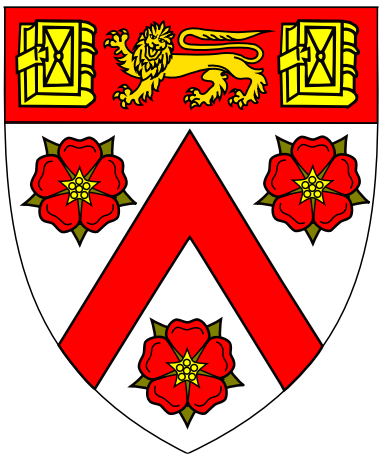
UNIVERSITY OF CAMBRIDGE

Systems analysis of Crassulacean acid metabolism (CAM) physiology and molecular biology



Methawi Chomthong

Supervisor: Prof. Howard Griffiths

Department of Plant Sciences

This thesis is submitted for the degree of

Doctor of Philosophy

Trinity college

January 2023

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as specified in the text. No part of this thesis has already been submitted for any degree or other qualification.

The thesis does not exceed the prescribed word limit of 60,000 words in length, exclusive of tables, footnotes, bibliography, and appendices as set by the Degree Committee for the Faculty of Biology.

Methawi Chomthong

January 2023

Summary

The main theme of this thesis was to explore the regulatory landscape of CAM using systems biology approaches. The scope of the regulatory landscape was drawn around the mesophyll metabolism of the dicotyledonous obligate CAM genus, *Kalanchoe,* even though the computational frameworks that had been developed here can also be applied to other species.

The first result chapter presented the Ordinary Differential Equation modelling of Crassulacean acid metabolism (Chapter 2) which was developed to capture the classical gas-exchange patterns as well as the responses to perturbation conditions. The findings suggested that the model was sufficient to explain the classical gas-exchange pattern whilst was also responsive to the perturbations. Nonetheless, the model parameters which represented the protein activities indirectly captured the upstream regulatory controls. Thus, the following result chapter shifted the focus to explore a more upstream level of regulations at the level of gene expression.

The second result chapter presented the Gene Regulatory Network Inference of *Kalanchoe fedtschenkoi* (Chapter 3). This chapter identified potential transcriptional regulators on different functional compartments of CAM including the following: Carboxylation subnetwork, Decarboxylation subnetwork, Circadian subnetwork and Stomatal subnetwork. This chapter highlighted the potential transcriptional regulators of key CAM genes, for example, PEPCarboxylase (PEPC), PEPCkinase (PPCK), pyruvate orthophosphate dikinase (PPDK), and pyruvate orthophosphate dikinase regulatory protein (PPDK-RP). Overall, the Gene Regulatory Network Inference provided the ranking of the potential transcriptional regulatory candidates on CAM genes. Hence, a reasonable step forward would be to probe for direct binding evidence through molecular approaches. The first step towards accessing the chromatin landscape with ATAC-sequencing technique was the Nuclei isolation followed by the flow cytometry separation technique for *Kalanchoe fedtschenkoi* which was presented as the final result chapter (Chapter 4).

To conclude, this thesis showed that the minimal mechanistic model at the level of protein functions can capture CAM gas-exchange patterns under various scenarios. Subsequently, a more upstream level of regulatory controls was explored across the genome with the Gene Regulatory Network Inference method. The key findings highlighted the potential transcriptional regulations of key CAM genes in addition to the regulations at the level of protein activities. Finally, the nuclei isolation was conducted as an initial step for a future molecular experiment to probe for chromatin accessibility for the CAM model species.

I dedicate this thesis to my supervisor,

whose knowledge and compassion are genuine and truly exceptional

Acknowledgements

First and foremost, I would like to thank Professor Howard Griffiths for giving me this opportunity to pursue my interest in computational and plant physiology from my undergraduate time until this very moment, and beyond. Your support and encouragement have been incredible throughout; you encourage me to challenge myself during my good time yet always make sure I can get through the difficult time. It has been a blessing to have a supervisor who listens to my ideas even though sometimes it has come out in *Methawi's language* as he calls it. Many fun and fruitful conversations about putting mathematical findings back into the context of plant physiology are at the top list of my good memories in Cambridge. You are the best supervisor I can ever ask for.

I would also like to thank Professor Nik Cunniffe for the discussion on modelling optimisation which pushed me forward when I thought that the ODE model may have hit the wall. Your support guided me to the finish line of this ODE project.

I am forever grateful to Dr Gitanjali Yadav for offering me a cupcake at a random tea break. That cupcake-led conversation kicked start my interest in Gene Regulatory Network which led to major discoveries during my PhD in the following year. I also need to thank you for challenging my view about the Co-expression network versus GRN inference. That has formed my basis on algorithmic development and I still enjoy thinking about it in my free time.

