-0.5643632	0.00189553	AT3G55020.1	Ypt/Rab-GAP domain of gyp1p superfamily protein	noTP	
Sevir.9G004300	-0.5001739	0.00276049	AT1G65000.1		noTP	
Sevir.7G005200	-3.9677456	0.00012364	AT2G38900.2	Serine protease inhibitor, potato inhibitor I-type family protein	noTP	M
Sevir.9G158200	-0.8468311	0.00338633	AT1G79050.1	recA DNA recombination family protein	cTP	BS
Sevir.7G323500	-0.3941811	0.00038018	AT5G52560.1	UDP-sugar pyrophosphorylase	noTP	
Sevir.7G163500	-1.0259571	0.00610326	AT1G13080.1	cytochrome P450, family 71, subfamily B, polypeptide 2	SP	M
Sevir.6G195300	-0.4415256	0.03879295	AT2G44730.1	Alcohol dehydrogenase transcription factor Myb/SANT-like family protein	noTP	

Sevir.8G028400	-0.6153575	0.04992661	AT1G74040.1	2-isopropylmalate synthase 1	cTP	M
Sevir.9G320400	-1.0281712	0.00466567	AT2G02040.1	peptide transporter 2	noTP	M
Sevir.8G080600	-0.4230472	0.03869897	AT1G14510.1	alfin-like 7	noTP	
Sevir.2G114200	-0.6088542	0.01497817	AT1G67420.1	Zn-dependent exopeptidases superfamily protein	noTP	M
Sevir.4G237600	-1.0224287	0.00128693	AT1G04250.1	AUX/IAA transcriptional regulator family protein	noTP	BS
Sevir.7G077900	-0.4097367	0.00656848	AT5G51300.2	splicing factor-related	noTP	
Sevir.9G403200	-0.7371641	0.01227212	AT2G41250.1	Haloacid dehalogenase-like hydrolase (HAD) superfamily protein	mTP	BS
Sevir.2G435000	-1.0503816	0.00618353	AT2G47900.3	tubby like protein 3	noTP	M
Sevir.4G032400	-0.6509661	0.0332219	AT5G23940.1	HXXXD-type acyl-transferase family protein	noTP	BS
Sevir.1G347400	-0.880778	0.04321839	AT3G14060.1		noTP	M
Sevir.9G074150	-1.2083636	0.00772198	AT3G13050.1	Major facilitator superfamily protein	noTP	M
Sevir.1G144433	-0.8040404	0.00737781			noTP	BS
Sevir.9G444101	-1.5954961	0.01138011			noTP	BS
Sevir.4G288200	-1.0243919	0.01170586	AT1G31350.1	KAR-UP F-box 1	noTP	BS
Sevir.1G064600	-0.8512901	0.00556175	AT4G30310.2	FGGY family of carbohydrate kinase	noTP	
Sevir.2G283100	-0.3889861	0.03016562	AT1G05910.1	cell division cycle protein 48-related / CDC48-related	noTP	BS
Sevir.3G074200	-0.7686172	0.00208489	AT4G32870.1	Polyketide cyclase/dehydrase and lipid transport superfamily protein	noTP	M
Sevir.5G230700	-0.5309711	0.00302214	AT3G54980.1	Pentatricopeptide repeat (PPR) superfamily protein	noTP	BS
Sevir.2G424700	-0.7046095	0.03356364	AT3G06270.1	Protein phosphatase 2C family protein	noTP	M
Sevir.5G409000	-0.826358	0.00555353	AT1G21410.1	F-box/RNI-like superfamily protein	noTP	BS
Sevir.2G124100	-0.5029667	0.02107793	AT2G20340.1	Pyridoxal phosphate (PLP)-dependent transferases superfamily protein	noTP	M
Sevir.2G234700	-0.5075293	0.00187825	AT4G13420.1	high affinity K+ transporter 5	noTP	BS
Sevir.3G206000	-0.8589393	0.00065051	AT2G38640.1	Protein of unknown function (DUF567)	noTP	
Sevir.7G087600	-0.5906127	0.02063145	AT1G29880.1	glycyl-tRNA synthetase / glycine--tRNA ligase	cTP	
Sevir.1G157300	-0.4852887	0.03887709	AT1G50430.1	Ergosterol biosynthesis ERG4/ERG24 family	noTP	BS
Sevir.9G049300	-0.922515	0.0308617	AT5G40010.1	AAA-ATPase 1	noTP	M
Sevir.6G051400	-0.6031039	0.02741607	AT2G32780.1	ubiquitin-specific protease 1	noTP	BS
Sevir.9G121900	-0.7292137	0.02246479			noTP	BS
Sevir.3G261400	-0.3800023	0.00832674	AT4G17620.1	glycine-rich protein	noTP	BS
Sevir.5G115100	-0.3573291	0.04431043	AT5G01960.1	RING/U-box superfamily protein	noTP	BS
Sevir.3G231600	-1.0589503	2.21E-05	AT1G06620.1	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein	noTP	BS
Sevir.8G027749	-0.4144325	0.0254844	AT1G51190.1	Integrase-type DNA-binding superfamily protein	noTP	BS
Sevir.2G346000	-0.6649831	0.0239918	AT5G54840.1	Ras-related small GTP-binding family protein	noTP	BS
Sevir.3G256600	-0.8351406	0.00183289	AT3G23600.1	alpha/beta-Hydrolases superfamily protein	noTP	
Sevir.6G238000	-0.489367	0.00088069	AT3G59990.4	methionine aminopeptidase 2B	noTP	
Sevir.7G156300	-0.7588521	1.34E-05	AT1G59740.1	Major facilitator superfamily protein	noTP	M
Sevir.5G098000	-0.4216739	0.00044671	AT3G12380.1	actin-related protein 5	noTP	
Sevir.1G043300	-2.4476493	0.00027605	AT5G11260.1	Basic-leucine zipper (bZIP) transcription factor family protein	noTP	M
Sevir.6G035300	-0.5687491	0.0118045	AT1G11650.2	RNA-binding (RRM/RBD/RNP motifs) family protein	noTP	M

Sevir.9G310100	-0.4681121	0.01264779			noTP	BS
Sevir.2G217100	-0.6378999	0.01164943			noTP	
Sevir.7G199625	-0.9039875	0.00817513	AT5G45340.1	cytochrome P450, family 707, subfamily A, polypeptide 3	noTP	M
Sevir.2G335600	-0.6176217	0.04838469	AT1G07250.1	UDP-glucosyl transferase 71C4	noTP	M
Sevir.3G237900	-2.2111289	7.46E-07	AT1G35420.2	alpha/beta-Hydrolases superfamily protein	cTP	M
Sevir.2G330400	-0.463753	0.04785894	AT1G02120.1	GRAM domain family protein	noTP	
Sevir.8G027201	-1.7020652	0.00010479	AT5G54290.2	cytochrome c biogenesis protein family	cTP	M
Sevir.6G074300	-0.4861229	0.03440491	AT1G30540.1	Actin-like ATPase superfamily protein	noTP	M
Sevir.7G215100	-0.5590833	0.017361	AT4G11670.1	Protein of unknown function (DUF810)	noTP	BS
Sevir.8G230028	-7.03767	5.31E-56			SP	
Sevir.5G409400	-1.5133979	0.01196472	AT1G04290.1	Thioesterase superfamily protein	noTP	BS
Sevir.1G183900	-0.8700369	0.02507679	AT3G10130.1	SOUL heme-binding family protein	cTP	M
Sevir.1G089000	-0.815638	0.01019586	AT4G16130.1	arabinose kinase	noTP	BS
Sevir.3G261200	-6.6441136	1.29E-53			noTP	
Sevir.5G475300	-1.2563883	3.37E-05	AT2G39670.1	Radical SAM superfamily protein	cTP	M
Sevir.9G484250	-0.652681	0.00187825			noTP	
Sevir.5G138400	-3.0555043	9.91E-22	AT2G19930.1	RNA-dependent RNA polymerase family protein	noTP	BS
Sevir.5G141601	-1.8792951	7.32E-09	AT1G78600.1	light-regulated zinc finger protein 1	noTP	
Sevir.7G300042	-0.5960459	0.00117877			noTP	BS
Sevir.7G098100	-1.7563616	0.00022057	AT3G52740.1		noTP	
Sevir.9G486800	-0.9955613	0.02535944	AT3G49810.1	ARM repeat superfamily protein	noTP	BS
Sevir.9G196500	-0.9100569	0.00221536	AT3G50670.1	U1 small nuclear ribonucleoprotein-70K	noTP	BS
Sevir.5G182200	-1.2781106	0.02329584	AT1G68300.1	Adenine nucleotide alpha hydrolases-like superfamily protein	noTP	
Sevir.5G005800	-0.806102	0.03826021	AT3G15355.1	ubiquitin-conjugating enzyme 25	noTP	BS
Sevir.2G355700	-0.8484107	0.01279135	AT5G42310.1	Pentatricopeptide repeat (PPR-like) superfamily protein	noTP	BS
Sevir.4G208400	-1.6456117	0.01138011	AT5G54770.1	thiazole biosynthetic enzyme, chloroplast (ARA6) (THI1) (THI4)	cTP	
Sevir.5G044800	-1.1855303	0.01122913	AT2G39040.1	Peroxidase superfamily protein	SP	M
Sevir.1G158200	-2.5610688	5.86E-16			noTP	
Sevir.5G020400	-0.8495292	0.03743638	AT3G06850.1	2-oxoacid dehydrogenases acyltransferase family protein	mTP	BS
Sevir.7G033500	-0.3864549	0.03222776	AT5G15610.2	Proteasome component (PCI) domain protein	noTP	
Sevir.7G173700	-0.9916141	0.00281464	AT4G13670.1	plastid transcriptionally active 5	cTP	BS
Sevir.7G012500	-0.5736947	0.03282433	AT2G01460.1	P-loop containing nucleoside triphosphate hydrolases superfamily protein	noTP	BS
Sevir.4G051000	-0.7427861	0.00208489	AT4G38520.1	Protein phosphatase 2C family protein	noTP	
Sevir.4G006300	-1.4532971	1.32E-08	AT3G61460.1	brassinosteroid-responsive RING-H2	noTP	BS
Sevir.1G142800	-0.903036	0.00466567	AT1G09900.1	Pentatricopeptide repeat (PPR-like) superfamily protein	mTP	BS
Sevir.1G343100	-1.4980938	1.74E-10	AT4G19100.1	Protein of unknown function (DUF3464)	cTP	BS
Sevir.7G300900	-0.5182632	0.04404903			noTP	BS
Sevir.9G446300	-0.3597222	0.02753924	AT5G50840.2		noTP	
Sevir.2G271332	-0.385394	0.04003294			noTP	BS
Sevir.5G169050	-5.003898	3.52E-103	AT3G11290.1		noTP	
Sevir.9G540800	-0.7172131	0.0026417	AT5G09880.1	Splicing factor, CC1-like	noTP	BS

Sevir.5G409500	-0.514436	0.04722263	AT1G71900.1	Protein of unknown function (DUF803)	noTP	M
Sevir.9G540700	-1.0329321	0.00466567	AT5G05320.1	FAD/NAD(P)-binding oxidoreductase family protein	noTP	BS
Sevir.2G239700	-0.8784261	0.01064615	AT1G49720.1	abscisic acid responsive element-binding factor 1	noTP	BS
Sevir.7G074000	-0.785476	0.01917109	AT3G01860.1		noTP	
Sevir.2G433500	-0.6401089	0.01885627	AT1G56110.1	homolog of nucleolar protein NOP56	noTP	BS
Sevir.3G245600	-0.7288179	0.0053492			noTP	M
Sevir.5G156200	-0.8995526	0.00184804	AT1G56670.1	GDSL-like Lipase/Acylhydrolase superfamily protein	SP	
Sevir.5G054800	-0.477676	0.01026083	AT1G75660.1	5\'-3\' exoribonuclease 3	noTP	
Sevir.4G171400	-0.5469135	0.02862891	AT2G17420.1	NADPH-dependent thioredoxin reductase A	mTP	BS
Sevir.3G278200	-0.5265561	0.04686661	AT3G16210.1	F-box family protein	noTP	BS
Sevir.9G531300	-3.0508248	0.01858159	AT5G39530.1	Protein of unknown function (DUF1997)	cTP	BS
Sevir.9G232600	-0.6140483	0.00119424	AT1G08600.3	P-loop containing nucleoside triphosphate hydrolases superfamily protein	noTP	BS
Sevir.3G246000	-1.2490775	0.00493327	AT2G31840.1	Thioredoxin superfamily protein	cTP	
Sevir.3G277766	-0.3575509	0.0486095	AT3G11290.1		noTP	
Sevir.3G322000	-0.6501782	0.02107793	AT3G14470.1	NB-ARC domain-containing disease resistance protein	noTP	M
Sevir.4G249800	-0.9090517	0.04044911	AT1G10970.1	zinc transporter 4 precursor	noTP	M
Sevir.9G425200	-1.1015881	0.00165041	AT5G22340.1		cTP	M
Sevir.7G329100	-0.5079786	0.0071525	AT1G51450.1	TRAUCO	noTP	BS
Sevir.9G581500	-0.4669867	0.01918039	AT4G17360.1	Formyl transferase	noTP	BS
Sevir.8G167100	-0.8003107	0.00607796	AT5G25610.1	BURP domain-containing protein	SP	BS
Sevir.5G410400	-2.8554062	0.01858692	AT4G25420.1	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein	noTP	M
Sevir.5G177200	-1.533337	0.00777879	AT3G02430.1	Protein of unknown function (DUF679)	noTP	
Sevir.4G150500	-0.6903644	0.04431043	AT1G58360.1	amino acid permease 1	noTP	M
Sevir.9G079300	-0.7140853	0.00261298	AT2G23070.1	Protein kinase superfamily protein	cTP	M
Sevir.3G038700	-0.672702	0.00562339	AT2G39090.1	tetratricopeptide repeat (TPR)-containing protein	noTP	M
Sevir.9G461300	-1.5845654	8.35E-05	AT4G37300.1	maternal effect embryo arrest 59	noTP	
Sevir.3G116800	-0.5035927	0.01284307	AT2G25760.1	Protein kinase family protein	noTP	BS
Sevir.9G330400	-0.4310075	0.03440491	AT1G17020.1	senescence-related gene 1	noTP	BS
Sevir.8G209600	-0.6065197	0.01138011	AT3G07040.1	NB-ARC domain-containing disease resistance protein	noTP	BS
Sevir.6G113800	-0.4099333	0.01642585	AT3G03190.1	glutathione S-transferase F11	noTP	BS
Sevir.9G176600	-0.6941096	0.01593855	AT1G50575.1	Putative lysine decarboxylase family protein	cTP	M
Sevir.8G168500	-0.4739681	0.03629459	AT5G22010.1	replication factor C1	noTP	M
Sevir.5G153600	-1.3103362	0.00012018	AT1G23400.1	RNA-binding CRS1 / YhbY (CRM) domain-containing protein	noTP	BS
Sevir.2G265700	-0.7580924	0.04430823	AT3G57980.1	DNA-binding bromodomain-containing protein	noTP	
Sevir.2G446300	-0.6268743	0.00404498	AT2G40770.1	zinc ion binding;DNA binding;helicases;ATP binding;nucleic acid binding	noTP	
Sevir.3G272800	-0.5456447	0.03096201			noTP	
Sevir.5G262400	-1.3798569	3.27E-07	AT5G62840.1	Phosphoglycerate mutase family protein	noTP	M
Sevir.5G047300	-0.8192717	0.00446782	AT5G11720.1	Glycosyl hydrolases family 31 protein	noTP	M
Sevir.3G039700	-1.5411486	0.00033021	AT3G62760.1	Glutathione S-transferase family protein	noTP	M

Sevir.3G368300	-0.4608557	0.02851099	AT4G24460.1	CRT (chloroquine-resistance transporter)-like transporter 2	noTP	BS
Sevir.9G338200	-0.5146709	0.01164943	AT1G76630.2	Tetratricopeptide repeat (TPR)-like superfamily protein	SP	BS
Sevir.2G033300	-0.3725722	0.0463757	AT3G18100.1	myb domain protein 4r1	noTP	
Sevir.9G319766	-2.5126707	0.00055804			noTP	BS
Sevir.8G000101	-1.1896536	1.36E-08			noTP	
Sevir.4G285800	-0.5046381	0.02741607	AT3G02410.1	alpha/beta-Hydrolases superfamily protein	noTP	
Sevir.5G379900	-0.5622533	0.01473419	AT3G46060.3	RAB GTPase homolog 8A	noTP	
Sevir.2G293100	-1.5351905	0.00070387	AT4G27700.1	Rhodanese/Cell cycle control phosphatase superfamily protein	cTP	
Sevir.7G157200	-0.9559282	0.00453959	AT3G59410.1	protein kinase family protein	noTP	BS
Sevir.3G252301	-0.5957763	4.17E-05	AT4G03430.1	pre-mRNA splicing factor-related	noTP	M
Sevir.4G255900	-1.0467355	0.03914805	AT5G57180.2	chloroplast import apparatus 2	noTP	M
Sevir.3G331300	-0.5813141	0.02499886	AT3G20420.1	RNAse THREE-like protein 2	noTP	BS
Sevir.5G119500	-1.0913865	0.04781534	AT5G11260.1	Basic-leucine zipper (bZIP) transcription factor family protein	noTP	
Sevir.5G060640	-0.553324	0.00195571			noTP	
Sevir.6G193200	-1.1529152	0.02591865	AT4G09040.1	RNA-binding (RRM/RBD/RNP motifs) family protein	cTP	M
Sevir.4G275500	-0.584952	0.01690638	AT4G20930.1	6-phosphogluconate dehydrogenase family protein	mTP	
Sevir.5G379400	-1.6421024	0.04077548	AT5G02580.1	Plant protein 1589 of unknown function	noTP	BS
Sevir.1G300400	-0.6201332	0.03222776	AT3G51520.1	diacylglycerol acyltransferase family	noTP	
Sevir.9G125900	-0.5345882	0.03878735	AT4G02725.1		cTP	
Sevir.4G104900	-0.6580807	0.03081755	AT1G10510.1	RNI-like superfamily protein	cTP	
Sevir.5G138500	-2.2892012	5.30E-30	AT2G19930.1	RNA-dependent RNA polymerase family protein	noTP	BS
Sevir.4G056600	-0.5655061	0.03890995	AT3G01300.1	Protein kinase superfamily protein	noTP	BS
Sevir.6G155300	-1.9372386	0.00492384	AT1G60010.1		noTP	BS
Sevir.7G326600	-1.3734127	0.00582057	AT5G54290.2	cytochrome c biogenesis protein family	cTP	M
Sevir.8G260500	-0.5843584	0.03332284	AT2G29050.1	RHOMBOID-like 1	noTP	BS
Sevir.5G137500	-1.4128518	7.63E-06	AT1G55140.1	Ribonuclease III family protein	cTP	M
Sevir.9G481700	-0.5427855	0.02281631	AT2G32840.1	proline-rich family protein	noTP	M
Sevir.2G253900	-0.901236	0.00187825	AT3G13180.1	NOL1/NOP2/sun family protein / antitermination NusB domain-containing protein	noTP	M
Sevir.4G118903	-0.6813595	5.39E-05	AT5G10240.1	asparagine synthetase 3	noTP	BS
Sevir.3G399400	-0.5501146	0.03705571	AT1G10570.1	Cysteine proteinases superfamily protein	noTP	BS
Sevir.5G321500	-1.0711635	0.00138626	AT5G45710.1	winged-helix DNA-binding transcription factor family protein	noTP	
Sevir.4G058900	-0.7873814	0.03629459	AT5G52010.1	C2H2-like zinc finger protein	cTP	
Sevir.5G013700	-1.0276545	5.78E-05	AT1G79900.1	Mitochondrial substrate carrier family protein	noTP	BS
Sevir.3G226400	-0.7727162	0.02165174	AT5G05930.1	guanylyl cyclase 1	noTP	M
Sevir.3G306000	-0.6187737	0.03478193	AT3G07230.1	wound-responsive protein-related	noTP	M
Sevir.6G239500	-0.6244864	0.01741642	AT3G14460.1	LRR and NB-ARC domains-containing disease resistance protein	noTP	M
Sevir.9G027100	-1.1433014	0.00044725	AT5G01470.2	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein	noTP	M
Sevir.1G363700	-0.5362703	0.04842269	AT5G61020.2	evolutionarily conserved C-terminal region 3	noTP	BS

Sevir.5G307400	-0.7861329	0.03142645	AT1G78620.1	Protein of unknown function DUF92, transmembrane	cTP	
Sevir.5G435900	-0.6918769	0.03879295	AT3G27180.1	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein	noTP	M
Sevir.7G290500	-0.5015886	0.02598526	AT5G25800.1	Polynucleotidyl transferase, ribonuclease H-like superfamily protein	noTP	
Sevir.5G286900	-2.437498	0.0024265	AT4G26890.1	mitogen-activated protein kinase kinase kinase 16	noTP	
Sevir.1G064300	-1.161288	0.03851102	AT2G24762.1	glutamine dumper 4	noTP	BS
Sevir.2G323100	-0.9602248	0.0087444	AT4G35780.1	ACT-like protein tyrosine kinase family protein	noTP	BS
Sevir.9G160900	-1.4237318	0.00075569	AT5G28750.1	Bacterial sec-independent translocation protein mttA/Hcf106	cTP	M
Sevir.7G223750	-0.6393171	0.04069036			noTP	BS
Sevir.5G177100	-0.5166737	0.00848058	AT1G72390.1		noTP	BS
Sevir.2G119700	-1.3833026	0.00040962	AT1G67580.2	Protein kinase superfamily protein	noTP	
Sevir.6G167700	-0.7135806	0.01974789	AT5G11000.1	Plant protein of unknown function (DUF868)	noTP	
Sevir.9G237400	-0.7147407	0.03651899	AT3G04550.1		cTP	BS
Sevir.3G198900	-0.5676065	0.04321839	AT3G12480.1	nuclear factor Y, subunit C11	noTP	BS
Sevir.1G001500	-0.5581453	0.03222776	AT1G20060.1	ATP binding microtubule motor family protein	noTP	BS
Sevir.8G039300	-0.9779054	0.01490126	AT1G06140.1	Pentatricopeptide repeat (PPR) superfamily protein	mTP	
Sevir.7G227700	-0.7456175	0.01289053	AT1G44750.2	purine permease 11	noTP	M
Sevir.5G468000	-0.8689579	0.01165687	AT1G34770.1		noTP	
Sevir.9G411700	-0.4744704	0.02573574	AT4G15880.1	Cysteine proteinases superfamily protein	noTP	
Sevir.1G241200	-0.5012018	0.03871733	AT2G20490.1	nucleolar RNA-binding Nop10p family protein	noTP	
Sevir.9G436500	-0.9532435	0.01167906			noTP	BS
Sevir.5G369100	-2.160369	0.00111619	AT2G45260.1	Plant protein of unknown function (DUF641)	noTP	
Sevir.9G502500	-0.7082085	8.53E-05			noTP	M
Sevir.3G383000	-1.3320251	0.00937375	AT3G19553.1	Amino acid permease family protein	noTP	BS
Sevir.9G094601	-0.615536	0.02597041			noTP	M
Sevir.5G126800	-0.9484183	0.0308618	AT2G36960.3	TSL-kinase interacting protein 1	noTP	
Sevir.5G235700	-0.5832979	0.03102076	AT5G18980.1	ARM repeat superfamily protein	noTP	BS
Sevir.7G057200	-0.4899644	0.04029646	AT1G05840.1	Eukaryotic aspartyl protease family protein	SP	
Sevir.8G028650	-0.6945128	0.01338772	AT3G57080.1	Eukaryotic rpb5 RNA polymerase subunit family protein	noTP	M
Sevir.3G038300	-0.7769222	0.01332492	AT5G27680.1	RECQ helicase SIM	noTP	
Sevir.2G305400	-0.6501933	0.04256751	AT3G02910.1	AIG2-like (avirulence induced gene) family protein	cTP	BS
Sevir.4G122300	-1.8162494	3.11E-07			noTP	BS
Sevir.7G253800	-0.7338717	0.01113588			noTP	BS
Sevir.5G305300	-1.2478178	0.00079141	AT5G11580.1	Regulator of chromosome condensation (RCC1) family protein	noTP	BS
Sevir.7G185100	-1.0664137	0.02107793	AT1G28150.1		cTP	M
Sevir.5G333500	-0.6850569	0.02493055	AT4G37560.1	Acetamidase/Formamidase family protein	noTP	M
Sevir.3G178600	-1.0802826	0.03356364	AT3G16360.2	HPT phosphotransmitter 4	noTP	BS
Sevir.7G237200	-0.6593172	0.03611582	AT2G38250.1	Homeodomain-like superfamily protein	noTP	M
Sevir.4G009400	-0.7653977	0.01813566	AT5G27270.1	Tetratricopeptide repeat (TPR)-like superfamily protein	noTP	M
Sevir.9G090700	-0.9437278	0.02828091	AT5G37480.1		noTP	
Sevir.2G047000	-0.634418	0.01845563	AT1G04580.1	aldehyde oxidase 4	noTP	BS