I would also like to extend my gratitude towards other members, old and new, in Howard-Hibberd-Kromdijk groups, especially Nick, Jamie, Yi, Moritz, and Jessica who were there since I entered the lab for the first time, as well as Leonie, Pallavi, Myriam, Indu, Rachael, and Tanya. All enthusiastic conversations and interactions have contributed towards my academic environment and the better understanding of plant physiology.

I would like to thank IPST institute for providing the initial sponsorship for my academic progress outside of Thailand. I acknowledge the departmental grants/funding and philosophical society grants for covering the rest of my PhD funding.

I am forever grateful to my Trinity family; Adam, Clare, Xin Ying, Nakarin, Neptune, Golf, Teoy, Tang, and Gift, as well as our neighbouring rival from St John, Gunn, for continuously supporting me throughout these years. We know it has not been easy but your presence has lifted a huge burden off my shoulders.

It would be impossible to not mention the City Road house (Kawin, Yokk, Peerapat, Ponpavi). You were there for me during my most confusing time. All the food and the laugh were memorable.

I would also like to thank the group of friends from my high school (Non, Ohm, Autt, Pang, Pung, and Fon). Having us from four different time zones in the same group chat means that we are always available for each other 24 hours. Maybe we should set up a call-centre.

Towards the final push of my thesis write-up phase, I sincerely thank Great, Bonita, Punpun, and Ice for checking that I do not forget to eat. I also sincerely thank Bhurint and Tossapol for listening to my whining and reminding me of the true inner peace.

Last but not least, I would like to thank my family, especially my dad, my mom, and my aunts for being forever supportive of me and my love of sciences and maths. Your love and understanding are always warm and never-failing. Thank you.

Table of contents

List of Figures

Chapter 2

Chapter 3

Chapter 4

List of Tables

Chapter 1

Chapter 2

List of Footnotes

Chapter 1

Introduction

Classical knowledge of Crassulacean Acid Metabolism

Crassulacean Acid Metabolism (CAM) species are known for their remarkable photosynthetic traits, including night-time primary CO_2-fixation by enzyme Phosphoenolpyruvate Carboxylase (PEPC), which conserves water when stomata are closed for part of the day as opposed to the daytime stomatal opening and atmospheric CO_2 fixation in C3 species. Although common CAM plants are typically associated with the succulent traits, not all succulent species perform CAM pathway and not all CAM plants are succulent (Kluge and Ting 1978b). The presence of the storage water within succulent tissues (i.e. leaf or stem) provides the ecological adaptive value for plants to occupy habitats with limited water supply in both semi-arid desert environments or the epiphytic niches in tropical rainforests. The combination of these beneficial traits led to remarkable global diversity arising from multiple independent origins of CAM over the last 5 million years and accounted for at least 6% of the modern Angiosperm species (Dodd *et al.* 2002; Borland *et al.* 2018; Wai and VanBuren 2018; Edwards 2019).

Although the earliest record of Crassulacean acid metabolism (CAM) the author is aware of was dated back to the Roman time where someone noticed the acid taste in house plants in the morning, the official origin of the scientific study of CAM was in the Late Enlightenment period (Black and Osmond 2003). The earliest academically documented CAM phenomenon was done by de Saussure[1] in 1804 where the ability of the *Opuntia* stem to remove CO_2 from the atmosphere at night was observed. In a separate observation of garden plants in India, the organic acid accumulation in *Bryophyllum calycinum* leaves was documented by Heyne[2] in 1815. Because the discovery of CO_2 fixation at night by de Saussure was very unexpected at that time, this observation did not fit into the developing concept of photosynthesis and was forgotten for more than a century. As a result, the malic acid accumulation documented by Heyne remained an unsolved puzzle despite the fact that mechanistic explanation of the acid accumulation had been documented 11 years earlier. These two aspects of CAM, the night time CO_2 fixation and malic acid accumulation, were not linked together as one phenomenon until Thomas and Wolf made this connection in 1949 (de Saussure, 1804; Heyne, 1815; Thomas, 1949; Wolf, 1949; Kluge & Ting, 1978b).

In terms of the taxonomical distribution, CAM pathway have been identified in more than 30 Angiosperm families, namely Agavaceae, Aizoaceae, Asclepiadaceae, Asteraceae, Bromeliaceae, Cactaceae, Crassulaceae, Cucurbitaceae, Didiereaceae, Euphorbiaceae, Geraniaceae, Labiatae, Liliaceae, Oxalidaceae, Orchidaceae, Piperaceae, Portulacaceae and Vitaceae. Although these families are not taxonomically close, they are all tropical or subtropical in origins and most of their CAM members are either stem or leaf succulents. Suffice it to say, many physiological, anatomical, biochemical and molecular features of CAM have developed convergently to regulate the expression of the CAM in a wide range of families.