Sevir.7G330700	-0.9178028	0.01880295	AT4G28706.3	pfkB-like carbohydrate kinase family protein	noTP	
Sevir.9G515300	-0.9730914	0.01927257	AT1G22440.1	Zinc-binding alcohol dehydrogenase family protein	noTP	
Sevir.3G320000	-0.7059142	0.04908699	AT5G66930.2		noTP	
Sevir.9G389400	-0.5253173	0.0450756	AT4G17910.1	transferases, transferring acyl groups	noTP	BS
Sevir.5G193500	-1.1692634	1.86E-07	AT2G24960.1		noTP	
Sevir.8G057200	-3.6925974	4.06E-06			noTP	BS
Sevir.3G138900	-1.9922415	0.00897833	AT2G25625.1		noTP	BS
Sevir.8G132200	-3.2495023	1.16E-27			noTP	M
Sevir.6G155500	-1.0974757	0.04941374	AT4G00980.1	zinc knuckle (CCHC-type) family protein	noTP	
Sevir.3G041200	-0.774561	0.00419703	AT5G17930.1	MIF4G domain-containing protein / MA3 domain-containing protein	noTP	
Sevir.3G027201	-1.0171454	0.00748364			noTP	M
Sevir.4G078800	-1.2854529	0.00414375	AT4G25200.1	mitochondrion-localized small heat shock protein 23.6	noTP	M
Sevir.5G360300	-0.6098836	0.04390438	AT5G04940.1	SU(VAR)3-9 homolog 1	noTP	BS
Sevir.7G248000	-1.0692134	0.00276049	AT1G64810.2	Arabidopsis thaliana protein of unknown function (DUF794)	noTP	
Sevir.5G070700	-0.7791205	0.00447454			noTP	M
Sevir.9G203900	-0.9700729	0.00449869	AT3G51000.1	alpha/beta-Hydrolases superfamily protein	noTP	M
Sevir.7G199800	-0.8343614	0.0215605	AT4G17486.1	PPPDE putative thiol peptidase family protein	noTP	M
Sevir.7G051200	-1.2744417	0.00111619	AT2G26970.1	Polynucleotidyl transferase, ribonuclease H-like superfamily protein	noTP	
Sevir.9G371300	-0.8162424	0.02860534	AT4G26630.1	DEK domain-containing chromatin associated protein	noTP	BS
Sevir.3G390500	-1.2640321	0.02803298	AT3G18950.1	Transducin/WD40 repeat-like superfamily protein	noTP	
Sevir.1G273500	-0.6912971	0.0109613	AT1G09290.1		noTP	
Sevir.5G018800	-1.0076519	0.01550592	AT5G36000.1		noTP	
Sevir.3G419200	-1.0656221	0.00110435	AT5G49555.1	FAD/NAD(P)-binding oxidoreductase family protein	mTP	
Sevir.5G251300	-1.1763286	0.00034322	AT1G11290.1	Pentatricopeptide repeat (PPR) superfamily protein	noTP	
Sevir.2G116200	-1.5214058	0.02103792	AT4G15610.1	Uncharacterised protein family (UPF0497)	noTP	M
Sevir.5G386400	-1.2412718	0.00022019	AT3G46790.1	Tetratricopeptide repeat (TPR)-like superfamily protein	cTP	BS
Sevir.2G044400	-0.8342357	0.03763076	AT5G06570.2	alpha/beta-Hydrolases superfamily protein	noTP	BS
Sevir.6G136108	-0.6824079	0.01170523			noTP	BS
Sevir.1G001000	-1.0924815	0.00109613	AT2G39130.1	Transmembrane amino acid transporter family protein	noTP	
Sevir.6G078750	-0.8874643	0.0025196			noTP	
Sevir.9G276201	-3.2556091	6.32E-17			noTP	
Sevir.3G057100	-0.5685927	0.04842269	AT5G52070.1	Agenet domain-containing protein	noTP	BS
Sevir.2G438600	-1.3648311	0.00381731	AT4G15770.1	RNA binding	noTP	M
Sevir.3G384700	-0.7294542	0.01140669	AT1G47240.1	NRAMP metal ion transporter 2	noTP	
Sevir.2G343200	-1.5177577	0.00021323	AT4G21870.1	HSP20-like chaperones superfamily protein	noTP	BS
Sevir.9G309800	-6.7987084	3.21E-13	AT3G53480.1	pleiotropic drug resistance 9	noTP	BS
Sevir.1G014600	-1.5614794	0.02043658			noTP	M
Sevir.2G206100	-0.5793621	0.04928563	AT2G01730.1	cleavage and polyadenylation specificity factor 73 kDa subunit-II	noTP	
Sevir.1G113900	-0.9956476	0.02955371	AT5G09920.1	RNA polymerase II, Rpb4, core protein	noTP	
Sevir.6G089200	-1.5131777	0.03081755	AT4G33420.1	Peroxidase superfamily protein	SP	BS

Sevir.3G145100	-0.9051481	0.02813579	AT1G75220.1	Major facilitator superfamily protein	noTP	
Sevir.5G387000	-0.9735101	0.02279625	AT5G59080.1		noTP	BS
Sevir.2G152300	-1.0234473	0.01264779	AT5G25120.1	ytochrome p450, family 71, subfamily B, polypeptide 11	SP	M
Sevir.8G158300	-3.8671914	3.26E-07			noTP	
Sevir.2G042200	-1.0684914	0.00247485	AT5G22050.2	Protein kinase superfamily protein	mTP	BS
Sevir.7G173000	-0.8368318	0.00573073	AT1G61820.1	beta glucosidase 46	SP	BS
Sevir.3G139000	-3.0581345	0.03356364	AT2G25625.1		cTP	BS
Sevir.7G090400	-0.7022696	0.03779765	AT3G26680.2	DNA repair metallo-beta-lactamase family protein	noTP	
Sevir.3G055600	-0.9180087	0.00823763	AT1G08530.1		luTP	
Sevir.3G014900	-0.9863285	5.53E-05	AT2G20740.1	Tetraspanin family protein	noTP	BS
Sevir.9G306900	-1.7550614	0.02803298			noTP	BS
Sevir.1G181100	-0.9440875	0.04293335	AT2G27830.1		noTP	M
Sevir.5G267101	-0.7023965	0.01956661	AT5G27260.1		noTP	
Sevir.9G259600	-3.1046294	0.01279135	AT2G27830.1		noTP	M
Sevir.8G234701	-2.2054959	9.84E-13	AT1G61360.1	S-locus lectin protein kinase family protein	mTP	M
Sevir.9G352000	-1.8371553	0.00176169	AT1G10360.1	glutathione S-transferase TAU 18	noTP	M
Sevir.6G004300	-0.917621	0.02892715	AT1G22030.1		noTP	BS
Sevir.4G055700	-0.7253671	0.02160335	AT1G48360.2	zinc ion binding;nucleic acid binding;hydrolases, acting on acid anhydrides, in phosphorus-containing anhydrides	noTP	
Sevir.2G272100	-0.8343659	0.01167906	AT4G26965.1	NADH:ubiquinone oxidoreductase, 17.2kDa subunit	noTP	M
Sevir.9G083101	-0.9885604	0.0025196	AT4G38180.1	FAR1-related sequence 5	noTP	M
Sevir.5G088800	-1.7860828	2.90E-05	AT5G38290.2	Peptidyl-tRNA hydrolase family protein	cTP	M
Sevir.2G118800	-0.759505	0.02352781	AT5G60340.1	P-loop containing nucleoside triphosphate hydrolases superfamily protein	noTP	M
Sevir.8G235200	-3.8653886	6.32E-17	AT1G61390.1	S-locus lectin protein kinase family protein	noTP	M
Sevir.2G411500	-0.7950002	0.01145562	AT1G14350.2	Duplicated homeodomain-like superfamily protein	noTP	BS
Sevir.2G348200	-0.8215938	0.03697524	AT1G11545.1	xyloglucan endotransglucosylase/hydrolase 8	SP	BS
Sevir.1G069200	-1.0598696	0.04271793	AT4G15110.1	cytochrome P450, family 97, subfamily B, polypeptide 3	cTP	
Sevir.9G228300	-1.3768868	0.00126669	AT1G04500.1	CCT motif family protein	noTP	BS
Sevir.4G205902	-0.7658684	0.0411738			noTP	
Sevir.2G138900	-0.8302233	0.00748162	AT5G20130.1		cTP	M
Sevir.8G247040	-1.1978886	0.04996957	AT4G36690.4	U2 snRNP auxilliary factor, large subunit, splicing factor	noTP	BS
Sevir.4G082500	-1.0642428	0.03208732	AT1G44800.1	nodulin MtN21 /EamA-like transporter family protein	noTP	BS
Sevir.9G352100	-1.5569807	0.02508546	AT3G01910.1	sulfite oxidase	mTP	M
Sevir.1G121100	-0.7214893	0.03343331	AT4G29010.1	Enoyl-CoA hydratase/isomerase family	noTP	BS
Sevir.8G001900	-1.0899961	0.00112486	AT5G60410.2	DNA-binding protein with MIZ/SP-RING zinc finger, PHD-finger and SAP domain	noTP	
Sevir.7G080302	-1.2254045	1.64E-05	AT3G53900.2	uracil phosphoribosyltransferase	noTP	
Sevir.9G545700	-0.8111538	0.00466567	AT2G27970.1	CDK-subunit 2	noTP	
Sevir.8G010225	-0.6881205	0.04164058	AT3G61790.1	Protein with RING/U-box and TRAF-like domains	noTP	
Sevir.3G226100	-1.1398444	0.03879295	AT5G02010.1	RHO guanyl-nucleotide exchange factor 7	noTP	BS
Sevir.1G127100	-0.9100226	0.0169993	AT4G19210.1	RNAse l inhibitor protein 2	noTP	BS

Sevir.5G424100	-0.9565264	0.02861489	AT5G62410.1	structural maintenance of chromosomes 2	noTP	BS
Sevir.3G327400	-1.5546763	0.04302355	AT3G46530.1	NB-ARC domain-containing disease resistance protein	noTP	BS
Sevir.1G312800	-1.1221176	0.04167221	AT3G48000.1	aldehyde dehydrogenase 2B4	mTP	BS
Sevir.7G124600	-0.7826852	0.02561778	AT1G14140.1	Mitochondrial substrate carrier family protein	noTP	
Sevir.5G333600	-0.9179907	0.01164943	AT4G21065.1	Tetratricopeptide repeat (TPR)-like superfamily protein	noTP	
Sevir.9G027200	-4.3831632	1.76E-05	AT1G19530.1		noTP	
Sevir.7G269700	-1.1383952	0.03887709			noTP	BS
Sevir.3G417600	-3.0409759	0.04785894	AT4G11650.1	osmotin 34	SP	M
Sevir.9G410450	-1.4214777	0.0118045			noTP	
Sevir.9G290301	-1.0952373	0.00644667			cTP	BS
Sevir.3G093600	-1.5408242	0.00152334	AT1G05894.1		noTP	BS
Sevir.9G247900	-1.8725535	8.06E-05			SP	M
Sevir.9G352200	-2.7510857	0.01360067	AT1G26690.1	emp24/gp25L/p24 family/GOLD family protein	SP	M
Sevir.7G004900	-3.4337463	0.00029637	AT2G38900.2	Serine protease inhibitor, potato inhibitor I-type family protein	noTP	M
Sevir.1G227400	-0.8010118	0.04473016	AT3G25250.1	AGC (cAMP-dependent, cGMP-dependent and protein kinase C) kinase family protein	noTP	BS
Sevir.1G257900	-0.9293752	0.02426315	AT2G46680.1	homeobox 7	noTP	
Sevir.2G084700	-0.9029451	0.03395568	AT4G36790.1	Major facilitator superfamily protein	noTP	BS
Sevir.7G080100	-2.0945486	0.03842914			noTP	
Sevir.7G070800	-0.894017	0.01586506	AT5G52450.1	MATE efflux family protein	noTP	M
Sevir.5G396300	-1.993204	0.00013394	AT4G21970.1	Protein of unknown function, DUF584	noTP	M
Sevir.8G158400	-1.1635077	0.03300847	AT3G47630.2		noTP	M
Sevir.9G443200	-0.9502942	0.02358282	AT3G58630.1	sequence-specific DNA binding transcription factors	noTP	BS
Sevir.9G037200	-1.6920732	2.35E-05			cTP	
Sevir.9G567400	-6.0775202	0.00026189	AT1G65910.1	NAC domain containing protein 28	noTP	BS
Sevir.7G269400	-0.7764845	0.02499886			noTP	
Sevir.9G309100	-0.9493132	0.02741607	AT5G09640.1	serine carboxypeptidase-like 19	SP	BS
Sevir.3G234600	-0.9210914	0.03440491	AT1G44414.1		noTP	
Sevir.5G351801	-4.264738	2.46E-11	AT5G11940.1	Subtilase family protein	SP	
Sevir.2G027700	-1.5685421	4.04E-05	AT1G13590.1	phytosulfokine 1 precursor	noTP	
Sevir.9G319500	-8.2865904	0.00062143			noTP	
Sevir.9G231100	-1.6821413	0.00579524			noTP	M
Sevir.4G067000	-1.3654552	0.03724123	AT5G45920.1	SGNH hydrolase-type esterase superfamily protein	noTP	M
Sevir.2G111800	-1.6893904	0.03713151	AT2G18680.1		noTP	
Sevir.9G419800	-1.0781947	0.02408001	AT2G37975.1	Yos1-like protein	SP	
Sevir.9G114000	-0.9860945	0.03887709	AT4G38180.1	FAR1-related sequence 5	noTP	BS
Sevir.2G265400	-1.3597107	0.00931753	AT2G42190.1		noTP	BS
Sevir.5G266401	-1.6956807	0.01672887			noTP	M
Sevir.9G439701	-1.2353548	0.01668165	AT3G10985.1	senescence associated gene 20	noTP	M
Sevir.6G234300	-1.5610625	0.01402074	AT5G61890.1	Integrase-type DNA-binding superfamily protein	noTP	
Sevir.2G165333	-6.048754	0.00247485	AT5G05490.1	Rad21/Rec8-like family protein	noTP	
Sevir.1G181200	-1.4174459	0.00302214	AT5G03720.1	heat shock transcription factor A3	noTP	M
Sevir.7G063400	-1.3023821	0.03222776	AT1G15960.1	NRAMP metal ion transporter 6	noTP	M
Sevir.8G017700	-1.2465956	0.02408001	AT1G07060.1		noTP	
Sevir.2G439200	-2.8883958	0.00094196	AT2G47750.1	putative indole-3-acetic acid-amido synthetase GH3.9	noTP	BS
Sevir.9G557700	-1.0762198	0.03792222	AT5G66005.3	Expressed protein	noTP	
Sevir.7G042320	-2.8755508	0.00421875			noTP	
Sevir.4G212800	-1.7594002	0.04996957	AT1G13080.1	cytochrome P450, family 71, subfamily B, polypeptide 2	noTP	M
Sevir.9G432350	-6.794076	0.00051323			noTP	

Sevir.2G431750	-1.9926857	0.01938738			noTP	M
Sevir.4G025700	-2.5157111	0.02741607	AT1G75580.1	SAUR-like auxin-responsive protein family	noTP	BS
Sevir.2G212900	-2.7458468	5.53E-05			noTP	M
Sevir.8G202400	-1.3145527	0.04941374			noTP	
Sevir.9G299500	-6.7732467	0.00848058	AT5G58490.1	NAD(P)-binding Rossmann-fold superfamily protein	noTP	BS
Sevir.1G089650	-27.660166	1.22E-09	AT2G07050.1	cycloartenol synthase 1	noTP	
Sevir.4G059600	-1.4538605	0.02828091	AT4G36930.1	basic helix-loop-helix (bHLH) DNA-binding superfamily protein	noTP	
Sevir.2G257400	-1.4171213	0.02197618	AT1G55760.1	BTB/POZ domain-containing protein	noTP	
Sevir.2G254701	-3.835964	0.00654695			noTP	M
Sevir.5G069600	-2.2400025	0.00038018	AT1G75280.1	NmrA-like negative transcriptional regulator family protein	noTP	BS
Sevir.7G045400	-1.6097407	0.04029646	AT2G34410.1	O-acetyltransferase family protein	SP	BS
Sevir.9G226400	-1.2680422	0.03043451	AT5G46040.1	Major facilitator superfamily protein	noTP	BS
Sevir.2G008800	-1.5002729	0.03441146	AT5G45540.1	Protein of unknown function (DUF594)	noTP	
Sevir.5G251400	-1.8136173	0.02512457			noTP	BS
Sevir.1G121200	-1.2788252	0.03059183			noTP	BS
Sevir.9G119400	-2.0234196	0.01474588	AT2G02850.1	plantacyanin	SP	M
Sevir.9G201733	-1.6582638	0.04431043			noTP	M
Sevir.2G452200	-1.6040423	0.02020379	AT4G40070.1	RING/U-box superfamily protein	SP	
Sevir.9G163100	-1.5644283	0.03356364	AT4G34370.1	RING/U-box superfamily protein	noTP	BS
Sevir.8G070401	-2.901025	0.0082858			noTP	M
Sevir.9G160601	-1.2922294	0.04992661	AT3G22942.1	G-protein gamma subunit 2	noTP	BS
Sevir.2G166000	-1.7224597	0.0486095	AT4G22320.1		noTP	M
Sevir.4G043700	-1.9711273	0.00848058	AT1G30550.1	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein	noTP	BS
Sevir.9G409100	-2.1105385	0.00498576	AT5G19830.1	Peptidyl-tRNA hydrolase family protein	mTP	BS
Sevir.9G077966	-1.8220076	0.00668869			SP	BS
Sevir.4G236100	-2.3152116	0.00695198	AT2G26710.1	Cytochrome P450 superfamily protein	noTP	BS
Sevir.5G465400	-6.6099775	3.37E-05	AT5G25140.1	cytochrome P450, family 71, subfamily B, polypeptide 13	SP	
Sevir.4G176601	-2.205612	0.01167906			noTP	M
Sevir.7G320200	-1.9181553	0.01428823			noTP	
Sevir.6G108260	-1.5231837	0.04075194			SP	
Sevir.1G025900	-1.5955948	0.01762442			noTP	
Sevir.3G246550	-2.3361164	0.04430823			noTP	BS
Sevir.3G072500	-2.1238866	0.03563695	AT1G76680.1	12-oxophytodienoate reductase 1	noTP	
Sevir.8G080700	-4.0835279	0.04016398			noTP	BS
Sevir.8G031033	-6.052985	0.00241069			SP	
Sevir.5G175300	-4.890549	0.02079323	AT2G14610.1	pathogenesis-related gene 1	SP	
Sevir.8G225400	-2.9548498	0.01361554			noTP	
Sevir.4G098900	-4.3570979	0.0112053	AT1G74460.1	GDSL-like Lipase/Acylhydrolase superfamily protein	SP	

Table A.84: Significantly downregulated genes in *HATCH1* knock-out compared to wild-type plants

Gene	log2Fold Change	p_{adj}	Arabidopsis accession	Arabidopsis gene function	Localisation	Cell type