Footnote:

1 *"…The cactus takes up CO_2 and oxygen in the same proportions when the former is present in small amounts. Experiments reported in A. on oxygen fixation have been performed in air previously deprived of CO_2, and the question arises whether this latter is taken up by the leaves in the normal air preferably to oxygen. I could show that a cactus placed in a jar with air enriched in CO_2 fixes CO_2 and oxygen at similar rates. Fixation was faster than in the absence of added CO_2. The experiment was as follows: 153.6 cm3 (7.75 cubic inches) of cactus were placed in the afternoon in a jar containing 43.5 cubic inches of air over mercury; the air contained 74% nitrogen, 19% oxygen and 7% CO_2 in volume. After 12h darkness, the volume of the air was reduced by the plant to 81% of the initial volume. Hence, air taken up by the leaves was about 1.25 times of their volume. I found by analysing the remaining air that the leaves absorbed 13.5 parts of oxygen and 5.5 parts of CO_2. The two gases were thereafter fixed in similar proportions, about ¾ of their initial amount. After their fixation the leaves were saturated and no further fixation could be observed in continuing darkness…"* (de Saussure, 1804, quoted from Kluge & Ting, 1978b)

2 *"…The leaves of the Cotyledon calcina, a plant called by Mr. Salisbury Bryophyllum calycinum, which on the whole have herbaceous taste, are in the morning as acid as sorrel, if not more so; as the day advances, the lose their acidity, and are tasteless about noon and become bitterish towards evening…"* (Heyne, 1815, quoted from Kluge & Ting, 1978c)

The distinct regulation of guard cells allows stomatal opening at night, directing CO_2 flux into the primary carboxylation process by PEPC during Phase I of CAM (Osmond 1978). This temporal shift in initial carboxylation by PEPC is mediated by the action of a circadian clock-controlled protein kinase called Phosphoenolpyruvate Carboxylase Kinase (PEPC kinase) which phosphorylates the PEPC and thus reduces its sensitivity to malate-inhibition at night (Hartwell *et al.* 1996; Hartwell *et al.* 1999; Nimmo 2000; Dodd *et al.* 2002). The final product of the primary carboxylation process catalysed by the active form of PEPC is malic acid which is then sequestered into large storage vacuoles that can occupy up to 95% of the mesophyll cell volume. During predominant part of the subsequent light period (Phase III), the malic acid is released from the vacuole and enters the decarboxylation pathway to release CO_2 for the secondary carboxylation by Ribulose-1,5-bisphosphate Carboxylase/Oxygenase (Rubisco). The CO_2 is concentrated around Rubisco behind closed stomata during the day-time, allowing CAM Rubisco to delay activation and to adopt a more relaxed specificity factor (selectivity of CO_2 over O_2 lower, compared to C3 plants) albeit with the retention of a high affinity for CO_2 (low Km,CO_2) (Griffiths *et al.* 2002, 2007, 2008). This temporal separation between primary and secondary CO_2 fixation within a single mesophyll cell enhances an overall operating carboxylation efficiency while water use is lower compared to C3 species (Borland *et al.* 2009, 2018).

A number of complex CAM phenotypes demonstrate remarkable physiological and ecological plasticity which still engender intriguing research questions (Yang *et al.* 2015; Winter *et al.* 2015; Heyduk *et al.* 2019; Winter 2019). In strong CAM species, such as *Kalanchoe daigremontiana, Agave tequilana* and *Tillandsia usneoides*, the strength and duration of the intermediate Phases II and IV (i.e. early morning and late afternoon, respectively) are adversely affected by degree of succulence and drought, but can be restored quickly upon rehydration (Owen and Griffiths 2013; Owen *et al.* 2016). Plasticity as represented by facultative CAM has also been documented for woody dicotylenous species such as *Clusia minor*, and temperate perennial herbs such as *Sedum telephium*, and many members of the Aizoaceae which can rapidly and reversibly undergo CAM induction (Borland and Griffiths, 1990; Haag-Kerwer *et al.* 1992; Borland *et al.* 1992; Dodd *et al.* 2002). Finally, there is the well-defined CAM induction triggered by abiotic stress in *Mesembryanthumum crystallinum* (Adams *et al.* 1998; Cushman 2001; Winter and Holtum 2007) or during leaf development and aging in *Kalanchoe* species (Hartwell et al., 2016). These examples of CAM inducibility and Phase II and IV responsiveness suggest that CAM regulatory mechanisms drive a core temporal switch which

can then be modified depending on environmental conditions and/or internal metabolic feedback. The latest supporting evidence comes from the extent that the circadian expression of the CAM cycle can be disrupted when key metabolic enzymes have been manipulated (Dever *et al.*, 2015; Hartwell *et al.*, 2016; Boxall *et al.*, 2017, 2020)

Modern -omics data and genetic manipulation tools of Crassulacean Acid Metabolism

Availability of CAM genomes have provided a springboard for analysis of orthologues and transcriptional control elements (Ming *et al.* 2015; Yang *et al.* 2017; Wai *et al.* 2019), whereas comparative multi-omics analysis have contributed to a more in-depth understanding of potential regulatory relationships in various CAM species (Zhang *et al.* 2016; Chiang *et al.* 2016; Abraham *et al.* 2016; Yin *et al.* 2018; Heyduk *et al.* 2019). Firstly, comparative transcriptomic studies have provided insights into the evolutionary trajectory of CAM and suggested that transcriptional regulation is associated with specific expression profile of key CAM genes (Ming *et al.* 2015; Yang *et al.* 2017; Wai *et al.* 2019). Secondly, additional analysis could extract lists of candidate CAM specific transcription factors which mediate CAM mode in obligate CAM species (*Kalanchoe fedtschenkoi* and *Agave americana*) (Moseley *et al.* 2018; Yin *et al.* 2018) and mediate the transition from C3 to CAM mode in facultative CAM species (*Mesembryanthemum crystallinum* and *Talinum triangulare*) (Brilhaus *et al.* 2016; Amin *et al.* 2019; Maleckova *et al.* 2019). Thirdly, additional layers of control are also being explored at the level of less conventional post-transcriptional regulation via microRNAs as well as long non-coding RNAs which could function competitively with endogenous RNAs to alter the expression profile of key CAM pathway genes such as PEPC and PPDK (Yang *et al.* 2015; Wai *et al.* 2017; Bai *et al.* 2019). Finally, transgenic manipulation experiments with RNA silencing approaches have been used to investigate the relationships between different genes in *Kalanchoe* species (Dever *et al.* 2015; Boxall *et al.* 2017; Boxall *et al.* 2020).

Comparative transcriptomic studies have provided a list of genes with specific CAM expression patterns with timing either phase-shifted or inverted between C3 and CAM. These genes with specific CAM expression patterns may play important roles in synchronising CAM mesophyll processes. Comparative transcriptomic studies have associated CAM evolution with specific regulation of transcript abundance pattern. Firstly, comparative studies within 13 Orchid species suggested that regulation at the transcription level of carboxylation and decarboxylation genes (PEPC, PPCK, and PPDK) was associated with CAM evolution within

this clade (Zhang *et al.* 2016). Secondly, transcriptomic analysis of two cultivated pineapple varieties and one wild pineapple relative showed that CAM-related genes demonstrated an inverse diel expression pattern in photosynthetic tissues (Ming *et al.* 2015). With the available pineapple genome data, they also suggested that CAM evolved not by gene duplication but through the modified expression of existing C3 genes. In addition, a key mechanism that controls the expression timing of CAM specific genes has been suggested to be the circadian clock, because the differentially regulated CAM pathway genes were enriched with clock-associated *cis*-regulatory elements (Ming *et al.* 2015).

The molecular basis to metabolic regulation associated with CAM

Carbohydrate and malate metabolism also show contrasting patterns between CAM and C3. Firstly, Kalanchoe genome analysis provided evidence for contrasting differential regulation of carbohydrate processing between CAM and C3 species. A network analysis compared families of carbohydrate 'active' gene families which were similar in number to those found in Arabidopsis but with altered expression profiles, with a possible role for a trehalose-6-phosphate synthase and an invertase, acting to alter carbohydrate partitioning between substrates for CAM and those needed for growth (Yang *et al.* 2017). Secondly, putative orthologues of chloroplast and vacuolar sugar transporters of pineapple have adopted a defined diel expression pattern, hence being implicated in the diel processing of carbohydrate in this soluble sugar processing CAM species (Borland *et al.* 2016). Thirdly, genomic studies of transcriptional patterns have also supported that malate channel ALMT could be one of the key candidate control point to allow for CAM evolution and function in both pineapple and *Talinum* (Brilhaus *et al.* 2016; Wai and VanBuren 2018). Finally, it is well-established that CAM induction in *Mesembryanthemum crystallinum* is associated with an increased in transcript abundance of a chloroplast carbohydrate transporter gene associated with a higher Glucose-6-phosphate transport rate in its CAM mode (Cushman *et al.* 2008).

The contribution of non-canonical post-transcriptional control must not be neglected. Emerging evidence from microRNA (miRNA) and long non-coding RNA (lncRNA) studies in pineapple have also offered additional control mechanism on several CAM-related genes at the post-transcriptional level (Wai *et al.* 2017; Bai *et al.* 2019). The evidence in pineapple (*Ananas comosus*) has suggested that long non-coding RNAs may compete with PPCK and PEPC genes for being transcribed in the photosynthetic tissue during the daytime, hence preventing these

nocturnal CAM enzymes from being transcribed during the day (Bai *et al.* 2019). A separate study also identified 20% of pineapple microRNAs with diel expression pattern and the specific binding sequence to CAM genes including PPCK1 and MDH (Wai *et al.* 2017).