Sevir.2G188500	0.60293517	0.03914805	AT2G25737.1	Sulfite exporter TauE/SafE family protein	SP	M
Sevir.7G105000	0.43212344	0.00222677			noTP	M
Sevir.9G512700	0.98562624	8.06E-05	AT2G22240.1	myo-inositol-1-phosphate synthase 2	noTP	M
Sevir.6G090300	0.49393937	0.00446782	AT1G16880.1	uridylyltransferase-related	cTP	BS
Sevir.8G033000	1.38798138	0.03144587	AT2G03200.1	Eukaryotic aspartyl protease family protein	SP	M
Sevir.1G195800	0.40518137	0.02598526	AT1G77490.1	thylakoidal ascorbate peroxidase	cTP	M
Sevir.4G248000	0.29456922	0.00556289	AT4G38510.5	ATPase, V1 complex, subunit B protein	noTP	
Sevir.4G038200	1.08542119	0.01918039	AT2G03200.1	Eukaryotic aspartyl protease family protein	SP	M
Sevir.6G113600	0.81217374	0.02598526	AT1G15980.1	NDH-dependent cyclic electron flow 1	cTP	BS
Sevir.3G423900	0.73045621	0.03655927	AT2G16430.2	purple acid phosphatase 10	SP	BS
Sevir.9G580400	0.29091467	8.06E-05	AT4G30600.1	signal recognition particle receptor alpha subunit family protein	noTP	
Sevir.3G072600	0.67370545	0.02098538	AT5G49810.1	methionine S-methyltransferase	noTP	BS
Sevir.3G021700	1.49757585	0.00030206	AT1G65840.1	polyamine oxidase 4	noTP	M
Sevir.2G447900	0.67773172	0.03563695	AT3G51950.2	Zinc finger (CCCH-type) family protein / RNA recognition motif (RRM)-containing protein	noTP	M
Sevir.9G520900	0.89168963	0.04351231	AT1G72360.3	Integrase-type DNA-binding superfamily protein	noTP	M
Sevir.4G287300	2.44723369	6.47E-15	AT4G19390.1	Uncharacterised protein family (UPF0114)	mTP	BS
Sevir.6G168700	0.66888349	0.00377338	AT2G24270.4	aldehyde dehydrogenase 11A3	noTP	
Sevir.5G036400	0.45186447	0.04321839			noTP	M
Sevir.9G272600	0.5077195	0.03193251	AT1G36160.2	acetyl-CoA carboxylase 1	noTP	BS
Sevir.1G036000	0.86167452	0.03407525			noTP	M
Sevir.2G055400	1.25955636	0.04941374	AT3G24480.1	Leucine-rich repeat (LRR) family protein	SP	M
Sevir.9G362400	0.33997924	0.00404498	AT5G21090.1	Leucine-rich repeat (LRR) family protein	SP	M
Sevir.2G454100	0.21364049	0.04941374	AT4G29040.1	regulatory particle AAA-ATPase 2A	noTP	
Sevir.3G361600	0.74286541	0.00819176	AT1G54630.1	acyl carrier protein 3	cTP	M
Sevir.9G405800	1.20264105	0.04302355	AT2G41430.1	dehydration-induced protein (ERD15)	noTP	
Sevir.7G279200	0.37552284	0.01279135	AT3G06720.2	importin alpha isoform 1	noTP	
Sevir.5G058100	0.6637876	0.00013447	AT3G16830.1	TOPLESS-related 2	noTP	BS
Sevir.9G084100	0.70795365	0.04044911	AT3G29360.1	UDP-glucose 6-dehydrogenase family protein	noTP	M
Sevir.9G023400	0.30438587	0.03193251	AT4G15940.1	Fumarylacetoacetate (FAA) hydrolase family	mTP	
Sevir.4G293700	0.80015005	0.02512457	AT4G24190.1	Chaperone protein htpG family protein	SP	BS
Sevir.3G026900	0.16709172	0.04431043	AT5G05780.1	RP non-ATPase subunit 8A	noTP	
Sevir.4G217300	0.86036566	0.00695459	AT5G38820.1	Transmembrane amino acid transporter family protein	noTP	
Sevir.7G225000	0.83834506	0.01162045	AT4G35310.1	calmodulin-domain protein kinase 5	noTP	M
Sevir.1G089900	0.23513625	0.03562518	AT2G07050.1	cycloartenol synthase 1	noTP	
Sevir.9G316400	1.01765111	0.00420246	AT4G17880.1	Basic helix-loop-helix (bHLH) DNA-binding family protein	noTP	
Sevir.2G396500	0.47408565	0.0089491	AT2G35840.2	Sucrose-6F-phosphate phosphohydrolase family protein	noTP	
Sevir.9G238300	1.00532951	0.02507679	AT2G38940.1	phosphate transporter 1;4	noTP	
Sevir.9G171900	0.7323268	0.01918039	AT5G15650.1	reversibly glycosylated polypeptide 2	noTP	M
Sevir.5G087500	0.46351153	0.03049782	AT1G13360.1	noTP		
Sevir.8G144600	0.74591614	0.03440491			noTP	M

Sevir.5G289200	0.52888153	0.01167906	AT5G47770.1	farnesyl diphosphate synthase 1	noTP	M
Sevir.6G033700	1.01116176	0.00190576	AT5G65630.1	global transcription factor group E7	noTP	M
Sevir.2G383700	0.34800359	0.04942228	AT3G26618.1	eukaryotic release factor 1-3	noTP	
Sevir.2G156900	0.31241102	0.01471991	AT2G04350.1	AMP-dependent synthetase and ligase family protein	noTP	BS
Sevir.6G220800	0.51995458	0.00544716	AT4G29950.1	Ypt/Rab-GAP domain of gyp1p superfamily protein	noTP	M
Sevir.1G160200	0.70126861	0.00838079	AT3G02750.1	Protein phosphatase 2C family protein	noTP	M
Sevir.1G177200	0.29324518	0.01861548	AT5G13430.1	Ubiquinol-cytochrome C reductase iron-sulfur subunit	mTP	BS
Sevir.5G369800	0.71493373	0.00848058			noTP	M
Sevir.1G036192	0.73262642	0.00441899			noTP	M
Sevir.6G162400	0.57336391	0.00290011	AT1G68580.2	agenet domain-containing protein / bromo-adjacent homology (BAH) domain-containing protein	noTP	M
Sevir.5G259800	0.51375799	0.00421367	AT1G64230.4	ubiquitin-conjugating enzyme 28	noTP	BS
Sevir.1G207300	0.78455506	0.00013447	AT5G64300.1	GTP cyclohydrolase II	cTP	M
Sevir.3G349900	0.56999398	0.00516391	AT3G62420.1	basic region/leucine zipper motif 53	noTP	
Sevir.9G259800	0.68876913	0.00722352	AT4G09670.1	Oxidoreductase family protein	noTP	M
Sevir.3G079600	0.28942887	0.02550509	AT1G64230.2	ubiquitin-conjugating enzyme 28	noTP	BS
Sevir.1G305400	0.56629476	1.64E-05	AT5G10770.1	Eukaryotic aspartyl protease family protein	SP	M
Sevir.9G300500	0.76914672	0.00935686	AT4G30100.1	P-loop containing nucleoside triphosphate hydrolases superfamily protein	noTP	BS
Sevir.2G393300	0.44951485	0.03447876	AT1G69120.1	K-box region and MADS-box transcription factor family protein	noTP	
Sevir.5G174500	0.64510644	0.03193251	AT2G35840.1	Sucrose-6F-phosphate phosphohydrolase family protein	noTP	
Sevir.9G199800	0.30667747	0.00251017	AT4G00570.1	NAD-dependent malic enzyme 2	mTP	BS
Sevir.1G275100	0.43412058	0.03356364	AT1G63050.1	MBOAT (membrane bound O-acyl transferase) family protein	cTP	M
Sevir.7G087800	0.38614336	0.01471991	AT5G61840.1	Exostosin family protein	SP	
Sevir.9G311200	0.44958837	0.00458319	AT2G02040.1	peptide transporter 2	noTP	M
Sevir.3G384600	1.74554517	0.01270622	AT3G19540.1	Protein of unknown function (DUF620)	noTP	M
Sevir.4G300900	0.85676289	0.02813579	AT5G17230.2	PHYTOENE SYNTHASE	noTP	BS
Sevir.3G026600	0.39212559	0.00754593	AT5G20990.1	molybdopterin biosynthesis CNX1 protein / molybdenum cofactor biosynthesis enzyme CNX1 (CNX1)	noTP	
Sevir.9G243600	0.6144219	0.01347029	AT3G01180.1	starch synthase 2	cTP	
Sevir.5G202100	0.57026222	0.00616685	AT3G02520.1	general regulatory factor 7	noTP	BS
Sevir.3G346900	0.38802711	0.01953081	AT1G09430.1	ATP-citrate lyase A-3	noTP	
Sevir.5G219800	0.84466586	0.04321839	AT3G24420.1	alpha/beta-Hydrolases superfamily protein	noTP	M
Sevir.7G335900	0.4276477	0.01138011	AT5G03940.1	chloroplast signal recognition particle 54 kDa subunit	cTP	M
Sevir.7G160503	0.34609177	0.02043658	AT2G20580.1	26S proteasome regulatory subunit S2 1A	noTP	
Sevir.4G208300	0.69951385	0.01914555	AT5G35090.1	noTP	M	
Sevir.2G141100	0.63155055	0.00243474	AT2G45910.1	U-box domain-containing protein kinase family protein	noTP	M
Sevir.1G068400	0.41568584	0.03440491	AT1G79440.1	aldehyde dehydrogenase 5F1	mTP	
Sevir.5G433700	0.5019858	0.01402074	AT3G02830.1	zinc finger protein 1	noTP	BS
Sevir.5G254800	0.51786425	0.01918039	AT5G15900.1	TRICHOME BIREFRINGENCE-LIKE 19	SP	M

Sevir.9G272200	0.7100124	0.04491331	AT1G22040.1	Galactose oxidase/kelch repeat superfamily protein	noTP	M
Sevir.6G158600	0.65639322	0.02454494	AT1G68090.1	annexin 5	noTP	BS
Sevir.3G424800	0.55062855	0.02107793	AT5G62670.1	H(+)-ATPase 11	noTP	
Sevir.5G382700	0.49653208	0.02165174	AT2G37650.1	GRAS family transcription factor	noTP	BS
Sevir.8G120400	0.24317639	0.02327004	AT2G43070.1	SIGNAL PEPTIDE PEPTIDASE-LIKE 3	SP	
Sevir.9G461900	0.70343684	0.02209198	AT4G34660.1	SH3 domain-containing protein	noTP	M
Sevir.1G145100	0.50769697	0.02598526	AT1G70780.1	noTP	M	
Sevir.5G438700	0.50091569	0.02490061	AT4G22950.1	AGAMOUS-like 19	noTP	BS
Sevir.5G378400	0.65528551	0.01940776	AT5G18520.1	Lung seven transmembrane receptor family protein	SP	M
Sevir.2G142600	0.77797276	0.02743894	AT5G23950.1	Calcium-dependent lipid-binding (CaLB domain) family protein	noTP	M
Sevir.9G033600	0.29216754	0.02598526	AT5G59840.1	Ras-related small GTP-binding family protein	noTP	M
Sevir.5G324200	0.55595241	0.00610326	AT1G28520.2	vascular plant one zinc finger protein	noTP	BS
Sevir.1G243000	0.95153602	0.00077399	AT5G44060.1	noTP	BS	
Sevir.9G464500	0.62301447	0.02329584	AT1G08540.1	RNApolymerase sigma subunit 2	cTP	M
Sevir.7G005700	0.53953315	0.00847741	AT3G04580.1	Signal transduction histidine kinase, hybrid-type, ethylene sensor	SP	M
Sevir.9G348800	0.36769113	0.00152334	AT4G01370.1	MAP kinase 4	noTP	M
Sevir.2G095360	0.63937033	0.02741607			noTP	M
Sevir.7G246800	0.85045076	1.05E-09	AT2G35060.1	K+ uptake permease 11	noTP	
Sevir.9G076800	0.33481569	0.04271793	AT1G47640.1	noTP		
Sevir.2G202700	1.03388034	0.04430823	AT5G62560.1	RING/U-box superfamily protein with ARM repeat domain	noTP	M
Sevir.2G005000	0.70837736	1.35E-10	AT1G16270.2	Protein kinase superfamily protein with octicosapeptide/Phox/Bem1p domain	noTP	BS
Sevir.9G147900	0.70888147	0.01162045	AT5G47430.1	DWNN domain, a CCHC-type zinc finger	noTP	BS
Sevir.4G101900	0.55594628	0.02573574	AT3G48040.1	RHO-related protein from plants 10	noTP	M
Sevir.5G421700	0.91785937	0.01263797	AT1G05805.1	basic helix-loop-helix (bHLH) DNA-binding superfamily protein	noTP	M
Sevir.9G120600	0.38050687	0.02330961	AT1G18180.1	Protein of unknown function (DUF1295)	noTP	M
Sevir.8G117400	0.29618237	0.04044911	AT3G48420.1	Haloacid dehalogenase-like hydrolase (HAD) superfamily protein	cTP	M
Sevir.9G006500	0.65386387	0.03496139	AT4G17890.1	ARF-GAP domain 8	noTP	M
Sevir.8G083300	0.65809675	0.03629459	AT5G49900.1	Beta-glucosidase, GBA2 type family protein	noTP	
Sevir.7G003400	0.27510843	0.03043451	AT2G45980.1	noTP		
Sevir.6G113500	0.71703007	0.00868025	AT5G25100.2	Endomembrane protein 70 protein family	noTP	
Sevir.5G427300	1.46034278	0.00187905	AT5G14420.3	RING domain ligase2	noTP	M
Sevir.1G125600	0.79039213	0.02561778	AT3G07040.1	NB-ARC domain-containing disease resistance protein	noTP	M
Sevir.3G020100	1.01113258	0.03142645			noTP	
Sevir.3G054200	0.5941775	0.0125593	AT2G47060.4	Protein kinase superfamily protein	noTP	M
Sevir.3G002700	0.43837174	0.01138011	AT4G33480.1	cTP	M	
Sevir.9G504500	1.82432786	0.00695504	AT4G37640.1	calcium ATPase 2	noTP	
Sevir.9G016400	0.91554193	0.04842269	AT2G44060.1	Late embryogenesis abundant protein, group 2	noTP	M
Sevir.2G140900	0.70077173	0.04491331	AT2G45910.1	U-box domain-containing protein kinase family protein	noTP	M
Sevir.1G206300	0.41153832	0.02851099	AT1G47380.1	Protein phosphatase 2C family protein	noTP	M

Sevir.4G244902	0.73169506	0.00956459	AT5G60900.1	receptor-like protein kinase 1	SP	M
Sevir.1G150800	0.41297188	0.0082858	AT1G20080.1	Calcium-dependent lipid-binding (CaLB domain) family protein	noTP	M
Sevir.5G132200	0.64973319	0.00848058	AT1G35190.1	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein	noTP	BS
Sevir.9G029600	1.44641698	0.00371404			SP	
Sevir.7G299900	0.66512845	0.00045398	AT3G52950.1	CBS / octicosapeptide/Phox/Bemp1 (PB1) domains-containing protein	noTP	M
Sevir.1G296600	0.55102461	0.01413029	AT3G27960.1	Tetratricopeptide repeat (TPR)-like superfamily protein	noTP	M
Sevir.6G120075	0.4335369	0.01565353	AT5G11100.1	Calcium-dependent lipid-binding (CaLB domain) family protein	noTP	M
Sevir.5G411500	0.70603163	0.04563702	AT1G57680.1	noTP	M	
Sevir.9G568200	0.55561554	0.04325233	AT4G11820.2	hydroxymethylglutaryl-CoA synthase / HMG-CoA synthase / 3-hydroxy-3-methylglutaryl coenzyme A synthase	noTP	M
Sevir.5G385800	0.40809236	0.04403794	AT2G37790.1	NAD(P)-linked oxidoreductase superfamily protein	noTP	
Sevir.5G433000	0.6493675	0.0455544	AT5G47430.1	DWNN domain, a CCHC-type zinc finger	noTP	M
Sevir.1G111800	0.87156481	0.00701708	AT2G17220.1	Protein kinase superfamily protein	noTP	M
Sevir.2G453866	0.63754363	0.00417664	AT3G10985.1	senescence associated gene 20	noTP	
Sevir.9G470000	1.05906069	0.04941374	AT5G19500.1	Tryptophan/tyrosine permease	cTP	BS
Sevir.3G006600	0.32364281	0.03431828	AT1G25280.1	tubby like protein 10	noTP	BS
Sevir.9G563800	0.67276395	0.03563695	AT4G00300.2	fringe-related protein	SP	M
Sevir.7G134900	0.47322435	0.04430823	AT1G25220.1	anthranilate synthase beta subunit 1	cTP	BS
Sevir.5G359500	0.67003469	0.00200344	AT3G56840.1	FAD-dependent oxidoreductase family protein	mTP	M
Sevir.1G169400	1.07644416	2.34E-11	AT1G67480.1	Galactose oxidase/kelch repeat superfamily protein	noTP	BS
Sevir.6G078500	1.15467389	3.02E-05	AT5G40780.1	lysine histidine transporter 1	noTP	M
Sevir.7G052000	0.55031835	0.03440491	AT2G01290.1	ribose-5-phosphate isomerase 2	noTP	BS
Sevir.1G021700	0.81903776	0.04510193	AT5G45800.1	Leucine-rich repeat protein kinase family protein	noTP	
Sevir.5G054100	0.92545258	0.04842269	AT1G65910.1	NAC domain containing protein 28	noTP	M
Sevir.5G432900	0.5954461	0.04588781	AT4G14030.2	selenium-binding protein 1	noTP	M
Sevir.8G088400	1.1317354	0.04067646	AT3G07040.1	NB-ARC domain-containing disease resistance protein	noTP	M
Sevir.2G239100	1.19608729	0.03135907	AT3G48690.1	alpha/beta-Hydrolases superfamily protein	noTP	M
Sevir.3G030800	0.41920434	0.02327004	AT5G65930.2	kinesin-like calmodulin-binding protein (ZWICHEL)	noTP	M
Sevir.1G354000	0.55388666	0.00480599	AT4G33400.1	Vacuolar import/degradation, Vid27-related protein	noTP	BS
Sevir.3G111000	1.8612613	0.01179611	AT5G26170.1	WRKY DNA-binding protein 50	noTP	
Sevir.7G331800	0.52367258	0.02569082	AT3G44610.1	Protein kinase superfamily protein	noTP	M
Sevir.9G144900	0.59482944	0.03982227	AT5G45110.1	NPR1-like protein 3	noTP	
Sevir.3G172200	0.72749543	0.01784482	AT1G28240.1	Protein of unknown function (DUF616)	noTP	M
Sevir.2G201600	0.56411421	5.78E-05	AT5G62890.2	Xanthine/uracil permease family protein	noTP	BS
Sevir.5G428500	3.37648884	0.01164943			noTP	BS
Sevir.3G211800	0.58789588	0.01774498	AT3G10980.1	PLAC8 family protein	noTP	M
Sevir.1G352300	1.37514947	0.01094664	AT1G51660.1	mitogen-activated protein kinase kinase 4	noTP	M

Sevir.9G422900	0.59604283	0.01668165	AT3G04030.3	Homeodomain-like superfamily protein	noTP	BS
Sevir.3G105000	1.20750084	0.00103663	AT4G32410.1	cellulose synthase 1	noTP	M
Sevir.9G008400	0.51037791	0.0215605	AT5G48840.1	homolog of bacterial PANC	noTP	
Sevir.3G341100	1.25488845	0.03631643	AT1G07530.1	SCARECROW-like 14	noTP	BS
Sevir.5G016000	0.61176649	0.04840133			noTP	
Sevir.6G093100	0.66224685	0.01162045	AT4G05200.1	cysteine-rich RLK (RECEPTOR-like protein kinase) 25	SP	BS
Sevir.7G227000	0.99574346	0.01026083	AT5G47060.1	Protein of unknown function (DUF581)	noTP	M
Sevir.1G274200	1.42250981	0.00590506	AT2G27230.2	transcription factor-related	noTP	
Sevir.3G003100	0.77835643	0.0332219	AT1G22610.1	C2 calcium/lipid-binding plant phosphoribosyltransferase family protein	noTP	BS
Sevir.3G041600	1.34387325	0.04075194	AT2G19130.1	S-locus lectin protein kinase family protein	SP	M
Sevir.9G423700	0.9626975	0.0332219	AT3G14470.1	NB-ARC domain-containing disease resistance protein	noTP	
Sevir.9G062000	1.39911458	0.04885176	AT5G04870.1	calcium dependent protein kinase 1	noTP	M
Sevir.3G418200	0.50601604	0.03102076	AT3G46730.1	NB-ARC domain-containing disease resistance protein	noTP	M
Sevir.6G236100	1.39002839	0.03283136	AT1G27170.1	transmembrane receptors;ATP binding	noTP	BS
Sevir.1G261900	0.66660657	0.01145562	AT5G47230.1	ethylene responsive element binding factor 5	noTP	M
Sevir.8G260300	0.88574151	0.03433417	AT5G24318.1	O-Glycosyl hydrolases family 17 protein	SP	
Sevir.1G354700	1.16444632	0.02591865	AT4G30440.1	UDP-D-glucuronate 4-epimerase 1	noTP	M
Sevir.5G439200	0.59947708	0.01430455	AT1G67310.1	Calmodulin-binding transcription activator protein with CG-1 and Ankyrin domains	noTP	BS
Sevir.3G295100	1.45539458	0.00635202	AT3G56400.1	WRKY DNA-binding protein 70	noTP	M
Sevir.8G200500	0.57563013	0.03427615	AT3G46530.1	NB-ARC domain-containing disease resistance protein	noTP	
Sevir.3G390200	1.66128823	0.00992948			noTP	M
Sevir.9G132550	0.97062171	0.00015157	AT3G61130.1	galacturonosyltransferase 1	SP	M
Sevir.8G057600	0.94910227	0.00015995	AT1G65870.1	Disease resistance-responsive (dirigent-like protein) family protein	SP	M
Sevir.2G336000	1.31440235	4.04E-05	AT3G16520.3	UDP-glucosyl transferase 88A1	noTP	M
Sevir.3G080600	1.15982758	0.0215605	AT3G24550.1	proline extensin-like receptor kinase 1	noTP	
Sevir.4G129900	1.17867178	0.00153759	AT3G14270.1	phosphatidylinositol-4-phosphate 5-kinase family protein	noTP	
Sevir.1G110500	0.57856267	0.03851102	AT3G51300.1	RHO-related protein from plants 1	noTP	M
Sevir.4G090000	0.85850077	0.00290011	AT4G08950.1	Phosphate-responsive 1 family protein	SP	M
Sevir.9G289000	1.68789458	4.38E-12	AT1G32780.1	GroES-like zinc-binding dehydrogenase family protein	noTP	
Sevir.5G058000	0.55276035	0.03235198	AT3G66654.3	Cyclophilin-like peptidyl-prolyl cis-trans isomerase family protein	noTP	M
Sevir.3G355600	2.23608488	0.03440491	AT5G57580.1	Calmodulin-binding protein	noTP	M
Sevir.8G178700	0.72630439	0.04798473	AT2G38860.2	Class I glutamine amidotransferase-like superfamily protein	noTP	BS
Sevir.9G469300	0.72057874	0.02828091	AT2G28470.1	beta-galactosidase 8	SP	M
Sevir.2G055300	0.73849032	0.02798898	AT4G14770.1	TESMIN/TSO1-like CXC 2	noTP	M
Sevir.9G022000	1.37055259	5.12E-05	AT5G28050.1	Cytidine/deoxycytidylate deaminase family protein	noTP	
Sevir.9G221200	0.49369188	0.04842269	AT2G37035.1	noTP		
Sevir.3G326500	1.14962724	0.03447876			noTP	BS