Recent transgenic manipulation experiments suggest a close link between circadian and metabolic control in CAM. Downregulating the carboxylation activity (PEPC) or decarboxylation activity (mitochondrial NAD-Malic enzyme (NAD-ME)) have been shown to disrupt CAM rhythmicity in other mesophyll components distant to the knockdown point, in addition to altering the oscillation pattern of circadian clock genes through unidentified metabolic feedbacks (Dever *et al.* 2015; Boxall *et al.* 2020). At the carboxylation pathway, the regulation of PEPC has been recently investigated using two recent RNA silencing approaches. In the first approach, the PEPC kinase of *Kalanchoe fedtschenkoi* was silenced, resulting in reduced dark phosphorylation of PEPC and a 66% reduction in night-time CO_2 assimilation (Boxall *et al.* 2017). This study emphasised the importance of PEPC phosphorylation to maintain PEPC activity through the dark period, which in turn affected circadian clock stability and overall CAM productivity. Subsequently, this group developed transgenic lines of *K. laxiflora* in which PEPC was down-regulated using RNAi (Boxall *et al.* 2020). The plants suffered a complete loss of dark CO_2 assimilation, fixing CO_2 directly only in the light, and the stomata revert back to C3 daytime opening. Additionally, transcripts for PPCK were down-regulated at night and peaked at the start of the light period, and gas exchange suffered arrhythmia under constant light, despite the transcript abundances of some core circadian clock genes being enhanced (Boxall *et al.* 2020).

Considering the decarboxylation component of the CAM cycle, RNA silencing has been used to reduce the activities of two CAM enzymes, NAD-ME and PPDK, which are both involved in the decarboxylation of malate and release of CO_2 internally during the light period (Dever *et al.* 2015). The expression of CAM was suppressed in both sets of plants, which tended to take up the majority of CO_2 directly in the light. There was a reduction in transcripts of many genes associated with the CAM cycle, including PEPC and genes associated with storage carbohydrate trafficking, including PPDK (in NAD-ME mutant) and glucan water dikinase and glucose-6-phosphate translocator. Whilst the expression patterns of PPCK and core clock gene transcripts were not affected, the circadian cycle became arrhythmic under continuous light, again suggesting a role for metabolite feedback in resetting the CAM cycle on a daily basis (Dever *et al.* 2015).

Although the transgenic perturbation approaches did not target the carbon storage element of CAM directly, the manipulation of PPCK and decarboxylation pathway both resulted in altered cellular carbohydrate balances (Dever *et al.* 2015; Boxall *et al*. 2017). Starch accumulation, which normally precedes the dark period to provide carbon skeletons for carboxylation, was generally reduced. In addition, there was a distinct correlation between a peak of sucrose and the expression of a circadian response-regulator transcripts. This provided more evidence for the connection between carbohydrate supply with circadian clock expression (Borland and Griffiths, 1997; Boxall *et al.* 2017). Moreover, the reduction in PEPC expression may have suggested additional insights for the transition between C3 and CAM modes of assimilation (Boxall *et al.* 2020). Transcript for genes associated with the partitioning of starch in C3 systems, such as the amylolytic pathway at night, were upregulated in the *K. laxiflora* transgenic lines developed by Boxall et al. 2020, as compared to the phosphorolytic route normally associated with starch-storing CAM species (Borland *et al*. 2009, 2016).

Remaining questions of Crassulacean Acid Metabolism

Remaining questions of CAM can be grouped into 2 categories; A) How are enzymes and their related pathways become active/inactive at the right time of the day?, and B) how have they reached that stage from the starting material in ancestral species?

In terms of the temporal control of the CAM pathway, the most complete explanation lies at the Phosphoenolpyruvate Carboxylase (PEPC) which is phosphorylated by a nocturnally expressed enzyme, PEPCkinase (PPCK) (Hartwell *et al.* 1999). The phosphorylated form of PEPC has lower sensitivity to malate inhibition, thus it actively catalyses the initial carboxylation at night. The regulation of decarboxylation pathway is, however, less clear in CAM species. In C3 and C4 species, the enzyme pyruvate orthophosphate dikinase (PPDK) is active in the dephosphorylated-form whilst its phosphorylation and dephosphorylation reactions are catalysed by the same bi-functional enzyme, PPDK-Regulatory protein (PPDK-RP). This regulatory relationship between the PPDK-RP and PPDK has not been verified in CAM species. The most direct evidence is from timeseries immunoblotting in *K. fedtschenkoi* which detected phosphorylated-form of PPDK from 4 hours before dusk into 8 hours after dusk (Dever *et al.* 2015). The presence of the inactive phosphorylated-form of PPDK over the first two-thirds of the dark period is consistent with with the lack of decarboxylation activity during the night. This suggested that the phosphorylation/dephosphorylation reaction may be partly

responsible for the diurnal activation of PPDK in *K. fedtschenkoi*. Nonetheless, the conditions that facilitate the alternating function of PPDK-RP between kinase and phosphorylase are not known for CAM. In addition, there is another level of complexity due to the compartmentalisation of PPDK, whereby two-thirds of this protein is localised in the cytosol, whereas the remaining one-third is in chloroplasts (Kondo *et al.* 2000). It is evident that activities of different compartments of CAM must be highly orchestrated (Borland *et al.* 2016; Ceusters *et al.* 2021). However, there is a big gap in the understanding of how genes across different compartments or pathways are synchronised. Are different genes under the same or unique regulators? Are these regulators responding directly to physical parameters including light intensity, wavelength distribution, intercellular CO_2 concentration (Ci), and humidity? How does the circadian network wire into the CAM system? How different is CAM circadian network from the C3 circadian network?

It remains challenging to date to dissect these questions not only because CAM traits are complex, but also because CAM operates through the interlocking loops. Stomatal conductance and the strength of carboxylation activity influences the CO_2 diffusion rate. Subsequently, the CO_2 that has entered the leaf influences the degree of stomatal opening and carbohydrate production through the intercellular CO_2 (Ci)-sensing pathway and (de)carboxylation-Calvin cycle, respectively. Finally, carbohydrate is recycled to produce the 3-carbon compound, Phosphoenolpyruvate, which is the reagent for the primary carboxylation of the atmospheric CO_2, hence influencing the next round of CO_2 fixation and so on (Borland *et al.* 2016; Borland *et al.* 2018). Disturbing one specific point can affect the whole system, but some components might have ability to respond directly to the perturbation signal or adjust upstream components. For example, the series of genetic perturbation experiments have demonstrated that genes in unrelated pathways were responding, but this may have been signalled by the level of metabolites or the circadian clock arrhythmia (Dever *et al.* 2015; Boxall *et al.* 2017, 2020). One of the more specific examples is from the genetic perturbation of gene coding for NAD-ME. In the NAD-ME knockdown line, the phosphorylated-form of PPDK was detected throughout a 24-hour period, implying the constitutive down-regulation of PPDK activity (Dever *et al.* 2015). In other words, the lack of NAD-ME, which has a normal function to catalyse the first step in the decarboxylation pathway, resulted in the downregulation of the enzyme PPDK, which catalyses the subsequent step within the same decarboxylation pathway.

Stomatal behaviour is part of the complex CAM traits The puzzle regarding CAM stomatal function persists as follows. Firstly, the evidence for CAM stomatal response to light is still

inconclusive. Initial reports on facultative CAM species (*Mesembryanthemum crystallinum* and *Portulacaria afra*) suggested that stomata no longer respond to light signals once induced into CAM mode (Lee and Assmann 1992; Tallman *et al.* 1997). However, the stomata of obligate CAM species (*Kalanchoe pinnata* and *K. daigremontiana*) have been shown to open in response to blue-light, similar to C3 stomata (Gotoh *et al.* 2019). These contradictory observations provide a focus for the question originally posed by Osmond (1978), and the interplay between light cues with the CAM pathway. How can the canonical pathway of stomatal opening in responding to blue-light be overridden during the Phase III of CAM? Secondly, nocturnal opening behaviour of CAM is hypothesised to be mediated by the low intercellular CO_2 (Ci) as a result of PEPC activity, but any precise mechanism has not been elucidated to date (Borland and Griffiths 1997; Drennan and Nobel 2000; Ceusters *et al.* 2008; von Caemmerer and Griffiths 2009). Thirdly, it is not clear whether guard cell Phosphoenolpyruvate Carboxylase (PEPC) activity at night contributes to the stomatal behaviour in CAM species by providing malate as one of the major solutes driving the opening process (Santelia and Lawson 2016; Males and Griffiths 2017). Finally, the analysis of transcriptomic data from *Agave* showed that ABA biosynthesis and signaling components are enriched in the expression cluster that shows peak expression at 0 and 3 hours before dawn (Abraham *et al.* 2016). This suggested a potential role for ABA in being synthesised and perceived at the precise time that CAM plants enter Phase III stomatal closure. However, the direct effect of ABA on the inverted timing of CAM stomatal behaviour remains elusive (Holtum and Winter 1982; Chu *et al.* 1990; Bastide *et al.* 1993; Dai *et al.* 1994; Taybi *et al.* 1995; Taybi and Cushman 1999).