Sevir.8G075700	1.46528334	0.01055649	AT1G18480.1	Calcineurin-like metallo-phosphoesterase superfamily protein	noTP	
Sevir.3G002900	0.58983553	0.03871733	AT1G77610.1	EamA-like transporter family protein	noTP	M
Sevir.9G387000	1.6676039	0.04911956	AT5G64080.2	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein	SP	M
Sevir.9G447600	1.52050309	0.03697524	AT2G38905.1	Low temperature and salt responsive protein family	SP	M
Sevir.5G022900	1.5934805	0.04287719	AT2G41380.1	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein	noTP	
Sevir.2G379900	1.67723548	0.00701708	AT3G18830.1	polyol/monosaccharide transporter 5	noTP	
Sevir.3G115500	1.20821692	0.00109192	AT1G17620.1	Late embryogenesis abundant (LEA) hydroxyproline-rich glycoprotein family	noTP	M
Sevir.7G215700	1.37191631	0.01921232	AT2G29050.1	RHOMBOID-like 1	noTP	M
Sevir.5G042432	0.79763618	0.01964014	AT1G48910.1	Flavin-containing monooxygenase family protein	noTP	BS
Sevir.9G050100	2.32856574	0.03142645	AT1G59710.1	Protein of unknown function (DUF569)	noTP	M
Sevir.5G144400	0.81338927	0.00935686	AT5G45820.1	CBL-interacting protein kinase 20	noTP	
Sevir.4G189200	0.69588598	0.04431043	AT4G11000.1	Ankyrin repeat family protein	noTP	M
Sevir.7G202300	0.7474659	0.03871733	AT4G00030.1	Plastid-lipid associated protein PAP / fibrillin family protein	cTP	M
Sevir.6G101900	1.99955129	9.65E-07	AT3G27870.1	ATPase E1-E2 type family protein / haloacid dehalogenase-like hydrolase family protein	noTP	
Sevir.9G021600	0.96272996	0.04622037	AT1G48350.1	Ribosomal L18p/L5e family protein	cTP	M
Sevir.4G259800	1.43747017	0.01914555	AT2G18980.1	Peroxidase superfamily protein	SP	BS
Sevir.3G197300	0.87087518	0.01413029	AT3G57270.1	beta-1,3-glucanase 1	noTP	M
Sevir.9G544400	1.07540192	0.03865743	AT4G00330.1	calmodulin-binding receptor-like cytoplasmic kinase 2	noTP	M
Sevir.1G294700	1.02790854	0.02543641	AT4G34530.1	cryptochrome-interacting basic-helix-loop-helix 1	noTP	
Sevir.4G113500	1.21254371	0.02238253	AT5G37710.1	alpha/beta-Hydrolases superfamily protein	noTP	M
Sevir.5G314600	1.54013982	0.04503226	AT4G29310.1	Protein of unknown function (DUF1005)	noTP	M
Sevir.1G035000	0.87685168	0.00195571	AT4G38510.5	ATPase, V1 complex, subunit B protein	noTP	BS
Sevir.3G407000	0.73052406	0.02005705	AT3G52060.2	Core-2/I-branching beta-1,6-N-acetylglucosaminyltransferase family protein	SP	
Sevir.2G351600	0.45952024	0.03826021	AT4G23180.1	cysteine-rich RLK (RECEPTOR-like protein kinase) 10	SP	M
Sevir.5G257600	1.49206989	0.00347433	AT4G23810.1	WRKY family transcription factor	noTP	M
Sevir.2G014400	0.61479962	0.00641196	AT1G69550.1	disease resistance protein (TIR-NBS-LRR class)	noTP	M
Sevir.3G079000	0.78361305	0.00090425	AT2G46220.1	Uncharacterized conserved protein (DUF2358)	cTP	M
Sevir.2G186800	1.31727808	0.01167906	AT4G23810.1	WRKY family transcription factor	noTP	M
Sevir.3G371000	0.96804736	0.01279135	AT5G67390.1	noTP	M	
Sevir.9G314500	0.78885858	0.00492397	AT5G16760.1	Inositol 1,3,4-trisphosphate 5/6-kinase family protein	noTP	BS
Sevir.9G483800	0.92139298	0.00023118	AT4G34500.1	Protein kinase superfamily protein	noTP	
Sevir.2G049300	0.87472998	0.02063653	AT2G19440.1	O-Glycosyl hydrolases family 17 protein	SP	BS

Sevir.8G164300	1.53619954	0.00556175	AT2G23770.1	protein kinase family protein / peptidoglycan-binding LysM domain-containing protein	noTP	M
Sevir.5G254500	1.22881166	0.00466567	AT1G29050.1	TRICHOME BIREFRINGENCE-LIKE 38	SP	M
Sevir.8G254700	0.95048115	8.35E-05	AT1G50180.1	NB-ARC domain-containing disease resistance protein	noTP	M
Sevir.9G498200	1.0914358	0.03607398	AT5G15470.1	galacturonosyltransferase 14	noTP	M
Sevir.8G091100	1.24226566	0.04940707	AT3G07040.1	NB-ARC domain-containing disease resistance protein	noTP	M
Sevir.7G025600	1.49015816	0.01938738	AT1G05675.1	UDP-Glycosyltransferase superfamily protein	noTP	
Sevir.7G125600	1.04291889	0.03563695	AT3G60520.1	noTP	M	
Sevir.6G013900	1.6002195	0.01138011	AT5G57550.1	xyloglucan endotransglucosylase/hydrolase 25	SP	
Sevir.9G180000	1.30640569	0.03214553	AT5G55050.1	GDSL-like Lipase/Acylhydrolase superfamily protein	SP	M
Sevir.6G216800	0.84899041	0.03974513	AT5G66600.1	Protein of unknown function, DUF547	noTP	BS
Sevir.7G048300	0.66114353	0.02554008	AT3G14460.1	LRR and NB-ARC domains-containing disease resistance protein	SP	BS
Sevir.7G102400	1.02307736	0.0072445	AT5G54200.1	Transducin/WD40 repeat-like superfamily protein	noTP	BS
Sevir.8G022400	0.7056122	0.03427615	AT5G58560.1	Phosphatidate cytidylyltransferase family protein	noTP	
Sevir.5G294500	1.15996371	0.00556175	AT2G42690.1	alpha/beta-Hydrolases superfamily protein	noTP	BS
Sevir.2G030000	0.99193793	0.00480599	AT3G53810.1	Concanavalin A-like lectin protein kinase family protein	SP	M
Sevir.7G091400	1.13233833	0.00466567	AT3G12920.1	SBP (S-ribonuclease binding protein) family protein	noTP	BS
Sevir.9G180200	0.89781118	0.0190984	AT3G14470.1	NB-ARC domain-containing disease resistance protein	noTP	
Sevir.3G201500	1.2602234	0.00024218	AT4G18380.1	F-box family protein	noTP	M
Sevir.7G307300	1.7193044	0.00544716	AT2G35930.1	plant U-box 23	noTP	M
Sevir.1G379100	0.88133177	0.01497817	AT3G53140.1	O-methyltransferase family protein	noTP	M
Sevir.7G068700	1.01238625	0.04184021	AT1G21270.1	wall-associated kinase 2	SP	M
Sevir.1G072000	0.75768056	0.00848058			noTP	
Sevir.5G096600	1.12980239	0.01813566	AT3G54140.1	peptide transporter 1	noTP	
Sevir.1G123000	1.32301465	0.04011464	AT4G08850.1	Leucine-rich repeat receptor-like protein kinase family protein	SP	BS
Sevir.8G196600	0.75289364	0.02278812	AT3G07040.1	NB-ARC domain-containing disease resistance protein	noTP	M
Sevir.8G042300	0.76996766	0.04992661	AT5G09810.1	actin 7	noTP	M
Sevir.9G428300	1.77030005	0.02107025			noTP	BS
Sevir.5G033000	0.7880361	0.02091221	AT1G09540.1	myb domain protein 61	noTP	M
Sevir.9G056400	2.11263317	0.00243474	AT3G22800.1	Leucine-rich repeat (LRR) family protein	SP	M
Sevir.2G418300	0.62065368	0.03842914	AT3G17440.1	novel plant snare 13	noTP	M
Sevir.2G006800	1.23479834	0.03563695	AT3G02130.1	receptor-like protein kinase 2	SP	M
Sevir.1G307300	0.99048268	0.00556289	AT3G07650.4	CONSTANS-like 9	noTP	BS
Sevir.9G132000	0.90355027	0.01095717	AT1G48320.1	Thioesterase superfamily protein	noTP	
Sevir.3G205200	0.60972777	0.04633885	AT2G38650.2	galacturonosyltransferase 7	noTP	M
Sevir.9G108400	0.85166068	0.04491331	AT1G58520.1	lipases;hydrolases, acting on ester bonds	noTP	
Sevir.1G169500	1.02647615	0.00115109			noTP	BS
Sevir.7G278800	1.1616	0.01953081	AT1G66340.1	Signal transduction histidine kinase, hybrid-type, ethylene sensor	noTP	BS
Sevir.9G580500	0.93633207	0.0224126	AT4G00150.1	GRAS family transcription factor	noTP	BS

Sevir.2G313400	2.43527275	0.0089491	AT5G65980.1	Auxin efflux carrier family protein	noTP	M
Sevir.9G154800	0.78554858	0.02512457	AT5G37870.1	Protein with RING/U-box and TRAF-like domains	noTP	M
Sevir.9G556300	0.90451287	0.02561778	AT1G20510.1	OPC-8:0 CoA ligase1	noTP	BS
Sevir.3G164800	0.50556592	0.03007251	AT5G47770.1	farnesyl diphosphate synthase 1	noTP	
Sevir.5G022400	1.1805035	0.00706981	AT3G56891.1	Heavy metal transport/detoxification superfamily protein	noTP	BS
Sevir.3G061300	1.45953348	0.01471991	AT2G19130.1	S-locus lectin protein kinase family protein	noTP	M
Sevir.2G392100	1.36887473	0.00012018	AT1G74160.1	noTP	M	
Sevir.5G459500	1.27771019	0.00575457	AT3G09270.1	glutathione S-transferase TAU 8	noTP	
Sevir.1G031900	1.03615401	0.03058013	AT5G56750.1	N-MYC downregulated-like 1	noTP	M
Sevir.6G194100	0.71996489	0.01753533			noTP	M
Sevir.3G107100	1.41118118	0.00309501	AT4G13395.1	ROTUNDIFOLIA like 12	noTP	M
Sevir.3G199000	0.87508878	0.03923067	AT2G40890.1	cytochrome P450, family 98, subfamily A, polypeptide 3	cTP	
Sevir.5G254400	1.40242064	0.03447876	AT4G18570.1	Tetratricopeptide repeat (TPR)-like superfamily protein	noTP	M
Sevir.3G384650	1.69335402	0.00241069			noTP	M
Sevir.6G220100	1.04670574	0.04618322	AT3G20240.1	Mitochondrial substrate carrier family protein	noTP	M
Sevir.7G102600	0.93386503	0.00823763			noTP	
Sevir.7G226900	1.61322116	0.00162916	AT5G47060.1	Protein of unknown function (DUF581)	noTP	M
Sevir.7G144500	0.75388484	0.04345329	AT5G11310.1	Pentatricopeptide repeat (PPR) superfamily protein	noTP	
Sevir.7G245100	1.20103255	0.00379342	AT1G79460.1	Terpenoid cyclases/Protein prenyltransferases superfamily protein	noTP	BS
Sevir.7G261100	0.87936379	0.03081755	AT1G61720.1	NAD(P)-binding Rossmann-fold superfamily protein	noTP	M
Sevir.2G364900	1.08151157	0.03447876	AT1G61250.1	secretory carrier 3	noTP	M
Sevir.6G132100	1.86187398	0.0263521	AT3G50150.1	Plant protein of unknown function (DUF247)	noTP	BS
Sevir.1G071500	1.28961169	0.00030206	AT2G35330.1	RING/U-box superfamily protein	noTP	
Sevir.1G201700	0.87454543	0.01332492	AT1G76690.1	12-oxophytodienoate reductase 2	noTP	
Sevir.3G307500	1.08667007	0.02741607	AT1G67580.1	Protein kinase superfamily protein	noTP	
Sevir.4G048100	1.35030162	0.00701708	AT5G50380.1	exocyst subunit exo70 family protein F1	noTP	
Sevir.3G080700	1.77459441	0.02851099	AT2G19130.1	S-locus lectin protein kinase family protein	SP	M
Sevir.5G413950	1.3507011	0.00138626			noTP	M
Sevir.9G369350	0.80826308	0.02561778			noTP	M
Sevir.4G101700	0.99998021	0.01305214	AT1G60060.1	Serine/threonine-protein kinase WNK (With No Lysine)-related	noTP	M
Sevir.8G027800	1.85191586	0.02927213	AT3G10190.1	Calcium-binding EF-hand family protein	noTP	
Sevir.2G385650	3.78182977	0.00012279			SP	M
Sevir.5G262700	1.20580841	0.03887244	AT1G60550.1	enoyl-CoA hydratase/isomerase D	noTP	BS
Sevir.2G263200	1.84140695	0.02327004	AT2G42320.2	nucleolar protein gar2-related	noTP	M
Sevir.7G077800	1.20173318	0.00111619	AT1G67340.1	HCP-like superfamily protein with MYND-type zinc finger	noTP	
Sevir.3G056200	1.9427939	0.00602284	AT3G05880.1	Low temperature and salt responsive protein family	noTP	M
Sevir.9G439800	1.39202648	0.04404903	AT3G51680.1	NAD(P)-binding Rossmann-fold superfamily protein	noTP	M
Sevir.3G060400	0.91337523	0.04840133	AT2G35840.1	Sucrose-6F-phosphate phosphohydrolase family protein	noTP	BS
Sevir.4G248900	1.46272039	0.01953081	AT5G21105.1	Plant L-ascorbate oxidase	SP	

Sevir.9G557400	1.70601203	0.00641196	AT2G17220.1	Protein kinase superfamily protein	noTP	M
Sevir.1G101300	3.4053503	1.49E-05	AT1G18485.1	Pentatricopeptide repeat (PPR) superfamily protein	noTP	
Sevir.6G039300	1.3152254	0.00076705			noTP	
Sevir.1G104500	1.19413591	0.00556289	AT1G29950.2	basic helix-loop-helix (bHLH) DNA-binding superfamily protein	noTP	BS
Sevir.6G225000	0.83915551	0.03418098	AT2G31680.1	RAB GTPase homolog A5D	noTP	M
Sevir.1G262600	1.33170229	0.01927257	AT5G07620.1	Protein kinase superfamily protein	noTP	M
Sevir.9G367900	1.56760216	0.00848058	AT4G00340.1	receptor-like protein kinase 4	SP	M
Sevir.5G408200	1.35702587	0.00215813	AT2G25170.1	chromatin remodeling factor CHD3 (PICKLE)	noTP	
Sevir.8G247020	0.82266381	0.03842914			noTP	BS
Sevir.9G355800	1.52777729	0.00754593	AT1G74300.1	alpha/beta-Hydrolases superfamily protein	noTP	BS
Sevir.5G072800	0.90391099	0.02109769	AT5G38560.1	Protein kinase superfamily protein	noTP	
Sevir.1G204500	0.8949161	0.03582328	AT1G09900.1	Pentatricopeptide repeat (PPR-like) superfamily protein	cTP	M
Sevir.6G198500	1.75667905	0.02741607	AT1G13635.2	DNA glycosylase superfamily protein	noTP	BS
Sevir.6G008600	1.76664281	0.0076771	AT5G47850.1	CRINKLY4 related 4	SP	BS
Sevir.4G194600	0.96087195	0.04618322	AT1G13600.1	basic leucine-zipper 58	noTP	BS
Sevir.7G157700	0.81126794	0.02913935	AT4G03290.1	EF hand calcium-binding protein family	noTP	M
Sevir.7G217700	1.56846492	0.02020379	AT4G22990.1	Major Facilitator Superfamily with SPX (SYG1/Pho81/XPR1) domain-containing protein	noTP	M
Sevir.3G089200	1.21667524	0.00859595	AT3G48510.1	noTP	BS	
Sevir.4G005100	1.65026258	0.01888469	AT5G24000.1	Protein of unknown function (DUF819)	cTP	
Sevir.3G377400	1.31272017	0.01431413	AT1G78160.1	pumilio 7	noTP	BS
Sevir.7G174800	1.00798023	0.00670013			noTP	BS
Sevir.1G261800	1.0809032	0.00840435	AT2G44850.2	noTP		
Sevir.3G004300	1.14025632	0.00695504	AT1G77850.1	auxin response factor 17	noTP	M
Sevir.5G304800	1.41497483	0.04842269	AT3G28345.1	ABC transporter family protein	noTP	BS
Sevir.1G346500	1.30568582	0.0118045	AT3G43850.1	noTP	M	
Sevir.5G354900	5.3514701	0.00051323	AT5G04010.1	F-box family protein	noTP	
Sevir.4G225000	2.37425534	0.00039151	AT3G02600.1	lipid phosphate phosphatase 3	noTP	BS
Sevir.9G109000	4.95832843	0.00010405			noTP	
Sevir.7G220400	1.38495741	0.00023118	AT1G33265.1	Transmembrane proteins 14C	cTP	
Sevir.1G286100	1.29594242	0.02591865	AT4G25960.1	P-glycoprotein 2	noTP	BS
Sevir.8G071400	1.57839509	0.00338633	AT5G13750.1	zinc induced facilitator-like 1	noTP	M
Sevir.5G264900	1.71315156	0.02512457	AT1G06475.1	noTP	M	
Sevir.1G146300	1.4833312	0.03868945	AT3G59890.1	Dihydrodipicolinate reductase, bacterial/plant	cTP	
Sevir.4G084000	2.37075758	0.02512457			noTP	
Sevir.7G333700	2.71616315	0.03697524	AT2G27690.1	cytochrome P450, family 94, subfamily C, polypeptide 1	noTP	M
Sevir.1G316500	2.3029625	0.00220412			noTP	
Sevir.7G225050	2.18097024	0.00055804			noTP	
Sevir.2G451900	1.38393545	0.02281631	AT2G41010.1	calmodulin (CAM)-binding protein of 25 kDa	noTP	M
Sevir.7G238600	1.41509169	0.0025196	AT4G30160.1	villin 4	cTP	
Sevir.7G050300	2.26327007	0.04937071	AT5G57150.1	basic helix-loop-helix (bHLH) DNA-binding superfamily protein	noTP	BS
Sevir.5G453000	2.13004472	0.00829678	AT4G03110.1	RNA-binding protein-defense related 1	noTP	M
Sevir.7G142000	1.65744702	0.04672623	AT4G26690.1	PLC-like phosphodiesterase family protein	SP	M
Sevir.1G160100	1.22655745	0.01964014			noTP	M
Sevir.5G468400	1.28330417	0.04563702			noTP	
Sevir.2G241100	1.16222404	0.02814104	AT5G42750.1	BRI1 kinase inhibitor 1	noTP	M

Sevir.1G284300	1.55135493	0.03440491	AT3G52450.1	plant U-box 22	noTP	M
Sevir.7G025400	2.47608616	0.00200344	AT1G05680.1	Uridine diphosphate glycosyltransferase 74E2	cTP	
Sevir.1G079600	1.79268853	0.02550509	AT3G54690.1	Sugar isomerase (SIS) family protein	noTP	
Sevir.5G242800	2.34054001	0.00829678	AT3G14610.1	cytochrome P450, family 72, subfamily A, polypeptide 7	noTP	
Sevir.1G365700	2.25292754	0.01164943	AT1G21270.1	wall-associated kinase 2	SP	M
Sevir.6G148400	1.54619231	0.03994616			noTP	
Sevir.6G035760	1.49356733	0.00660624	AT5G38960.1	RmLC-like cupins superfamily protein	noTP	
Sevir.3G128400	1.28479981	0.04964285	AT1G53250.1	noTP	M	
Sevir.5G348900	1.70828019	0.02107793	AT1G11580.1	methylesterase PCR A	noTP	M
Sevir.5G289400	1.55816732	0.01940319	AT4G26090.1	NB-ARC domain-containing disease resistance protein	noTP	BS
Sevir.1G237000	1.48503454	0.03401129	AT3G47090.1	Leucine-rich repeat protein kinase family protein	noTP	M
Sevir.7G224900	2.16334176	0.02926549	AT1G29340.1	plant U-box 17	noTP	
Sevir.2G258700	1.87878652	0.03684169	AT5G25610.1	BURP domain-containing protein	noTP	BS

A.9.3 Significantly enriched GO terms in significantly upregulated or downregulated genes in *HATCH1* knock-out plants compared to wild-type

Table A.85 GO terms significantly enriched in genes significantly upregulated in *HATCH1* knock-out plants compared to wild-type

GO ID	Term	Annotated	Count	Expected
GO:0006396	RNA processing	373	21	6.93
GO:0015996	chlorophyll catabolic process	16	4	0.3
GO:0046149	pigment catabolic process	16	4	0.3
GO:0008380	RNA splicing	90	8	1.67
GO:1901565	organonitrogen compound catabolic process	116	9	2.15
GO:0043617	cellular response to sucrose starvation	2	2	0.04
GO:0006787	porphyrin-containing compound catabolic process	19	4	0.35
GO:0033015	tetrapyrrole catabolic process	19	4	0.35
GO:0051187	cofactor catabolic process	20	4	0.37
GO:0006551	leucine metabolic process	9	3	0.17
GO:0046700	heterocycle catabolic process	84	7	1.56
GO:0044270	cellular nitrogen compound catabolic process	85	7	1.58
GO:0044710	single-organism metabolic process	3533	87	65.61
GO:1901606	alpha-amino acid catabolic process	46	5	0.85
GO:0009269	response to desiccation	13	3	0.24
GO:0009063	cellular amino acid catabolic process	47	5	0.87
GO:0009098	leucine biosynthetic process	4	2	0.07

GO:0010114	response to red light	30	4	0.56
GO:0019439	aromatic compound catabolic process	125	8	2.32
GO:0016054	organic acid catabolic process	76	6	1.41
GO:0046395	carboxylic acid catabolic process	76	6	1.41
GO:0006397	mRNA processing	102	7	1.89
GO:1901361	organic cyclic compound catabolic process	130	8	2.41
GO:0015994	chlorophyll metabolic process	53	5	0.98
GO:0000375	RNA splicing, via transesterification reactions	58	5	1.08
GO:0000377	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	58	5	1.08
GO:0055114	oxidation-reduction process	1558	43	28.93
GO:1901605	alpha-amino acid metabolic process	207	10	3.84
GO:0009740	gibberellic acid mediated signaling pathway	20	3	0.37
GO:0007623	circadian rhythm	63	5	1.17
GO:0048511	rhythmic process	63	5	1.17
GO:0010476	gibberellin mediated signaling pathway	21	3	0.39
GO:0016071	mRNA metabolic process	152	8	2.82
GO:0071370	cellular response to gibberellin stimulus	22	3	0.41
GO:0032508	DNA duplex unwinding	8	2	0.15
GO:0009081	branched-chain amino acid metabolic process	24	3	0.45
GO:0009507	chloroplast	1441	42	26.54
GO:0044435	plastid part	873	29	16.08
GO:0044434	chloroplast part	856	28	15.77
GO:0009536	plastid	1523	43	28.05
GO:0009535	chloroplast thylakoid membrane	229	11	4.22
GO:0055035	plastid thylakoid membrane	231	11	4.25
GO:0043231	intracellular membrane-bounded organelle	4590	100	84.54
GO:0043227	membrane-bounded organelle	4592	100	84.57
GO:0042651	thylakoid membrane	253	11	4.66
GO:0034357	photosynthetic membrane	263	11	4.84
GO:0003852	2-isopropylmalate synthase activity	2	2	0.04
GO:0003723	RNA binding	473	20	8.46
GO:0046912	transferase activity, transferring acyl groups, acyl groups converted into alkyl on transfer	9	3	0.16
GO:0016811	hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds, in linear amides	35	4	0.63
GO:0016491	oxidoreductase activity	1651	44	29.52
GO:0016408	C-acyltransferase activity	7	2	0.13

GO:0004045	aminoacyl-tRNA hydrolase activity	8	2	0.14
GO:0061659	ubiquitin-like protein ligase activity	8	2	0.14
GO:0006396	RNA processing	373	21	6.93

Table A.86 GO terms significantly enriched in genes significantly downregulated in *HATCH1* knock-out plants compared to wild-type

GO ID	**Term**	**Annotated**	**Count**	**Expected**
GO:0005986	sucrose biosynthetic process	6	3	0.08
GO:0006796	phosphate-containing compound metabolic process	1915	45	26.52
GO:0006793	phosphorus metabolic process	1924	45	26.65
GO:0010243	response to organonitrogen compound	74	6	1.02
GO:0002376	immune system process	181	9	2.51
GO:0010200	response to chitin	55	5	0.76
GO:0044763	single-organism cellular process	4332	80	60
GO:0045087	innate immune response	156	8	2.16
GO:0006955	immune response	161	8	2.23
GO:0044262	cellular carbohydrate metabolic process	251	10	3.48
GO:0044699	single-organism process	6587	110	91.23
GO:0009933	meristem structural organization	42	4	0.58
GO:0044711	single-organism biosynthetic process	997	25	13.81
GO:0010072	primary shoot apical meristem specification	6	2	0.08
GO:0010264	myo-inositol hexakisphosphate biosynthetic process	6	2	0.08
GO:0033517	myo-inositol hexakisphosphate metabolic process	6	2	0.08
GO:0045893	positive regulation of transcription, DNA-templated	101	6	1.4
GO:1902680	positive regulation of RNA biosynthetic process	102	6	1.41
GO:1903508	positive regulation of nucleic acid-templated transcription	102	6	1.41
GO:0016310	phosphorylation	1538	34	21.3
GO:0051254	positive regulation of RNA metabolic process	107	6	1.48
GO:0005985	sucrose metabolic process	23	3	0.32
GO:0006020	inositol metabolic process	7	2	0.1
GO:0009306	protein secretion	7	2	0.1
GO:0006468	protein phosphorylation	1434	32	19.86
GO:2000028	regulation of photoperiodism, flowering	24	3	0.33
GO:0048532	anatomical structure arrangement	48	4	0.66
GO:0009617	response to bacterium	230	9	3.19

GO:0009631	cold acclimation	25	3	0.35
GO:0010557	positive regulation of macromolecule biosynthetic process	114	6	1.58
GO:0045935	positive regulation of nucleobase-containing compound metabolic process	114	6	1.58
GO:0032958	inositol phosphate biosynthetic process	8	2	0.11
GO:0007154	cell communication	720	19	9.97
GO:0044283	small molecule biosynthetic process	421	13	5.83
GO:0046351	disaccharide biosynthetic process	28	3	0.39
GO:0008299	isoprenoid biosynthetic process	86	5	1.19
GO:0010628	positive regulation of gene expression	123	6	1.7
GO:0090421	embryonic meristem initiation	10	2	0.14
GO:0031328	positive regulation of cellular biosynthetic process	127	6	1.76
GO:0048583	regulation of response to stimulus	298	10	4.13
GO:0051173	positive regulation of nitrogen compound metabolic process	128	6	1.77
GO:0051704	multi-organism process	758	19	10.5
GO:0042221	response to chemical	1388	30	19.22
GO:0016051	carbohydrate biosynthetic process	169	7	2.34
GO:0033180	proton-transporting V-type ATPase, V1 domain	8	2	0.12
GO:0032553	ribonucleotide binding	2897	67	37.86
GO:0001883	purine nucleoside binding	2848	66	37.22
GO:0032550	purine ribonucleoside binding	2848	66	37.22
GO:0032555	purine ribonucleotide binding	2848	66	37.22
GO:0032549	ribonucleoside binding	2853	66	37.28
GO:0001882	nucleoside binding	2854	66	37.3
GO:0017076	purine nucleotide binding	2855	66	37.31
GO:0097367	carbohydrate derivative binding	2926	67	38.24
GO:0043168	anion binding	3273	72	42.77
GO:0050307	sucrose-phosphate phosphatase activity	3	3	0.04
GO:0036094	small molecule binding	3539	75	46.25
GO:0032559	adenyl ribonucleotide binding	2625	60	34.3
GO:0030554	adenyl nucleotide binding	2629	60	34.36
GO:0000166	nucleotide binding	3457	73	45.18
GO:1901265	nucleoside phosphate binding	3457	73	45.18
GO:0004337	geranyltranstransferase activity	2	2	0.03
GO:0035639	purine ribonucleoside triphosphate binding	2494	53	32.59
GO:0050308	sugar-phosphatase activity	12	3	0.16
GO:0004672	protein kinase activity	1426	34	18.64

GO:0016773	phosphotransferase activity, alcohol group as acceptor	1561	36	20.4
GO:0005524	ATP binding	2271	47	29.68
GO:0043531	ADP binding	360	13	4.7
GO:0016301	kinase activity	1627	36	21.26
GO:0019203	carbohydrate phosphatase activity	18	3	0.24
GO:0043167	ion binding	5902	98	77.13
GO:0003824	catalytic activity	8351	129	109.13
GO:0000287	magnesium ion binding	125	6	1.63
GO:0004674	protein serine/threonine kinase activity	208	8	2.72
GO:0016772	transferase activity, transferring phosphorus-containing groups	1817	36	23.74
GO:0016833	oxo-acid-lyase activity	11	2	0.14
GO:1901363	heterocyclic compound binding	6519	103	85.19
GO:0097159	organic cyclic compound binding	6520	103	85.2
GO:0004842	ubiquitin-protein transferase activity	179	7	2.34
GO:0005986	sucrose biosynthetic process	6	3	0.08

A.10 *E. coli* pyruvate transporter knock-out generation

A.10.1 Primers used in generating *E. coli* pyruvate transporter knock-outs

Table A.87: Primers used for generating knock-out oligos

For primer sequence, blue = 50 bp target gene homology region, red = ~20 bp antibiotic resistance cassette homology region

Primer ID	Primer sequence (5'-3')
CstA/AmpR_for	AGAAACAAAATGTAACATCTCTATGGACACGCACACGGATAACA ACTATG GTTACCAATGCTTAATCAGTGAGGC
CstA/AmpR_rev	TTCGCCAACATCCCCCCCTCACTCTGACTTTAGTGTGCGCCTTTTG CCTG GTGCGCGGAACCCCTATTT
BtsT/SpecR_for	TAGAACGGCTTCGGCCAACTATTAATCAATACATGCCAGGTTTTA CTATG TCCGCTCATGAGACAATAACCCT
BtsT/SpecR_rev	AAGCCCCGCCGAAGCGGGGCTAAACACGGTTAGTGGTGCGAAG AGATCTT ACTCACGTTAAGGGATTTTGGTCA
YnfM/KanR_FRT_for	GGGTATTTACCGCAGTCCGGACTTATTTTTCAGGCGTGCAGACG ACGATGTAA TGTAGGCTGGAGCTGCTTCG
YnfM/KanR_FRT_rev	CCCTATGTATAAGCCTGATCTACAGGCATATTTAGCAAGGATTTC AAGTG ATTCCGGGGATCCGTCGACC
YhjX/KanR_FRT_for	GGTTGTCGGCAGAGATTTTTCCTTTTTATTACTGCAGGAATACTG CCATGTAA TGTAGGCTGGAGCTGCTTCG
YhjX/KanR_FRT_rev	GCTGAAGCCCAGTAGCTCGCGGCTGAGCATTAAAGGGAGCCAT GCGCCTC ATTCCGGGGATCCGTCGACC

Table A.88: Genotyping primers used to confirm *CstA* → *AmpR* mutation

Primer ID	Primer sequence (5'-3')
CstA_geno_for	GGACACGCACACGGATAACAAC
CstA_geno_rev	ACTCTGACTTTAGTGTGCGCCT
AmpR_geno_for	ATTGTTGCCGGGAAGCTAGAGT

Table A.89: Genotyping primers used to confirm *BtsT* → *SpecR* mutation

Primer ID	**Primer sequence (5'-3')**
BtsT_geno_for	CGGCCAACTATTAATCAATACATGCC
BtsT_geno_rev	ACGGTTAGTGGTGCGAAGAGAT
SpecR_geno_for	ACATAGCGTTGCCTTGGTAGGT

Table A.90: Genotyping primers used to confirm *KanR* in place of *YfeH* removal

Primer ID	**Primer sequence (5'-3')**
YfeH_geno_for	CTGTCCAGCAGGCGAACACCAG
YfeH_geno_rev	GCTGCGTTTGCTGGGTAGCTGA

Table A.91: Genotyping primers used to confirm *YnfM* → *KanR* mutation

Primer ID	**Primer sequence (5'-3')**
YnfM_geno_for	CCGGACTTATTTTTCAGGCGTGCA
YnfM_geno_rev	TGATCTACAGGCATATTTAGCAAGGA
KanR_geno_rev	GCCCATTCGACCACCAAGCGAA

Table A.92: Genotyping primers used to confirm *KanR* in place of *YnfM* removal

Primer ID	**Primer sequence (5'-3')**
YnfM_geno_for	See Table A.34
YnfM_geno_rev	See Table A.34

Table A.93: Genotyping primers used to confirm *YhjX* → *KanR* mutation

Primer ID	**Primer sequence (5'-3')**
YhjX_geno_for	GGTTGTCGGCAGAGATTTTTCCT
YhjX_geno_for	GCATTAAAGGGAGCCATGCGCC
KanR_geno_rev	See Table A.34

A.10.2 Colony PCR gel electrophoresis to confirm pyruvate transporter knock-out / antibiotic resistance cassette removal

Figure A.2: Schematic representation of PCR-based mutation verification strategy

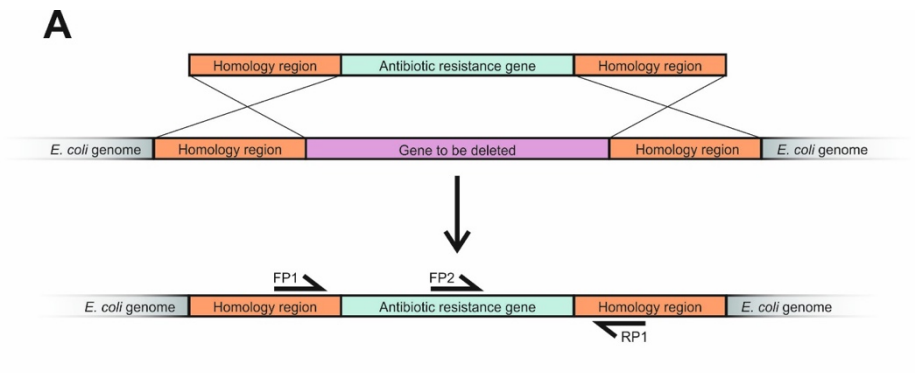


Figure A.3: Confirmation of *CstA* → *AmpR* mutation in 2m.

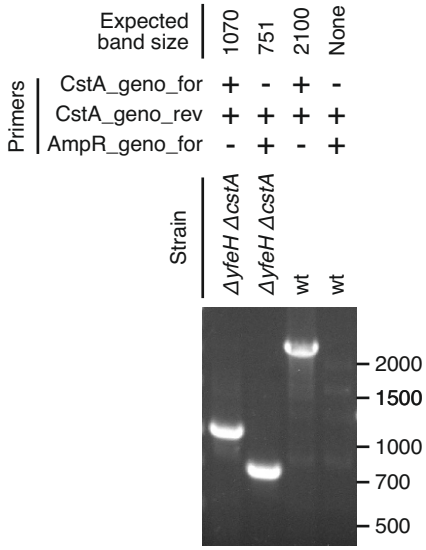


Figure A.4: Confirmation of *BtsT* → *SpecR* mutation in 3m.

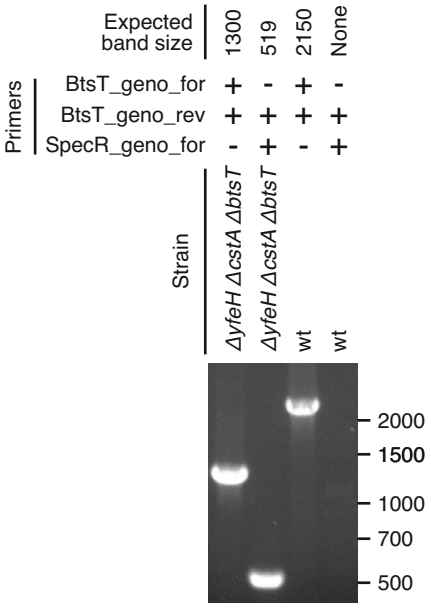


Figure A.5: Confirmation of *KanR* in place of *YfeH* removal.

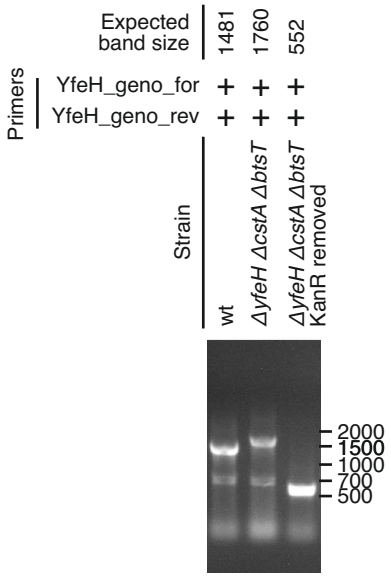


Figure A.6: Confirmation of *YnfM* → KanR mutation in 4m.

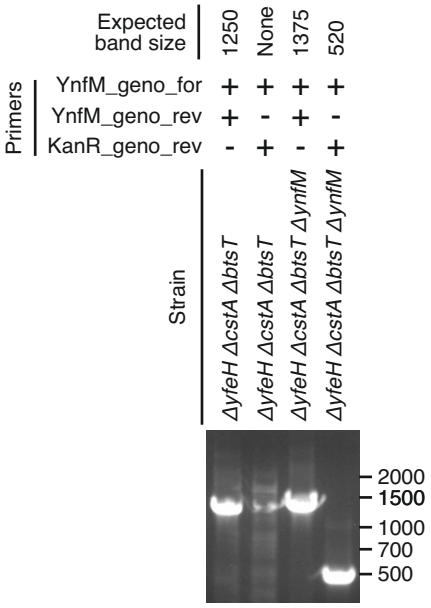


Figure A.7: Confirmation of *KanR* in place of *YnfM* removal.

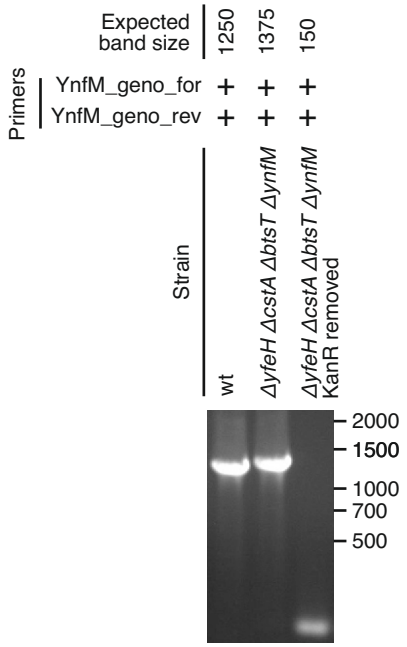
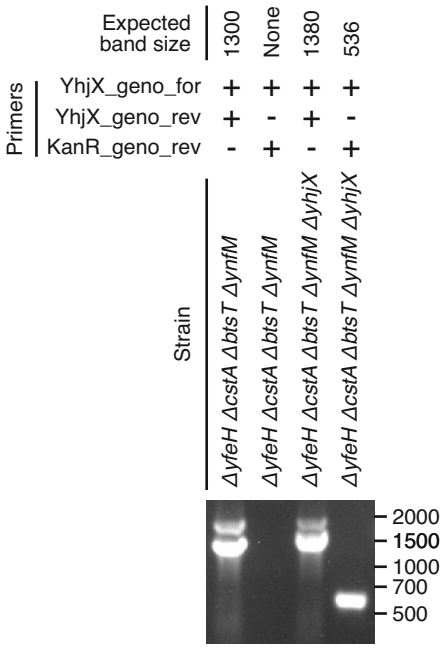


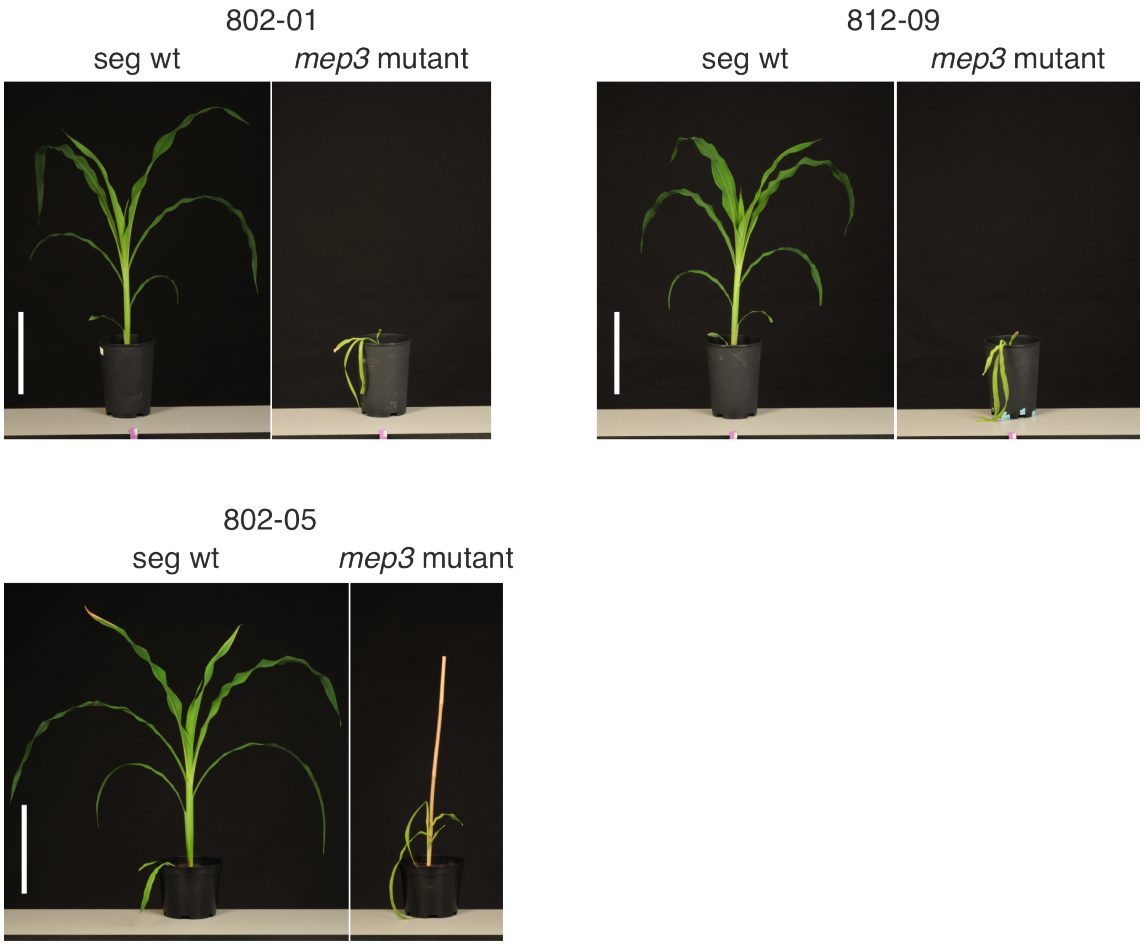
Figure A.8: Confirmation of *YhjX* → KanR mutation in 5m.



A.11 *Zea mays mep3* knock-out

A.11.1 Photos of *mep3* and segregating wild-type plants

Figure A.9: Photos of *mep3* and segregating wild-type plants from all three independent lines in T_2 generation. Scale bar = 20 cm.



A.12 *Zea mays* genotyping

A.12.1 PCR primers used to confirm segregating wild-type, heterozygous and homozygous mutants

Table A.94: Primers used to amplify mutated region of *HATCH3.1*

Primer ID	Primer sequence (5'-3')
ZmHATCH3.1_for	GTATGGCTGCTTCCGCTCCAAG
ZmHATCH3.1_rev	CAACGCTGACGACGATGATGCT

A.13 *Zea mays mpa3.1* knock-out

A.13.1 Photos of *mpa3.1* and segregating wild-type plants

Figure A.10: Photos of *mpa3.1* and segregating wild-type plants from two independent lines in T_3 generation. Scale bar = 20 cm.



Bibliography

1. Ehleringer, J. R., Sage, R. F., Flanagan, L. B. & Pearcy, R. W. Climate change and the evolution of C_4 photosynthesis. *Trends Ecol. Evol.* **6**, 95–99 (1991).
2. Peterhansel, C. *et al.* Photorespiration. *Arab. Book Am. Soc. Plant Biol.* **8**, (2010).
3. Walker, B. J., VanLoocke, A., Bernacchi, C. J. & Ort, D. R. The Costs of Photorespiration to Food Production Now and in the Future. *Annu. Rev. Plant Biol.* **67**, 107–129 (2016).
4. Dusenge, M. E., Duarte, A. G. & Way, D. A. Plant carbon metabolism and climate change: elevated CO_2 and temperature impacts on photosynthesis, photorespiration and respiration. *New Phytol.* **221**, 32–49 (2019).
5. Edwards, E. J. Evolutionary trajectories, accessibility and other metaphors: the case of C_4 and CAM photosynthesis. *New Phytol.* **223**, 1742–1755 (2019).
6. Heyduk, K., Moreno-Villena, J. J., Gilman, I. S., Christin, P.-A. & Edwards, E. J. The genetics of convergent evolution: insights from plant photosynthesis. *Nat. Rev. Genet.* **20**, 485–493 (2019).
7. Keeley, J. E. & Rundel, P. W. Evolution of CAM and C_4 Carbon-Concentrating Mechanisms. *Int. J. Plant Sci.* **164**, S55–S77 (2003).
8. Sage, R. F. A portrait of the C_4 photosynthetic family on the 50th anniversary of its discovery: species number, evolutionary lineages, and Hall of Fame. *J. Exp. Bot.* **67**, 4039–4056 (2016).
9. Sharwood, R. E., Sonawane, B. V., Ghannoum, O. & Whitney, S. M. Improved analysis of C_4 and C_3 photosynthesis via refined in vitro assays of their carbon fixation biochemistry. *J. Exp. Bot.* **67**, 3137–3148 (2016).
10. Winter, K. & Smith, J. A. C. CAM photosynthesis: the acid test. *New Phytol.* **233**, 599–609 (2022).
11. Yamori, W., Hikosaka, K. & Way, D. A. Temperature response of photosynthesis in C_3, C_4, and CAM plants: temperature acclimation and temperature adaptation. *Photosynth. Res.* **119**, 101–117 (2014).