In terms of the evolution from C3 ancestral species, it is hypothesised that key CAM genes are already present in C3 system, but the regulatory components are rewired to result in the specific temporal characteristic of CAM pathway (Ming *et al*. 2015). The analysis of the pineapple (*Ananas comosus*) genome showed that CAM has evolved from re-wiring of existing C3 components through regulatory neofunctionalisation of pre-existing genes, rather than by coding neofunctionalisation (Ming *et al.* 2015). Whole-genome duplication and single-gene duplication are common features in plant genome evolution. After duplication events, some duplicated genes may be lost through fractionation. Alternatively, the duplicated genes that are retained within the genome can either partition the original gene function (i.e. subfunctionalisation) or develop novel functions (neofunctionalisation). The two types of neofunctionalization are regulatory neofunctionalisation and coding functionalisation.

Regulatory neofunctionalisation is a result of expression divergence which allows the protein encoded from the duplicated gene to function in a different temporal or spatial environment as compared to the protein encoded from the original gene. On the contrary, the coding neofunctionalisation results in a novel protein function due to the gain-of-function mutation in the coding region of the duplicated gene (Hughes *et al.* 2014).

Key CAM genes are present in C3 species, but do not function in the primary CO_2 fixation pathway like in CAM. For instance, whilst β-CA is crucial for catalysing the formation of HCO_3- for primary CO_2 fixation in CAM, the roles of β-CA in C3 range across photosynthetic and non-photosynthetic tissues (Aubry *et al.* 2011). The β-CA systems have been implicated in the CO_2 -sensing pathway in *Arabidopsis* (Hu *et al.* 2010), in supplying CO_2 to carboxylation site in tobacco (Price *et al.* 1994), and in lipid biosynthesis in cotton seed (Hoang *et al.* 1999; Hoang and Chapman 2002) and legume nodules (Flemetakis *et al.* 2003; Kavroulakis *et al.* 2007). In C3 species, the enzyme PEPC functions in providing carbon skeletons for Krebs cycle and ammonium assimilation pathway (Miyao and Fukayama 2003; Masumoto *et al.* 2010), maintaining malate homeostasis, and regulating stomatal conductance (Aubry *et al.* 2011). Notably, the transcript abundance of PEPC genes increase in response to drought and salinity stress in *Arabidopsis* and wheat (González *et al.* 2003; Sánchez *et al.* 2006). In C3, the PEPC is not the primary CO_2 fixation enzyme although there have been reports of PEPC catalysing the fixation of respired CO_2 in rice (Imaizumi *et al.* 1997). Similarly, a PEPCkinase knockout experiment in *Arabidopsis* implied that the enzyme functions in regulating the Krebs cycle and photorespiration, as opposed to the main regulator of the primary carboxylation pathway, as in CAM. In these examples, it could be that β-CA activity is prompted by the supply and demand for HCO_3^- when PEPC is activated.

To increase the probability of the rewiring of the gene network, duplication events followed by changes in the protein sequence and regulatory motifs would provide evidence for the formation of novel regulatory relationships, which then dictate the distinct diel expression profiles of CAM system. These genome rearrangements have been identified in multiple CAM species (Yang *et al.* 2017; Heyduk *et al.* 2019a). The presence of multiple paralogues of key CAM genes may have facilitated the neofunctionalisation of the newly duplicated genes. For instance, there are 5 paralogous PEPC genes in *Kalanchoe fedtschenkoi* genome with a single copy being highly expressed specifically at night (Yang *et al.* 2017). However, the evidence of the changes of the regulatory motif in the promoter region of this duplicated PEPC gene is still lacking. In a different CAM model species, pineapple, there has been a search for regulatory

motifs in promoter regions of key CAM genes which include β-CA, PEPC, PEPCkinase, MDH, PEPCK and PPDK (Ming *et al.* 2015). This study in pineapple searched for 5 known circadian clock-related motif sequences: morning element, evening element, CCA1-binding site, G-box element and TCP15-binding motif. They identified the presence of at least 1 of these 5 motifs for all of these genes of interest, with the exception of PPDK. However, direct evidence for specific Transcription Factors (TF) and the conditions under which TF binding occurs, have yet to be verified.

In this thesis, the author set out to address the first category of the open questions for CAM: how are enzymes and their related pathways regulated to become active/inactive at the right time of the day-night cycle? Our methodology relies on systems biology approaches, including computational modelling and the handling of large scale transcriptomic datasets. We address how newly available large-scale data can be aligned with various computational techniques including systems models developed for CAM to make more rapid progress in addressing the fundamental physiological questions. By identifying key molecular targets associated with regulatory hubs controlling key elements of the CAM system, such an approach could be used to inform the experimental genetic manipulation approaches needed to define these responses empirically. In order to develop these approaches, the focus of this thesis is firstly on the mesophyll control of biochemistry which is highly coordinated across the CAM Phases. The effect of physical or genetic perturbations on the pattern of CAM Phases through mesophyll biochemistry were explored with Ordinary Differential Equations. Subsequently, the author expanded the scope of the modelling into the Gene Regulatory Network level to investigate the potential roles of candidate transcriptional regulators of key CAM genes.