12. Edwards, G. E., Franceschi, V. R. & Voznesenskaya, E. V. Single-cell C_4 photosynthesis versus the dual-cell (Kranz) paradigm. *Annu. Rev. Plant Biol.* **55**, 173–196 (2004).
13. von Caemmerer, S. & Furbank, R. T. The C_4 pathway: an efficient CO_2 pump. *Photosynth. Res.* **77**, 191–207 (2003).
14. Evans, J. R. & von Caemmerer, S. Would C_4 rice produce more biomass than C3 rice?**Sheehy JE, Mitchell PL, Hardy B, editors. 2000. Redesigning rice photosynthesis to increase yield. Proceedings of the Workshop on The Quest to Reduce Hunger: Redesigning Rice Photosynthesis, 30 Nov.-3 Dec. 1999, Los Baños, Philippines. Makati City (Philippines): International Rice Research Institute and Amsterdam (The Netherlands): Elsevier Science B.V. 293 p. in *Studies in Plant Science* (eds. Sheehy, J. E., Mitchell, P. L. & Hardy, B.) vol. 7 53–71 (Elsevier, 2000).
15. Sage, R. F., Pearcy, R. W. & Seemann, J. R. The Nitrogen Use Efficiency of C_3 and C_4 Plants : III. Leaf Nitrogen Effects on the Activity of Carboxylating Enzymes in Chenopodium album (L.) and Amaranthus retroflexus (L.). *Plant Physiol.* **85**, 355–359 (1987).
16. Way, D. A., Katul, G. G., Manzoni, S. & Vico, G. Increasing water use efficiency along the C_3 to C_4 evolutionary pathway: a stomatal optimization perspective. *J. Exp. Bot.* **65**, 3683–3693 (2014).
17. Hatch, M. D., Kagawa, T. & Craig, S. Subdivision of C_4-Pathway Species Based on Differing C_4 Acid Decarboxylating Systems and Ultrastructural Features. *Funct. Plant Biol.* **2**, 111–128 (1975).
18. Furbank, R. T. Evolution of the C_4 photosynthetic mechanism: are there really three C_4 acid decarboxylation types? *J. Exp. Bot.* **62**, 3103–3108 (2011).
19. Wang, Y., Bräutigam, A., Weber, A. P. M. & Zhu, X.-G. Three distinct biochemical subtypes of C_4 photosynthesis? A modelling analysis. *J. Exp. Bot.* **65**, 3567–3578 (2014).
20. Wingler, A., Walker, R. P., Chen, Z.-H. & Leegood, R. C. Phosphoenolpyruvate Carboxykinase Is Involved in the Decarboxylation of Aspartate in the Bundle Sheath of Maize. *Plant Physiol.* **120**, 539–546 (1999).

21. Facchinelli, F. & Weber, A. The Metabolite Transporters of the Plastid Envelope: An Update. *Front. Plant Sci.* **2**, (2011).
22. Linka, N. & Weber, A. P. M. Intracellular Metabolite Transporters in Plants. *Mol. Plant* **3**, 21–53 (2010).
23. Weber, A. P. & von Caemmerer, S. Plastid transport and metabolism of C_3 and C_4 plants—comparative analysis and possible biotechnological exploitation. *Curr. Opin. Plant Biol.* **13**, 256–264 (2010).
24. Majeran, W. *et al.* Consequences of C_4 Differentiation for Chloroplast Membrane Proteomes in Maize Mesophyll and Bundle Sheath Cells. *Mol. Cell. Proteomics MCP* **7**, 1609–1638 (2008).
25. Woo, K. C. *et al.* Deficient Photosystem II in Agranal Bundle Sheath Chloroplasts of C_4 Plants. *Proc. Natl. Acad. Sci. U. S. A.* **67**, 18–25 (1970).
26. Bräutigam, A. *et al.* Biochemical mechanisms driving rapid fluxes in C_4 photosynthesis. *bioRxiv* 387431 (2018) doi:10.1101/387431.
27. Lu, M. Structure and Mechanism of the Divalent Anion/Na^+ Symporter. *Int. J. Mol. Sci.* **20**, 440 (2019).
28. Renné, P. *et al.* The Arabidopsis mutant dct is deficient in the plastidic glutamate/malate translocator DiT2. *Plant J.* **35**, 316–331 (2003).
29. Taniguchi, M. *et al.* Identifying and Characterizing Plastidic 2-Oxoglutarate/Malate and Dicarboxylate Transporters in Arabidopsis thaliana. *Plant Cell Physiol.* **43**, 706–717 (2002).
30. Taniguchi, Y. *et al.* Differentiation of Dicarboxylate Transporters in Mesophyll and Bundle Sheath Chloroplasts of Maize. *Plant Cell Physiol.* **45**, 187–200 (2004).
31. Heldt, H. W. & Rapley, L. Specific transport of inorganic phosphate, 3-phosphoglycerate and dihydroxyacetonephosphate, and of dicarboxylates across the inner membrane of spinach chloroplasts. *FEBS Lett.* **10**, 143–148 (1970).
32. Lehner, K. & Heldt, H. W. Dicarboxylate transport across the inner membrane of the chloroplast envelope. *Biochim. Biophys. Acta* **501**, 531–544 (1978).

33. Woo, K. C., Flügge, U. I. & Heldt, H. W. A Two-Translocator Model for the Transport of 2-Oxoglutarate and Glutamate in Chloroplasts during Ammonia Assimilation in the Light 1. *Plant Physiol.* **84**, 624–632 (1987).
34. Dry, I. B. & Wiskich, J. T. Characterization of Dicarboxylate Stimulation of Ammonia, Glutamine, and 2-Oxoglutarate-Dependent O_2 Evolution in Isolated Pea Chloroplasts. *Plant Physiol.* **72**, 291–296 (1983).
35. Woo, K. C. & Osmond, C. B. Stimulation of Ammonia and 2-Oxoglutarate-Dependent O2 Evolution in Isolated Chloroplasts by Dicarboxylates and the Role of the Chloroplast in Photorespiratory Nitrogen Recycling. *Plant Physiol.* **69**, 591–596 (1982).
36. Anderson, J. W. & Walker, D. A. Oxygen evolution by a reconstituted spinach chloroplast system in the presence of l-glutamine and 2-oxoglutarate. *Planta* **159**, 77–83 (1983).
37. Menzlaff, E. & Flügge, U.-I. Purification and functional reconstitution of the 2-oxoglutarate/malate translocator from spinach chloroplasts. *Biochim. Biophys. Acta BBA - Biomembr.* **1147**, 13–18 (1993).
38. Weber, A. *et al.* The 2-oxoglutarate/malate translocator of chloroplast envelope membranes: molecular cloning of a transporter containing a 12-helix motif and expression of the functional protein in yeast cells. *Biochemistry* **34**, 2621–2627 (1995).
39. Flügge, I. U., Woo, K. C. & Heldt, H. W. Characteristics of 2-oxoglutarate and glutamate transport in spinach chloroplasts. *Planta* **174**, 534–541 (1988).
40. Somerville, S. C. & Ogren, W. L. An Arabidopsis thaliana mutant defective in chloroplast dicarboxylate transport. *Proc. Natl. Acad. Sci. U. S. A.* **80**, 1290 (1983).
41. Kinoshita, H. *et al.* The chloroplastic 2-oxoglutarate/malate transporter has dual function as the malate valve and in carbon/nitrogen metabolism. *Plant J.* **65**, 15–26 (2011).
42. Schneidereit, J., Häusler, R. E., Fiene, G., Kaiser, W. M. & Weber, A. P. M. Antisense repression reveals a crucial role of the plastidic 2-oxoglutarate/malate translocator

DiT1 at the interface between carbon and nitrogen metabolism. *Plant J.* **45**, 206–224 (2006).

43. Taniguchi, M. & Miyake, H. Redox-shuttling between chloroplast and cytosol: integration of intra-chloroplast and extra-chloroplast metabolism. *Curr. Opin. Plant Biol.* **15**, 252–260 (2012).
44. Scheibe, R. Malate valves to balance cellular energy supply. *Physiol. Plant.* **120**, 21–26 (2004).
45. Hatch, M. D., Dröscher, L., Flügge, U. I. & Heldt, H. W. A specific translocator for oxaloacetate transport in chloroplasts. *FEBS Lett.* **178**, 15–19 (1984).
46. Eisenhut, M., Pick, T. R., Bordych, C. & Weber, A. P. M. Towards closing the remaining gaps in photorespiration – the essential but unexplored role of transport proteins. *Plant Biol.* **15**, 676–685 (2013).
47. Shameer, S., Ratcliffe, R. G. & Sweetlove, L. J. Leaf Energy Balance Requires Mitochondrial Respiration and Export of Chloroplast NADPH in the Light. *Plant Physiol.* **180**, 1947–1961 (2019).
48. Chang, Y.-M. *et al.* Characterizing Regulatory and Functional Differentiation between Maize Mesophyll and Bundle Sheath Cells by Transcriptomic Analysis. *Plant Physiol.* **160**, 165–177 (2012).
49. Friso, G., Majeran, W., Huang, M., Sun, Q. & Wijk, K. J. van. Reconstruction of Metabolic Pathways, Protein Expression, and Homeostasis Machineries across Maize Bundle Sheath and Mesophyll Chloroplasts: Large-Scale Quantitative Proteomics Using the First Maize Genome Assembly. *Plant Physiol.* **152**, 1219–1250 (2010).
50. Li, P. *et al.* The developmental dynamics of the maize leaf transcriptome. *Nat. Genet.* **42**, 1060–1067 (2010).
51. Bräutigam, A., Hoffmann-Benning, S. & Weber, A. P. M. Comparative Proteomics of Chloroplast Envelopes from C_3 and C_4 Plants Reveals Specific Adaptations of the Plastid Envelope to C_4 Photosynthesis and Candidate Proteins Required for Maintaining C_4 Metabolite Fluxes. *Plant Physiol.* **148**, 568–579 (2008).

52. Weissmann, S. *et al.* Interactions of C_4 Subtype Metabolic Activities and Transport in Maize Are Revealed through the Characterization of DCT2 Mutants. *Plant Cell* **28**, 466–484 (2016).
53. Chapman, K. S. R. & Hatch, M. D. Aspartate stimulation of malate decarboxylation in Zea mays bundle sheath cells: Possible role in regulation of C_4 photosynthesis. *Biochem. Biophys. Res. Commun.* **86**, 1274–1280 (1979).
54. Jenkins, C. L. D. & Boag, S. Isolation of Bundle Sheath Cell Chloroplasts from the NADP-ME Type C_4 Plant Zea mays: Capacities for CO_2 Assimilation and Malate Decarboxylation. *Plant Physiol.* **79**, 84–89 (1985).
55. Emms, D. M., Covshoff, S., Hibberd, J. M. & Kelly, S. Independent and Parallel Evolution of New Genes by Gene Duplication in Two Origins of C_4 Photosynthesis Provides New Insight into the Mechanism of Phloem Loading in C_4 Species. *Mol. Biol. Evol.* **33**, 1796–1806 (2016).
56. John, C. R., Smith-Unna, R. D., Woodfield, H., Covshoff, S. & Hibberd, J. M. Evolutionary convergence of cell-specific gene expression in independent lineages of C_4 grasses. *Plant Physiol.* **165**, 62–75 (2014).
57. Weissmann, S. *et al.* DCT4—A New Member of the Dicarboxylate Transporter Family in C_4 Grasses. *Genome Biol. Evol.* **13**, evaa251 (2021).
58. Bräutigam, A. *et al.* An mRNA Blueprint for C_4 Photosynthesis Derived from Comparative Transcriptomics of Closely Related C_3 and C_4 Species. *Plant Physiol.* **155**, 142–156 (2011).
59. Rao, X. *et al.* Comparative cell-specific transcriptomics reveals differentiation of C_4 photosynthesis pathways in switchgrass and other C_4 lineages. *J. Exp. Bot.* **67**, 1649–1662 (2016).
60. Jack, D. L., Yang, N. M. & H. Saier Jr, M. The drug/metabolite transporter superfamily. *Eur. J. Biochem.* **268**, 3620–3639 (2001).
61. Fliege, R., Flügge, U.-I., Werdan, K. & Heldt, H. W. Specific transport of inorganic phosphate, 3-phosphoglycerate and triosephosphates across the inner membrane of the envelope in spinach chloroplasts. *Biochim. Biophys. Acta BBA - Bioenerg.* **502**, 232–247 (1978).

62. Häusler, R. E., Schlieben, N. H., Schulz, B. & Flügge, U.-I. Compensation of decreased triose phosphate/phosphate translocator activity by accelerated starch turnover and glucose transport in transgenic tobacco. *Planta* **204**, 366–376 (1998).
63. Riesmeier, J. W. *et al.* Antisense repression of the chloroplast triose phosphate translocator affects carbon partitioning in transgenic potato plants. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 6160–6164 (1993).
64. Schneider, A. *et al.* An Arabidopsis thaliana knock-out mutant of the chloroplast triose phosphate/phosphate translocator is severely compromised only when starch synthesis, but not starch mobilisation is abolished. *Plant J.* **32**, 685–699 (2002).
65. Heineke, D. *et al.* Effect of antisense repression of the chloroplast triose-phosphate translocator on photosynthetic metabolism in transgenic potato plants. *Planta* **193**, 174–180 (1994).
66. Gowik, U., Bräutigam, A., Weber, K. L., Weber, A. P. M. & Westhoff, P. Evolution of C_4 Photosynthesis in the Genus Flaveria: How Many and Which Genes Does It Take to Make C_4? *Plant Cell* **23**, 2087–2105 (2011).
67. Koteyeva, N. K., Voznesenskaya, E. V., Cousins, A. B. & Edwards, G. E. Differentiation of C_4 photosynthesis along a leaf developmental gradient in two Cleome species having different forms of Kranz anatomy. *J. Exp. Bot.* **65**, 3525–3541 (2014).
68. Flügge, U. I. & Weber, A. A rapid method for measuring organelle-specific substrate transport in homogenates from plant tissues. *Planta* **194**, 181–185 (1994).
69. Borchert, S., Harborth, J., Schunemann, D., Hoferichter, P. & Heldt, H. W. Studies of the Enzymic Capacities and Transport Properties of Pea Root Plastids. *Plant Physiol.* **101**, 303–312 (1993).
70. Neuhaus, H. E., Batz, O., Thom, E. & Scheibe, R. Purification of highly intact plastids from various heterotrophic plant tissues: analysis of enzymic equipment and precursor dependency for starch biosynthesis. *Biochem. J.* **296**, 395–401 (1993).

71. Kammerer, B. *et al.* Molecular Characterization of a Carbon Transporter in Plastids from Heterotrophic Tissues: The Glucose 6-Phosphate/Phosphate Antiporter. *Plant Cell* **10**, 105–117 (1998).
72. Leakey, A. D. B. *et al.* Genomic basis for stimulated respiration by plants growing under elevated carbon dioxide. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 3597–3602 (2009).
73. Kunz, H. H. *et al.* The role of plastidial glucose-6-phosphate/phosphate translocators in vegetative tissues of Arabidopsis thaliana mutants impaired in starch biosynthesis. *Plant Biol.* **12**, 115–128 (2010).
74. Lloyd, J. C. & Zakhleniuk, O. V. Responses of primary and secondary metabolism to sugar accumulation revealed by microarray expression analysis of the Arabidopsis mutant, pho3. *J. Exp. Bot.* **55**, 1221–1230 (2004).
75. Dyson, B. C. *et al.* Acclimation of metabolism to light in A rabidopsis thaliana: the glucose 6-phosphate/phosphate translocator GPT2 directs metabolic acclimation. *Plant Cell Environ.* **38**, 1404–1417 (2015).
76. Weise, S. E. *et al.* Transcriptional Regulation of the Glucose-6-Phosphate/Phosphate Translocator 2 Is Related to Carbon Exchange Across the Chloroplast Envelope. *Front. Plant Sci.* **10**, (2019).
77. Sharkey, T. D. & Weise, S. E. The glucose 6-phosphate shunt around the Calvin–Benson cycle. *J. Exp. Bot.* **67**, 4067–4077 (2016).
78. Eicks, M., Maurino, V., Knappe, S., Flügge, U.-I. & Fischer, K. The Plastidic Pentose Phosphate Translocator Represents a Link between the Cytosolic and the Plastidic Pentose Phosphate Pathways in Plants. *Plant Physiol.* **128**, 512–522 (2002).
79. Fischer, K. *et al.* A new class of plastidic phosphate translocators: a putative link between primary and secondary metabolism by the phosphoenolpyruvate/phosphate antiporter. *Plant Cell* **9**, 453–462 (1997).
80. Knappe, S. *et al.* Characterization of two functional phosphoenolpyruvate/phosphate translocator (PPT) genes in Arabidopsis–AtPPT1 may be involved in the provision of signals for correct mesophyll development. *Plant J.* **36**, 411–420 (2003).

81. Voll, L. *et al.* The phenotype of the Arabidopsis cue1 mutant is not simply caused by a general restriction of the shikimate pathway. *Plant J.* **36**, 301–317 (2003).
82. Li, Hm., Culligan, K., Dixon, R. A. & Chory, J. CUE1: A Mesophyll Cell-Specific Positive Regulator of Light-Controlled Gene Expression in Arabidopsis. *Plant Cell* **7**, 1599–1610 (1995).
83. Streatfield, S. J. *et al.* The Phosphoenolpyruvate/Phosphate Translocator Is Required for Phenolic Metabolism, Palisade Cell Development, and Plastid-Dependent Nuclear Gene Expression. *Plant Cell* **11**, 1609–1621 (1999).
84. Tamagnone, L. *et al.* Inhibition of Phenolic Acid Metabolism Results in Precocious Cell Death and Altered Cell Morphology in Leaves of Transgenic Tobacco Plants. *Plant Cell* **10**, 1801–1816 (1998).
85. Elkind, Y. *et al.* Abnormal plant development and down-regulation of phenylpropanoid biosynthesis in transgenic tobacco containing a heterologous phenylalanine ammonia-lyase gene. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 9057–9061 (1990).
86. Lyu, M.-J. A. *et al.* What Matters for C_4 Transporters: Evolutionary Changes of Phosphoenolpyruvate Transporter for C_4 Photosynthesis. *Front. Plant Sci.* **11**, (2020).
87. Bräutigam, A., Schliesky, S., Külahoglu, C., Osborne, C. P. & Weber, A. P. M. Towards an integrative model of C_4 photosynthetic subtypes: insights from comparative transcriptome analysis of NAD-ME, NADP-ME, and PEP-CK C4 species. *J. Exp. Bot.* **65**, 3579–3593 (2014).
88. Mansour, N. M., Sawhney, M., Tamang, D. G., Vogl, C. & Saier, M. H. The bile/arsenite/riboflavin transporter (BART) superfamily. *FEBS J.* **274**, 612–629 (2007).
89. Myo, T. *et al.* Genome-wide identification of the BASS gene family in four Gossypium species and functional characterization of GhBASSs against salt stress. *Sci. Rep.* **11**, 11342 (2021).

90. Huber, S. C. & Edwards, G. E. Transport in C_4 mesophyll chloroplasts. Evidence for an exchange of inorganic phosphate and phosphoenolpyruvate. *Biochim. Biophys. Acta BBA - Bioenerg.* **462**, 603–612 (1977).
91. Flügge, U., Stitt, M. & Heldt, H. W. Light-driven uptake of pyruvate into mesophyll chloroplasts from maize. *FEBS Lett.* **183**, 335–339 (1985).
92. Ohnishi, J. & Kanai, R. Light-Dependent Uptake of Pyruvate by Mesophyll Chloroplasts of a C_4 Plant, Panicum miliaceum L. *Plant Cell Physiol.* **28**, 243–251 (1987).
93. Ohnishi, J. & Kanai, R. Pyruvate Uptake by Mesophyll and Bundle Sheath Chloroplasts of a C_4 Plant, Panicum miliaceum L. *Plant Cell Physiol.* **28**, 1–10 (1987).
94. Aoki, N., Ohnishi, J. & Kanai, R. Two Different Mechanisms for Transport of Pyruvate into Mesophyll Chloroplasts of C_4 Plants—a Comparative Study. *Plant Cell Physiol.* **33**, 805–809 (1992).
95. Grass Phylogeny Working Group II. New grass phylogeny resolves deep evolutionary relationships and discovers C_4 origins. *New Phytol.* **193**, 304–312 (2012).
96. Ohnishi, J. & Kanai, R. Pyruvate uptake induced by a pH jump in mesophyll chloroplasts of maize and sorghum, NADP-malic enzyme type C_4 species. *FEBS Lett.* **269**, 122–124 (1990).
97. Ohnishi, J., Flügge, U.-I., Heldt, H. W. & Kanai, R. Involvement of Na^+ in Active Uptake of Pyruvate in Mesophyll Chloroplasts of Some C_4 Plants 1. *Plant Physiol.* **94**, 950–959 (1990).
98. Ohnishi, J. & Kanai, R. Na^+-induced uptake of pyruvate into mesophyll chloroplasts of a C_4 plant, Panicum miliaceum. *FEBS Lett.* **219**, 347–350 (1987).
99. Furumoto, T. *et al.* A plastidial sodium-dependent pyruvate transporter. *Nature* **476**, 472–475 (2011).
100. Cellier, F. *et al.* Characterization of AtCHX17, a member of the cation/H^+ exchangers, CHX family, from Arabidopsis thaliana suggests a role in K^+ homeostasis. *Plant J. Cell Mol. Biol.* **39**, 834–846 (2004).

101. Zhao, Y., Ai, X., Wang, M., Xiao, L. & Xia, G. A putative pyruvate transporter TaBASS2 positively regulates salinity tolerance in wheat via modulation of ABI4 expression. *BMC Plant Biol.* **16**, (2016).
102. Huang, L. *et al.* A plastidial pantoate transporter with a potential role in pantothenate synthesis. *Biochem. J.* **475**, 813–825 (2018).
103. South, P. F. *et al.* Bile Acid Sodium Symporter BASS6 Can Transport Glycolate and Is Involved in Photorespiratory Metabolism in Arabidopsis thaliana. *Plant Cell* **29**, 808–823 (2017).
104. Pick, T. R. *et al.* PLGG1, a plastidic glycolate glycerate transporter, is required for photorespiration and defines a unique class of metabolite transporters. *Proc. Natl. Acad. Sci.* **110**, 3185–3190 (2013).
105. Kuhnert, F. The Role of Organellar Transporters in C_3/C_4-Metabolism. PhD thesis. Heinrich Heine Universität Düsseldorf. University and State Library Publication service: https://docserv.uni-duesseldorf.de/servlets/DocumentServlet?id=54267. (2020).
106. Halestrap, A. P. The mitochondrial pyruvate carrier. Kinetics and specificity for substrates and inhibitors. *Biochem. J.* **148**, 85–96 (1975).
107. Herzig, S. *et al.* Identification and Functional Expression of the Mitochondrial Pyruvate Carrier. *Science* **337**, 93–96 (2012).
108. Halestrap, A. P. & Denton, R. M. Specific inhibition of pyruvate transport in rat liver mitochondria and human erythrocytes by alpha-cyano-4-hydroxycinnamate. *Biochem. J.* **138**, 313–316 (1974).
109. Bricker, D. K. *et al.* A Mitochondrial Pyruvate Carrier Required for Pyruvate Uptake in Yeast, Drosophila, and Humans. *Science* **337**, 96–100 (2012).
110. Li, C.-L., Wang, M., Ma, X.-Y. & Zhang, W. NRGA1, a Putative Mitochondrial Pyruvate Carrier, Mediates ABA Regulation of Guard Cell Ion Channels and Drought Stress Responses in Arabidopsis. *Mol. Plant* **7**, 1508–1521 (2014).
111. Shen, J.-L. *et al.* Mitochondrial pyruvate carrier 1 mediates abscisic acid-regulated stomatal closure and the drought response by affecting cellular pyruvate content in Arabidopsis thaliana. *BMC Plant Biol.* **17**, (2017).

112. He, L. *et al.* Mitochondrial Pyruvate Carriers Prevent Cadmium Toxicity by Sustaining the TCA Cycle and Glutathione Synthesis1. *Plant Physiol.* **180**, 198–211 (2019).
113. Le, X. H., Lee, C.-P. & Millar, A. H. The mitochondrial pyruvate carrier (MPC) complex mediates one of three pyruvate-supplying pathways that sustain Arabidopsis respiratory metabolism. *Plant Cell* **33**, 2776–2793 (2021).
114. Aubry, S., Kelly, S., Kümpers, B. M. C., Smith-Unna, R. D. & Hibberd, J. M. Deep Evolutionary Comparison of Gene Expression Identifies Parallel Recruitment of Trans-Factors in Two Independent Origins of C4 Photosynthesis. *PLoS Genet.* **10**, e1004365 (2014).
115. Haferkamp, I. & Schmitz-Esser, S. The Plant Mitochondrial Carrier Family: Functional and Evolutionary Aspects. *Front. Plant Sci.* **3**, (2012).
116. Palmieri, L. *et al.* Molecular identification of three Arabidopsis thaliana mitochondrial dicarboxylate carrier isoforms: organ distribution, bacterial expression, reconstitution into liposomes and functional characterization. *Biochem. J.* **410**, 621–629 (2008).
117. DeSantis, A., Arrigoni, O. & Palmieri, F. Carrier-mediated transport of metabolites in purified bean mitochondria. *Plant Cell Physiol.* **17**, 1221–1233 (1976).
118. Day, D. A. & Hanson, J. B. Effect of Phosphate and Uncouplers on Substrate Transport and Oxidation by Isolated Corn Mitochondria 1. *Plant Physiol.* **59**, 139–144 (1977).
119. Furbank, R. T., Agostino, A. & Hatch, M. D. C_4 acid decarboxylation and photosynthesis in bundle sheath cells of NAD-malic enzyme-type C_4 plants: Mechanism and the role of malate and orthophosphate. *Arch. Biochem. Biophys.* **276**, 374–381 (1990).
120. Kagawa, T. & Hatch, M. D. Mitochondria as a site of C_4 acid decarboxylation in C_4-pathway photosynthesis. *Arch. Biochem. Biophys.* **167**, 687–696 (1975).
121. Hamel, P. *et al.* Redundancy in the function of mitochondrial phosphate transport in Saccharomyces cerevisiae and Arabidopsis thaliana. *Mol. Microbiol.* **51**, 307–317 (2004).

122. Picault, N., Palmieri, L., Pisano, I., Hodges, M. & Palmieri, F. Identification of a Novel Transporter for Dicarboxylates and Tricarboxylates in Plant Mitochondria. Bacterial expression, reconstitution, functional characterization, and tissue distribution. *J. Biol. Chem.* **277**, 24204–24211 (2002).
123. Taniguchi, M. & Sugiyama, T. Isolation, characterization and expression of cDNA clones encoding a mitochondrial malate translocator from Panicum miliaceum L. *Plant Mol. Biol.* **30**, 51–64 (1996).
124. Taniguchi, M. & Sugiyama, T. The Expression of 2-Oxoglutarate/Malate Translocator in the Bundle-Sheath Mitochondria of Panicum miliaceum, a NAD-Malic Enzyme-Type C_4 Plant, Is Regulated by Light and Development. *Plant Physiol.* **114**, 285–293 (1997).
125. Foyer, C. H., Noctor, G. & Hodges, M. Respiration and nitrogen assimilation: targeting mitochondria-associated metabolism as a means to enhance nitrogen use efficiency. *J. Exp. Bot.* **62**, 1467–1482 (2011).
126. Lancien, M., *et al.* Simultaneous expression of NAD-dependent isocitrate dehydrogenase and other krebs cycle genes after nitrate resupply to short-term nitrogen-starved tobacco. *Plant Physiol.* **120**, 717–726 (1999).
127. Scheible, W. R. *et al.* Nitrate Acts as a Signal to Induce Organic Acid Metabolism and Repress Starch Metabolism in Tobacco. *Plant Cell* **9**, 783–798 (1997).
128. Nicholls, D. G. & Locke, R. M. Thermogenic mechanisms in brown fat. *Physiol. Rev.* **64**, 1–64 (1984).
129. Borecký, J. *et al.* The plant energy-dissipating mitochondrial systems: depicting the genomic structure and the expression profiles of the gene families of uncoupling protein and alternative oxidase in monocots and dicots. *J. Exp. Bot.* **57**, 849–864 (2006).
130. Vercesi, A. E. *et al.* PUMPing plants. *Nature* **375**, 24–24 (1995).
131. Kowaltowski, A. J., Costa, A. D. T. & Vercesi, A. E. Activation of the potato plant uncoupling mitochondrial protein inhibits reactive oxygen species generation by the respiratory chain. *FEBS Lett.* **425**, 213–216 (1998).

132. Maia, I. G. *et al.* AtPUMP: an Arabidopsis gene encoding a plant uncoupling mitochondrial protein. *FEBS Lett.* **429**, 403–406 (1998).
133. Bouillaud, F., Couplan, E., Pecqueur, C. & Ricquier, D. Homologues of the uncoupling protein from brown adipose tissue (UCP1): UCP2, UCP3, BMCP1 and UCP4. *Biochim. Biophys. Acta BBA - Bioenerg.* **1504**, 107–119 (2001).
134. Pinhero, R. G., Rao, M. V., Paliyath, G., Murr, D. P. & Fletcher, R. A. Changes in Activities of Antioxidant Enzymes and Their Relationship to Genetic and Paclobutrazol-Induced Chilling Tolerance of Maize Seedlings. *Plant Physiol.* **114**, 695–704 (1997).
135. Ito, K., Abe, Y., Johnston, S. D. & Seymour, R. S. Ubiquitous expression of a gene encoding for uncoupling protein isolated from the thermogenic inflorescence of the dead horse arum Helicodiceros muscivorus. *J. Exp. Bot.* **54**, 1113–1114 (2003).
136. Sweetlove, L. J. *et al.* Mitochondrial uncoupling protein is required for efficient photosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 19587–19592 (2006).
137. Monné, M. *et al.* Uncoupling proteins 1 and 2 (UCP1 and UCP2) from Arabidopsis thaliana are mitochondrial transporters of aspartate, glutamate and dicarboxylates. *J. Biol. Chem.* jbc.RA117.000771 (2018) doi:10.1074/jbc.RA117.000771.
138. Dry, I. B., Dimitriadis, E., Ward, A. D. & Wiskich, J. T. The photorespiratory hydrogen shuttle. Synthesis of phthalonic acid and its use in the characterization of the malate/aspartate shuttle in pea (Pisum sativum) leaf mitochondria. *Biochem. J.* **245**, 669–675 (1987).
139. Bianconi, M. E. *et al.* Continued Adaptation of C_4 Photosynthesis After an Initial Burst of Changes in the Andropogoneae Grasses. *Syst. Biol.* **69**, 445–461 (2020).
140. Day, D. A. & Hatch, M. D. Dicarboxylate transport in maize mesophyll chloroplasts. *Arch. Biochem. Biophys.* **211**, 738–742 (1981).
141. Paysan-Lafosse, T. *et al.* InterPro in 2022. *Nucleic Acids Res.* gkac993 (2022) doi:10.1093/nar/gkac993.
142. Almagro Armenteros, J. J. *et al.* Detecting sequence signals in targeting peptides using deep learning. *Life Sci. Alliance* **2**, e201900429 (2019).

143. Hallgren, J. *et al.* DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. 2022.04.08.487609 Preprint at https://doi.org/10.1101/2022.04.08.487609 (2022).
144. Kelly, S. *et al.* Wide sampling of natural diversity identifies novel molecular signatures of C_4 photosynthesis. *bioRxiv* 163097 (2017) doi:10.1101/163097.
145. Fairbairn, D. J. *et al.* Characterisation of two distinct HKT1-like potassium transporters from Eucalyptus camaldulensis. *Plant Mol. Biol.* **43**, 515–525 (2000).
146. Garciadeblas, B., Benito, B. & Rodríguez-Navarro, A. Molecular cloning and functional expression in bacteria of the potassium transporters CnHAK1 and CnHAK2 of the seagrass Cymodocea nodosa. *Plant Mol. Biol.* **50**, 623–633 (2002).
147. Kim, E. J., Kwak, J. M., Uozumi, N. & Schroeder, J. I. AtKUP1: An Arabidopsis Gene Encoding High-Affinity Potassium Transport Activity. *Plant Cell* **10**, 51–62 (1998).
148. Uozumi, N. *et al.* The Arabidopsis HKT1 Gene Homolog Mediates Inward Na+ Currents in Xenopus laevis Oocytes and Na^+ Uptake in Saccharomyces cerevisiae. *Plant Physiol.* **122**, 1249–1260 (2000).
149. Uozumi, N., Nakamura, T., Schroeder, J. I. & Muto, S. Determination of transmembrane topology of an inward-rectifying potassium channel from Arabidopsis thaliana based on functional expression in Escherichia coli. *Proc. Natl. Acad. Sci.* **95**, 9773–9778 (1998).
150. Tjaden, J., Schwöppe, C., Möhlmann, T., Quick, P. W. & Neuhaus, H. E. Expression of a Plastidic ATP/ADP Transporter Gene in Escherichia coli Leads to a Functional Adenine Nucleotide Transport System in the Bacterial Cytoplasmic Membrane. *J. Biol. Chem.* **273**, 9630–9636 (1998).
151. Jeong, J. *et al.* A Nodule-Specific Dicarboxylate Transporter from Alder Is a Member of the Peptide Transporter Family. *Plant Physiol.* **134**, 969–978 (2004).
152. Kay, W. W. & Kornberg, H. L. The Uptake of C_4-Dicarboxylic Acids by Escherichia coli. *Eur. J. Biochem.* **18**, 274–281 (1971).
153. Kreth, J., Lengeler, J. W. & Jahreis, K. Characterization of Pyruvate Uptake in Escherichia coli K-12. *PLoS ONE* **8**, e67125 (2013).

154. Unden, G., Strecker, A., Kleefeld, A. & Kim, O. B. C_4-Dicarboxylate Utilization in Aerobic and Anaerobic Growth. *EcoSal Plus* **7**, (2016).
155. Davies, S. J. *et al.* Inactivation and Regulation of the Aerobic C_4-Dicarboxylate Transport (dctA) Gene of Escherichia coli. *J. Bacteriol.* **181**, 5624–5635 (1999).
156. Ferro, M. *et al.* Integral membrane proteins of the chloroplast envelope: Identification and subcellular localization of new transporters. *Proc. Natl. Acad. Sci.* **99**, 11487–11492 (2002).
157. Christian, R. W., Hewitt, S. L., Nelson, G., Roalson, E. H. & Dhingra, A. Plastid transit peptides—where do they come from and where do they all belong? Multi-genome and pan-genomic assessment of chloroplast transit peptide evolution. *PeerJ* **8**, e9772 (2020).
158. Breuers, F. *et al.* Dynamic Remodeling of the Plastid Envelope Membranes – A Tool for Chloroplast Envelope in vivo Localizations. *Front. Plant Sci.* **3**, (2012).
159. Zhang, B. *et al.* Inner Envelope CHLOROPLAST MANGANESE TRANSPORTER 1 Supports Manganese Homeostasis and Phototrophic Growth in Arabidopsis. *Mol. Plant* **11**, 943–954 (2018).
160. Karinou, E., Compton, E. L. R., Morel, M. & Javelle, A. The Escherichia coli SLC26 homologue YchM (DauA) is a C_4-dicarboxylic acid transporter. *Mol. Microbiol.* **87**, 623–640 (2013).
161. Sá-Pessoa, J. *et al.* SATP (YaaH), a succinate–acetate transporter protein in Escherichia coli. *Biochem. J.* **454**, 585–595 (2013).
162. Zientz, E., Janausch, I. G., Six, S. & Unden, G. Functioning of DcuC as the C_4-Dicarboxylate Carrier during Glucose Fermentation by Escherichia coli. *J. Bacteriol.* **181**, 3716–3720 (1999).
163. Strecker, A., Schubert, C., Zedler, S., Steinmetz, P. & Unden, G. DcuA of aerobically grown Escherichia coli serves as a nitrogen shuttle (L-aspartate/fumarate) for nitrogen uptake. *Mol. Microbiol.* **109**, 801–811 (2018).
164. Wallace, B. J. & Young, I. G. Aerobic respiration in mutants of Escherichia coli accumulating quinone analogues of ubiquinone. *Biochim. Biophys. Acta BBA - Bioenerg.* **461**, 75–83 (1977).

165. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
166. Hegedűs, T., Geisler, M., Lukács, G. L. & Farkas, B. Ins and outs of AlphaFold2 transmembrane protein structure predictions. *Cell. Mol. Life Sci.* **79**, 73 (2022).
167. Evans, R. *et al.* Protein complex prediction with AlphaFold-Multimer. 2021.10.04.463034 Preprint at https://doi.org/10.1101/2021.10.04.463034 (2022).
168. Weiner, H., Burnell, J. N., Woodrow, I. E., Heldt, H. W. & Hatch, M. D. Metabolite diffusion into bundle sheath cells from C_4 plants: relation to C_4 photosynthesis and plasmodesmatal function. *Plant Physiol.* **88**, 815–822 (1988).
169. Vélez-Bermúdez, I. C. & Schmidt, W. The conundrum of discordant protein and mRNA expression. Are plants special? *Front. Plant Sci.* **5**, (2014).
170. Orth, J. D., Thiele, I. & Palsson, B. Ø. What is flux balance analysis? *Nat. Biotechnol.* **28**, 245–248 (2010).
171. Blätke, M.-A. & Bräutigam, A. Evolution of C_4 photosynthesis predicted by constraint-based modelling. *eLife* **8**, e49305 (2019).
172. Goodstein, D. M. *et al.* Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* **40**, D1178–D1186 (2012).
173. Puigbò, P., Guzmán, E., Romeu, A. & Garcia-Vallvé, S. OPTIMIZER: a web server for optimizing the codon usage of DNA sequences. *Nucleic Acids Res.* **35**, W126-131 (2007).
174. Mölder, F. *et al.* Sustainable data analysis with Snakemake. Preprint at https://doi.org/10.12688/f1000research.29032.1 (2021).
175. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
176. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008 (2021).
177. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).

178. Karp, P. D. *et al.* The EcoCyc Database. *EcoSal Plus* **6**, 10.1128/ecosalplus.ESP-0009–2013 (2014).
179. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
180. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017).
181. Richards, A. University of Oxford Advanced Research Computing. (2015) doi:10.5281/zenodo.22558.
182. Schulz, A., Knoetzel, J., Scheller, H. V. & Mant, A. Uptake of a Fluorescent Dye as a Swift and Simple Indicator of Organelle Intactness: Import-competent Chloroplasts from Soil-grown Arabidopsis. *J. Histochem. Cytochem.* **52**, 701–704 (2004).
183. Lichtenthaler, H. K. & Buschmann, C. Chlorophylls and Carotenoids: Measurement and Characterization by UV-VIS Spectroscopy. *Curr. Protoc. Food Anal. Chem.* **1**, F4.3.1-F4.3.8 (2001).
184. Mertz, R. A. & Brutnell, T. P. Bundle sheath suberization in grass leaves: multiple barriers to characterization. *J. Exp. Bot.* **65**, 3371–3380 (2014).
185. Caemmerer, S. von. Biochemical Models of Leaf Photosynthesis. (2000) doi:10.1071/9780643103405.
186. Evans, J. R., Kaldenhoff, R., Genty, B. & Terashima, I. A. Resistances along the CO_2 diffusion pathway inside leaves. *J. Exp. Bot.* **60**, 2235–2248 (2009).
187. von Caemmerer, S. & Evans, J. R. Temperature responses of mesophyll conductance differ greatly between species. *Plant Cell Environ.* **38**, 629–637 (2015).
188. Ubierna, N., Gandin, A. & Cousins, A. B. The response of mesophyll conductance to short-term variation in CO_2 in the C_4 plants Setaria viridis and Zea mays. *J. Exp. Bot.* **69**, 1159–1170 (2018).
189. Hatch, M. D. The C4-pathway of photosynthesis. Evidence for an intermediate pool of carbon dioxide and the identity of the donor C_4-dicarboxylic acid. *Biochem. J.* **125**, 425–432 (1971).
190. Meister, M., Agostino, A. & Hatch, M. D. The roles of malate and aspartate in C_4 photosynthetic metabolism of Flaveria bidentis (L.). *Planta* **199**, 262–269 (1996).

191. Walker, R. P., Acheson, R. M., Técsi, L. I. & Leegood, R. C. Phosphoenolpyruvate Carboxykinase in C_4 Plants: Its Role and Regulation. *Funct. Plant Biol.* **24**, 459–468 (1997).
192. Muhaidat, R. & McKown, A. D. Significant involvement of PEP-CK in carbon assimilation of C_4 eudicots. *Ann. Bot.* **111**, 577–589 (2013).
193. Sommer, M., Bräutigam, A. & Weber, A. P. M. The dicotyledonous NAD malic enzyme C_4 plant Cleome gynandra displays age-dependent plasticity of C_4 decarboxylation biochemistry. *Plant Biol. Stuttg. Ger.* **14**, 621–629 (2012).
194. Burnell, J. N. & Hatch, M. D. Photosynthesis in phosphoenolpyruvate carboxykinase-type C_4 plants: Pathways of C_4 acid decarboxylation in bundle sheath cells of Urochloa panicoides. *Arch. Biochem. Biophys.* **260**, 187–199 (1988).
195. Kromdijk, J., Ubierna, N., Cousins, A. B. & Griffiths, H. Bundle-sheath leakiness in C_4 photosynthesis: a careful balancing act between CO_2 concentration and assimilation. *J. Exp. Bot.* **65**, 3443–3457 (2014).
196. Pick, T. R. *et al.* Systems Analysis of a Maize Leaf Developmental Gradient Redefines the Current C_4 Model and Provides Candidates for Regulation. *Plant Cell* **23**, 4208–4220 (2011).
197. Bellasio, C. & Griffiths, H. The Operation of Two Decarboxylases, Transamination, and Partitioning of C_4 Metabolic Processes between Mesophyll and Bundle Sheath Cells Allows Light Capture To Be Balanced for the Maize C_4 Pathway. *Plant Physiol.* **164**, 466–480 (2014).
198. Cacefo, V. *et al.* Decarboxylation mechanisms of C_4 photosynthesis in Saccharum spp.: increased PEPCK activity under water-limiting conditions. *BMC Plant Biol.* **19**, 144 (2019).
199. Sales, C. R. G. *et al.* Flexibility of C_4 decarboxylation and photosynthetic plasticity in sugarcane plants under shading. *Environ. Exp. Bot.* **149**, 34–42 (2018).
200. Sharwood, R. E., Sonawane, B. V. & Ghannoum, O. Photosynthetic flexibility in maize exposed to salinity and shade. *J. Exp. Bot.* **65**, 3715–3724 (2014).
201. Brutnell, T. P. *et al.* Setaria viridis: A Model for C_4 Photosynthesis. *Plant Cell* **22**, 2537–2544 (2010).

202. Calace, P. *et al.* The C_4 cycle and beyond: diverse metabolic adaptations accompany dual-cell photosynthetic functions in Setaria. *J. Exp. Bot.* **72**, 7876–7890 (2021).
203. de Oliveira Dal'Molin, C. G. *et al.* Metabolic Reconstruction of Setaria italica: A Systems Biology Approach for Integrating Tissue-Specific Omics and Pathway Analysis of Bioenergy Grasses. *Front. Plant Sci.* **7**, 1138 (2016).
204. Washburn, J. D. *et al.* Distinct C_4 sub-types and C_3 bundle sheath isolation in the Paniceae grasses. *Plant Direct* **5**, e373 (2021).
205. Jeggo, P. A. DNA breakage and repair. *Adv. Genet.* **38**, 185–218 (1998).
206. Jinek, M. *et al.* A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**, 816–821 (2012).
207. Boch, J. *et al.* Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. *Science* **326**, 1509–1512 (2009).
208. Kim, Y. G., Cha, J. & Chandrasegaran, S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc. Natl. Acad. Sci.* **93**, 1156–1160 (1996).
209. Adli, M. The CRISPR tool kit for genome editing and beyond. *Nat. Commun.* **9**, 1911 (2018).
210. Mohr, S. E. *et al.* CRISPR guide RNA design for research applications. *Febs J.* **283**, 3232–3238 (2016).
211. Xie, K., Minkenberg, B. & Yang, Y. Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proc. Natl. Acad. Sci.* **112**, 3570–3575 (2015).
212. Fauser, F., Schiml, S. & Puchta, H. Both CRISPR/Cas-based nucleases and nickases can be used efficiently for genome engineering in Arabidopsis thaliana. *Plant J.* **79**, 348–359 (2014).
213. Martins, P. K., Ribeiro, A. P., Cunha, B. A. D. B. da, Kobayashi, A. K. & Molinari, H. B. C. A simple and highly efficient Agrobacterium-mediated transformation protocol for Setaria viridis. *Biotechnol. Rep.* **6**, 41–44 (2015).
214. Basso, M. F. *et al.* Efficient genome editing and gene knockout in Setaria viridis with CRISPR/Cas9 directed gene editing by the non-homologous end-joining pathway. *Plant Biotechnol. Tokyo Jpn.* **38**, 227–238 (2021).

215. Kohn, M. J. Carbon isotope compositions of terrestrial C_3 plants as indicators of (paleo)ecology and (paleo)climate. *Proc. Natl. Acad. Sci.* **107**, 19691–19695 (2010).
216. O'Leary, M. H. Carbon Isotopes in Photosynthesis: Fractionation techniques may reveal new aspects of carbon dynamics in plants. *BioScience* **38**, 328–336 (1988).
217. Eggels, S., Blankenagel, S., Schön, C.-C. & Avramova, V. The carbon isotopic signature of C_4 crops and its applicability in breeding for climate resilience. *Theor. Appl. Genet.* **134**, 1663–1675 (2021).
218. von Caemmerer, S. Updating the steady-state model of C_4 photosynthesis. *J. Exp. Bot.* **72**, 6003–6017 (2021).
219. Sternberg, L. D. S. L., Deniro, M. J., Sloan, M. E. & Black, C. C. Compensation Point and Isotopic Characteristics of C_3/C_4 Intermediates and Hybrids in Panicum1. *Plant Physiol.* **80**, 242–245 (1986).
220. Lopez-Calcagno, P., Howard, T. & Raines, C. The CP12 protein family: a thioredoxin-mediated metabolic switch? *Front. Plant Sci.* **5**, (2014).
221. Tola, A. J., Jaballi, A., Germain, H. & Missihoun, T. D. Recent Development on Plant Aldehyde Dehydrogenase Enzymes and Their Functions in Plant Development and Stress Signaling. *Genes* **12**, 51 (2020).
222. Kruger, N. J. & von Schaewen, A. The oxidative pentose phosphate pathway: structure and organisation. *Curr. Opin. Plant Biol.* **6**, 236–246 (2003).
223. Wang, C., Takahashi, H. & Shikanai, T. PROTON GRADIENT REGULATION 5 contributes to ferredoxin-dependent cyclic phosphorylation in ruptured chloroplasts. *Biochim. Biophys. Acta BBA - Bioenerg.* **1859**, 1173–1179 (2018).
224. Takagi, D. & Miyake, C. PROTON GRADIENT REGULATION 5 supports linear electron flow to oxidize photosystem I. *Physiol. Plant.* **164**, 337–348 (2018).
225. Binder, S. Branched-Chain Amino Acid Metabolism in Arabidopsis thaliana. *Arab. Book Am. Soc. Plant Biol.* **8**, e0137 (2010).
226. Brogna, S. & Wen, J. Nonsense-mediated mRNA decay (NMD) mechanisms. *Nat. Struct. Mol. Biol.* **16**, 107–113 (2009).

227. Chollet, R., Vidal, J. & O'Leary, M. H. PHOSPHOENOLPYRUVATE CARBOXYLASE: A Ubiquitous, Highly Regulated Enzyme in Plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **47**, 273–298 (1996).
228. Nimmo, H. G. Control of the phosphorylation of phosphoenolpyruvate carboxylase in higher plants. *Arch. Biochem. Biophys.* **414**, 189–196 (2003).
229. Yamagami, T. *et al.* Biochemical Diversity among the 1-Amino-cyclopropane-1-Carboxylate Synthase Isozymes Encoded by the Arabidopsis Gene Family. *J. Biol. Chem.* **278**, 49102–49112 (2003).
230. Concordet, J.-P. & Haeussler, M. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Res.* **46**, W242–W245 (2018).
231. Bae, S., Park, J. & Kim, J.-S. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* **30**, 1473–1475 (2014).
232. Hofacker, I. L. Vienna RNA secondary structure server. *Nucleic Acids Res.* **31**, 3429–3431 (2003).
233. Lorenz, R. *et al.* ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **6**, 26 (2011).
234. Schindele, P., Wolter, F. & Puchta, H. CRISPR Guide RNA Design Guidelines for Efficient Genome Editing. in *RNA Tagging* (ed. Heinlein, M.) vol. 2166 331–342 (Springer US, 2020).
235. Hahn, F., Korolev, A., Sanjurjo Loures, L. & Nekrasov, V. A modular cloning toolkit for genome editing in plants. *BMC Plant Biol.* **20**, 179 (2020).
236. Boyd, R. A., Gandin, A. & Cousins, A. B. Temperature Responses of C_4 Photosynthesis: Biochemical Analysis of Rubisco, Phosphoenolpyruvate Carboxylase, and Carbonic Anhydrase in Setaria viridis. *Plant Physiol.* **169**, 1850–1861 (2015).
237. DiMario, R. J. & Cousins, A. B. A single serine to alanine substitution decreases bicarbonate affinity of phosphoenolpyruvate carboxylase in C_4 Flaveria trinervia. *J. Exp. Bot.* **70**, 995–1004 (2019).

238. Farquhar, G. D. On the Nature of Carbon Isotope Discrimination in C_4 Species. *Funct. Plant Biol.* **10**, 205–226 (1983).
239. Yin, X., van der Putten, P. E. L., Driever, S. M. & Struik, P. C. Temperature response of bundle-sheath conductance in maize leaves. *J. Exp. Bot.* **67**, 2699–2714 (2016).
240. Ubierna, N., Gandin, A., Boyd, R. A. & Cousins, A. B. Temperature response of mesophyll conductance in three C_4 species calculated with two methods: ^{18}O discrimination and in vitro V_{pmax}. *New Phytol.* **214**, 66–80 (2017).
241. Alonso-Cantabrana, H. *et al.* Diffusion of CO_2 across the Mesophyll-Bundle Sheath Cell Interface in a C_4 Plant with Genetically Reduced PEP Carboxylase Activity. *Plant Physiol.* **178**, 72–81 (2018).
242. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).
243. Goodstein, D. M. *et al.* Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* **40**, D1178–D1186 (2012).
244. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
245. Tian, F., Yang, D.-C., Meng, Y.-Q., Jin, J. & Gao, G. PlantRegMap: charting functional regulatory maps in plants. *Nucleic Acids Res.* **48**, D1104–D1113 (2020).
246. Behr, S. *et al.* Identification of a High-Affinity Pyruvate Receptor in Escherichia coli. *Sci. Rep.* **7**, 1388 (2017).
247. Kristoficova, I., Vilhena, C., Behr, S. & Jung, K. BtsT, a Novel and Specific Pyruvate/H^+ Symporter in Escherichia coli. *J. Bacteriol.* **200**, e00599-17 (2018).
248. Hwang, S. *et al.* Peptide Transporter CstA Imports Pyruvate in Escherichia coli K-12. *J. Bacteriol.* **200**, e00771-17 (2018).
249. Schultz, J. E. & Matin, A. Molecular and functional characterization of a carbon starvation gene of Escherichia coli. *J. Mol. Biol.* **218**, 129–140 (1991).
250. Gasperotti, A., Göing, S., Fajardo-Ruiz, E., Forné, I. & Jung, K. Function and Regulation of the Pyruvate Transporter CstA in Escherichia coli. *Int. J. Mol. Sci.* **21**, 9068 (2020).

251. Behr, S., Fried, L. & Jung, K. Identification of a Novel Nutrient-Sensing Histidine Kinase/Response Regulator Network in Escherichia coli. *J. Bacteriol.* **196**, 2023–2029 (2014).
252. Fried, L., Behr, S. & Jung, K. Identification of a Target Gene and Activating Stimulus for the YpdA/YpdB Histidine Kinase/Response Regulator System in Escherichia coli. *J. Bacteriol.* **195**, 807–815 (2013).
253. Zhou, S. *et al.* YhjX Regulates the Growth of Escherichia coli in the Presence of a Subinhibitory Concentration of Gentamicin and Mediates the Adaptive Resistance to Gentamicin. *Front. Microbiol.* **10**, (2019).
254. Zhao, C., Lin, Z., Dong, H., Zhang, Y. & Li, Y. Reexamination of the Physiological Role of PykA in Escherichia coli Revealed that It Negatively Regulates the Intracellular ATP Levels under Anaerobic Conditions. *Appl. Environ. Microbiol.* **83**, e00316-17 (2017).
255. Kaberdina, A. C., Ruiz-Larrabeiti, O., Lin-Chao, S. & Kaberdin, V. R. Reprogramming of gene expression in Escherichia coli cultured on pyruvate versus glucose. *Mol. Genet. Genomics* **294**, 1359–1371 (2019).
256. Koita, K. & Rao, C. V. Identification and Analysis of the Putative Pentose Sugar Efflux Transporters in Escherichia coli. *PLoS ONE* **7**, e43700 (2012).
257. Göhler, A.-K. *et al.* More than just a metabolic regulator - elucidation and validation of new targets of PdhR in Escherichia coli. *BMC Syst. Biol.* **5**, 197 (2011).
258. Gonin, S. *et al.* Crystal structures of an Extracytoplasmic Solute Receptor from a TRAP transporter in its open and closed forms reveal a helix-swapped dimer requiring a cation for α-keto acid binding. *BMC Struct. Biol.* **7**, 11 (2007).
259. Pernil, R., Herrero, A. & Flores, E. A TRAP Transporter for Pyruvate and Other Monocarboxylate 2-Oxoacids in the Cyanobacterium Anabaena sp. Strain PCC 7120. *J. Bacteriol.* **192**, 6089–6092 (2010).
260. Hosie, A. H. F., Allaway, D. & Poole, P. S. A Monocarboxylate Permease of Rhizobium leguminosarum Is the First Member of a New Subfamily of Transporters. *J. Bacteriol.* **184**, 5436–5448 (2002).

261. Felmlee, M. A., Jones, R. S., Rodriguez-Cruz, V., Follman, K. E. & Morris, M. E. Monocarboxylate Transporters (SLC16): Function, Regulation, and Role in Health and Disease. *Pharmacol. Rev.* **72**, 466–485 (2020).
262. Casal, M., Paiva, S., Andrade, R. P., Gancedo, C. & Leão, C. The Lactate-Proton Symport of Saccharomyces cerevisiae Is Encoded by JEN1. *J. Bacteriol.* **181**, 2620–2623 (1999).
263. Lodi, T., Fontanesi, F., Ferrero, I. & Donnini, C. Carboxylic acids permeases in yeast: two genes in Kluyveromyces lactis. *Gene* **339**, 111–119 (2004).
264. Jolkver, E. *et al.* Identification and Characterization of a Bacterial Transport System for the Uptake of Pyruvate, Propionate, and Acetate in Corynebacterium glutamicum. *J. Bacteriol.* **191**, 940 (2009).
265. Armenteros, J. J. A. *et al.* Detecting sequence signals in targeting peptides using deep learning. *Life Sci. Alliance* **2**, (2019).
266. Emanuelsson, O., Nielsen, H., Brunak, S. & von Heijne, G. Predicting Subcellular Localization of Proteins Based on their N-terminal Amino Acid Sequence. *J. Mol. Biol.* **300**, 1005–1016 (2000).
267. Horton, P. *et al.* WoLF PSORT: protein localization predictor. *Nucleic Acids Res.* **35**, W585-587 (2007).
268. Bannai, H., Tamada, Y., Maruyama, O., Nakai, K. & Miyano, S. Extensive feature detection of N-terminal protein sorting signals. *Bioinformatics* **18**, 298–305 (2002).
269. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. L. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).
270. Käll, L., Krogh, A. & Sonnhammer, E. L. L. A combined transmembrane topology and signal peptide prediction method. *J. Mol. Biol.* **338**, 1027–1036 (2004).
271. Reynolds, S. M., Käll, L., Riffle, M. E., Bilmes, J. A. & Noble, W. S. Transmembrane topology and signal peptide prediction using dynamic bayesian networks. *PLoS Comput. Biol.* **4**, e1000213 (2008).

272. Wang, Q. *et al.* PureseqTM: efficient and accurate prediction of transmembrane topology from amino acid sequence only. *bioRxiv* 627307 (2019) doi:10.1101/627307.
273. Denton, A. K. *et al.* Freeze-quenched maize mesophyll and bundle sheath separation uncovers bias in previous tissue-specific RNA-Seq data. *J. Exp. Bot.* **68**, 147–160 (2017).
274. Tausta, S. L. *et al.* Developmental dynamics of Kranz cell transcriptional specificity in maize leaf reveals early onset of C_4-related processes. *J. Exp. Bot.* **65**, 3543–3555 (2014).
275. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
276. Pérez-Pérez, J. M. *et al.* Functional Redundancy and Divergence within the Arabidopsis RETICULATA-RELATED Gene Family. *Plant Physiol.* **162**, 589–603 (2013).
277. Walters, R. G., Shephard, F., Rogers, J. J. M., Rolfe, S. A. & Horton, P. Identification of Mutants of Arabidopsis Defective in Acclimation of Photosynthesis to the Light Environment. *Plant Physiol.* **131**, 472–481 (2003).
278. Chazaux, M. *et al.* ACCLIMATION OF PHOTOSYNTHESIS TO THE ENVIRONMENT 1 regulates Photosystem II Supercomplex dynamics in response to light in Chlamydomonas reinhardtii. 2020.02.26.966580 Preprint at https://doi.org/10.1101/2020.02.26.966580 (2020).
279. Duy, D., Stübe, R., Wanner, G. & Philippar, K. The Chloroplast Permease PIC1 Regulates Plant Growth and Development by Directing Homeostasis and Transport of Iron. *Plant Physiol.* **155**, 1709–1722 (2011).
280. Kikuchi, S. *et al.* A 1-Megadalton Translocation Complex Containing Tic20 and Tic21 Mediates Chloroplast Protein Import at the Inner Envelope Membrane. *Plant Cell* **21**, 1781–1797 (2009).
281. Fujiki, Y. *et al.* Functional identification of AtAVT3, a family of vacuolar amino acid transporters, in Arabidopsis. *FEBS Lett.* **591**, 5–15 (2017).

282. Schwacke, R. *et al.* ARAMEMNON, a Novel Database for Arabidopsis Integral Membrane Proteins. *Plant Physiol.* **131**, 16–26 (2003).
283. Baba, T. *et al.* Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2**, 2006.0008 (2006).
284. Berardini, T. Z. *et al.* The arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome. *genesis* **53**, 474–485 (2015).
285. Emanuelsson, O., Nielsen, H. & von Heijne, G. ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci. Publ. Protein Soc.* **8**, 978–984 (1999).
286. Sharan, S. K., Thomason, L. C., Kuznetsov, S. G. & Court, D. L. Recombineering: A Homologous Recombination-Based Method of Genetic Engineering. *Nat. Protoc.* **4**, 206–223 (2009).
287. Hossain, M. J., Thurlow, C. M., Sun, D., Nasrin, S. & Liles, M. R. Genome modifications and cloning using a conjugally transferable recombineering system. *Biotechnol. Rep.* **8**, 24–35 (2015).
288. Technical protocol for FLPe expression plasmids for E. coli. (2013) http://dev.genebridges.com/gb/pdf/A104-107-FLPe_Plasmids_for_prokaryotic_expression.pdf. Accessed 14 April 2023.
289. Sprouffske, K. & Wagner, A. Growthcurver: an R package for obtaining interpretable metrics from microbial growth curves. *BMC Bioinformatics* **17**, 172 (2016).
290. Aesaert, S. *et al.* Optimized Transformation and Gene Editing of the B104 Public Maize Inbred by Improved Tissue Culture and Use of Morphogenic Regulators. *Front. Plant Sci.* **13**, 883847 (2022).
291. Coussens, G. *et al.* Brachypodium distachyon promoters as efficient building blocks for transgenic research in maize. *J. Exp. Bot.* **63**, 4263–4273 (2012).
292. Wodniok, S. *et al.* Origin of land plants: Do conjugating green algae hold the key? *BMC Evol. Biol.* **11**, 104 (2011).
293. von Caemmerer, S. & Furbank, R. T. Strategies for improving C_4 photosynthesis. *Curr. Opin. Plant Biol.* **31**, 125–134 (2016).

294. Tang, S. *et al.* Pyruvate transporter BnaBASS2 impacts seed oil accumulation in Brassica napus. *Plant Biotechnol. J.* **20**, 2406–2417 (2022).
295. Proost, S. *et al.* PLAZA 3.0: an access point for plant comparative genomics. *Nucleic Acids Res.* **43**, D974–D981 (2015).
296. Li, F.-W. *et al.* Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nat. Plants* **4**, 460–472 (2018).
297. Li, F.-W. *et al.* Anthoceros genomes illuminate the origin of land plants and the unique biology of hornworts. *Nat. Plants* **6**, 259–272 (2020).
298. Howe, K. L. *et al.* Ensembl Genomes 2020—enabling non-vertebrate genomic research. *Nucleic Acids Res.* **48**, D689–D695 (2020).
299. Grigoriev, I. V. *et al.* PhycoCosm, a comparative algal genomics resource. *Nucleic Acids Res.* **49**, D1004–D1011 (2021).
300. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
301. Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
302. Jombart, T., Balloux, F. & Dray, S. adephylo: new tools for investigating the phylogenetic signal in biological traits. *Bioinformatics* **26**, 1907–1909 (2010).
303. Jomain-Baum, M., Schramm, V. L. & Hanson, R. W. Mechanism of 3-mercaptopicolinic acid inhibition of hepatic phosphoenolpyruvate carboxykinase (GTP). *J. Biol. Chem.* **251**, 37–44 (1976).
304. Authority (EFSA), E. F. S. Calcium acetate, calcium pyruvate, calcium succinate, magnesium pyruvate magnesium succinate and potassium malate added for nutritional purposes to food supplements. *EFSA J.* **7**, 1088 (2009).
305. Glenn, J. M. *et al.* Acute citrulline malate supplementation improves upper- and lower-body submaximal weightlifting exercise performance in resistance-trained females. *Eur. J. Nutr.* **56**, 775–784 (2017).
306. Saper, R. B., Eisenberg, D. M. & Phillips, R. S. Common Dietary Supplements for Weight Loss. *Am. Fam. Physician* **70**, 1731–1738 (2004).

307. Akita, H., Nakashima, N. & Hoshino, T. Pyruvate production using engineered Escherichia coli. *AMB Express* **6**, 94 (2016).
308. Park, H.-S., Lee, J.-Y. & Kim, H.-S. Production of L-DOPA(3,4-dihydroxyphenyl-L-alanine) from benzene by using a hybrid pathway. *Biotechnol. Bioeng.* **58**, 339–343 (1998).
309. Rosche, B., Sandford, V., Breuer, M., Hauer, B. & Rogers, P. Biotransformation of benzaldehyde into (R)-phenylacetylcarbinol by filamentous fungi or their extracts. *Appl. Microbiol. Biotechnol.* **57**, 309–315 (2001).
310. Zhang, Y. *et al.* An efficient method for N-acetyl-D-neuraminic acid production using coupled bacterial cells with a safe temperature-induced system. *Appl. Microbiol. Biotechnol.* **86**, 481–489 (2010).
311. Liu, J. *et al.* Biological production of l-malate: recent advances and future prospects. *World J. Microbiol. Biotechnol.* **34**, 6 (2017).
312. McKinlay, J. B., Vieille, C. & Zeikus, J. G. Prospects for a bio-based succinate industry. *Appl. Microbiol. Biotechnol.* **76**, 727–740 (2007).
313. Werpy, T. & Petersen, G. Top Value Added Chemicals from Biomass: Volume I -- Results of Screening for Potential Candidates from Sugars and Synthesis Gas. https://www.osti.gov/biblio/15008859 (2004) doi:10.2172/15008859.
314. Hatch, M. D. & Slack, C. R. Photosynthesis by sugar-cane leaves. *Biochem. J.* **101**, 103–111 (1966).
315. United Nations, Department of Economic and Social Affairs, Population Division. *World population prospects Highlights, 2019 revision Highlights, 2019 revision*. (2019).
316. Long, S. P., Zhu, X.-G., Naidu, S. L. & Ort, D. R. Can improvement in photosynthesis increase crop yields? *Plant Cell Environ.* **29**, 315–330 (2006).
317. Mitchell, P. L. & Sheehy, J. E. Supercharging rice photosynthesis to increase yield. *New Phytol.* **171**, 688–693 (2006).
318. von Caemmerer, S., Edwards, G. E., Koteyeva, N. & Cousins, A. B. Single cell C_4 photosynthesis in aquatic and terrestrial plants: A gas exchange perspective. *Aquat. Bot.* **118**, 71–80 (2014).

319. Holaday, A. S. & Bowes, G. C_4 Acid Metabolism and Dark CO_2 Fixation in a Submersed Aquatic Macrophyte (Hydrilla verticillata) 1. *Plant Physiol.* **65**, 331–335 (1980).
320. Caemmerer, S. V. C_4 photosynthesis in a single C_3 cell is theoretically inefficient but may ameliorate internal CO_2 diffusion limitations of C_3 leaves. *Plant Cell Environ.* **26**, 1191–1197 (2003).
321. Miyao, M., Masumoto, C., Miyazawa, S.-I. & Fukayama, H. Lessons from engineering a single-cell C_4 photosynthetic pathway into rice. *J. Exp. Bot.* **62**, 3021–3029 (2011).
322. Suzuki, S. *et al.* Carbon metabolism in transgenic rice plants that express phosphoenolpyruvate carboxylase and/or phosphoenolpyruvate carboxykinase. *Plant Sci.* **170**, 1010–1019 (2006).
323. Taniguchi, Y. *et al.* Overproduction of C_4 photosynthetic enzymes in transgenic rice plants: an approach to introduce the C_4-like photosynthetic pathway into rice. *J. Exp. Bot.* **59**, 1799–1809 (2008).
324. Huber, S. C. & Edwards, G. E. Transport in C_4 mesophyll chloroplasts. Characterization of the pyruvate carrier. *Biochim. Biophys. Acta BBA - Bioenerg.* **462**, 583–602 (1977).
325. Gerrard Wheeler, M. C. *et al.* Arabidopsis thaliana NADP-malic enzyme isoforms: high degree of identity but clearly distinct properties. *Plant Mol. Biol.* **67**, 231–242 (2008).
326. Maier, A., Zell, M. B. & Maurino, V. G. Malate decarboxylases: evolution and roles of NAD(P)-ME isoforms in species performing C_4 and C_3 photosynthesis. *J. Exp. Bot.* **62**, 3061–3069 (2011).
327. Saigo, M., Alvarez, C. E., Andreo, C. S. & Drincovich, M. F. Plastidial NADP-malic enzymes from grasses: Unraveling the way to the C_4 specific isoforms. *Plant Physiol. Biochem.* **63**, 39–48 (2013).
328. Fahnenstich, H. *et al.* Alteration of Organic Acid Metabolism in Arabidopsis Overexpressing the Maize C_4 NADP-Malic Enzyme Causes Accelerated Senescence during Extended Darkness. *Plant Physiol.* **145**, 640–652 (2007).

329. Badia, M. B. *et al.* Enhanced cytosolic NADP-ME2 activity in A. thaliana affects plant development, stress tolerance and specific diurnal and nocturnal cellular processes. *Plant Sci.* **240**, 193–203 (2015).
330. Geerlof, A. M9 mineral medium. (2010). https://static.igem.org/mediawiki/2019/2/20/T--Tuebingen--M9_recipe.pdf. Accessed 14 April 2023.
331. Weber, E., Engler, C., Gruetzner, R., Werner, S. & Marillonnet, S. A Modular Cloning System for Standardized Assembly of Multigene Constructs. *PLoS ONE* **6**, e16765 (2011).
332. Geddes, B. A., Mendoza-Suárez, M. A. & Poole, P. S. A Bacterial Expression Vector Archive (BEVA) for Flexible Modular Assembly of Golden Gate-Compatible Vectors. *Front. Microbiol.* **9**, (2019).
333. Grant, K. Engineering rhizobacteria as synthetic biology chassis. PhD thesis. University of Oxford. Oxford University Research Archive. (2019).
334. Engler, C. *et al.* A Golden Gate Modular Cloning Toolbox for Plants. *ACS Synth. Biol.* **3**, 839–843 (2014).
335. Grützner, R. *et al.* High-efficiency genome editing in plants mediated by a Cas9 gene containing multiple introns. *Plant Commun.* **2**, 100135 (2021).