Large scale approaches (Section 1): An overview

A recent review covered the advancement in CAM modelling and the newly available '-omics' data with the venue for expansion/development (Chomthong and Griffiths 2020). A bottom-up approach builds on existing mechanistic models (Owen and Griffiths, 2013; Hartzell *et al.* 2018) and metabolic network models (Cheung *et al.* 2014; Shameer *et al.* 2018). Such models posit a null state, in which the only a minimal circadian regulation is imposed on selective model components. At this stage, the experiments reporting manipulated gene expression and associated physiological outputs are fruitful resources for developing an improved versions of the existing CAM mesophyll metabolic models (Boxall *et al.* 2017; Boxall *et al.* 2019; Dever *et al.* 2015). The advantage of this bottom-up approach is that it does not require prior knowledge about an entire regulatory network. In contrast, the top-down approaches such as the construction of gene regulatory network and protein interaction network require genome-wide input data (Emmert-Streib *et al.* 2014). The challenge of the top-down approach is to connect the complete information back to the classical mechanistic understanding of the CAM system and distinguish between causation and correlation.

Ultimately the top-down and bottom-up approaches converge in allowing identification of a specific gene knockdown/knockout targets which could be introduced into tractable transformation systems, such as *Kalanchoe* (Hartwell *et al.* 2016; Liu *et al.* 2019). The refinement of pre-existing mesophyll metabolic models would allow for an *in silico* quantification of the effect that a combination of candidate control points have on CAM physiological output. In addition, multiple versions of model developed specifically for different CAM condition may capture the transition from C3 into CAM behaviour which could suggest multiple possible routes for evolutionary, developmental or stress-induced transition from C3 into CAM. The responses to the genetic perturbation can be used to validate the models and potentially be informative for a future CAM biodesign attempt (Lim *et al.* 2019).

Large scale approaches (Section 2) : Modelling approaches with a Systems Dynamic Model and Ordinary Differential Equations

CAM mesophyll metabolism integrates complex traits coupled via feedback effects, and control of gene expression. The diel pattern of CAM mesophyll metabolism relies on circadian controls which set the diel phases of carboxylation processes while metabolites provide another layer of refinement (Borland *et al.* 2016). Circadian regulation of the activation of key enzymes, including carboxylases PEPC and Rubisco, has been shown to play an important role in governing the temporal separation of CAM photosynthetic activity (Carter *et al.* 1991; Hartwell *et al.* 1996; Hartwell *et al.* 1999; Griffiths *et al.* 2002; Maxwell *et al.* 2002; Dodd *et al.* 2002; Davies and Griffiths 2012). The most prominent integration point between circadian control and metabolite control is at PEPC which is phosphorylated by PPCK at night leading to the reduced sensitivity to malate allosteric inhibition, whilst the PPCK abundance itself is under transcriptional control from circadian clock outputs (Carter *et al*. 1991; Hartwell *et al.* 1999). In addition, PPCK gene expression can be modified, with increased malate accumulation exerting negative feedback on PPCK gene expression, possibly as a secondary effect via the circadian clock (Borland *et al.* 1999; Nimmo 2000). Thus, modelling techniques that allow for the interaction of various components in the system have been proved useful for CAM.

System Dynamics modelling was a common tool in modelling continuous nonlinear systems with interlocking feedback interactions. The approach was firstly developed in 1961 for complex problems in economics and world population (Olson 2003). The System Dynamics modelling was developed for CAM in 2013 by Owen and Griffiths (Owen and Griffiths 2013).The model for CAM imposed regulation at flow junctions which include stomatal conductance, mesophyll conductance, and malic acid transport across vacuole while these junctions were responding to feedback control from stomatal aperture, malate-inhibition of PEPC and enzyme kinetics (Owen and Griffiths 2013).

Ordinary Differential Equations (ODEs) are equations in which the derivatives are taken with respect to one independent variable. The term ‘Ordinary’ distinguishes the approach from Partial Differential Equations (PDEs), where the derivatives are taken with respect to multiple variables. The ODE system is a more mathematically rigorous form of writing the system dynamics model while keeping feedback interactions embedded in the system of equations. In

the case of CAM modelling, the single variable where all other state variables were taken derivative with respect to would be the 'time' variable.

The System Dynamics model and Ordinary Differential Equations have been developed to be parameterised by experimental data. Thus, the emergence of genetic and physiological manipulation experiments can be integrated into these modelling tools to investigate the effect of perturbation on the system responses through comparing measured and predicted changes in expression of the canonical CAM pathway.

Large scale approaches (Section 3): The nature of the timeseries transcriptome data and its potential for further use

The study that acted as a springboard for subsequent detailed analysis of CAM in *Kalanchoe* spp. was the genome sequencing and transcriptome sequencing project done by Yang et al in 2017 (Yang *et al.* 2017). This study set-off by sequencing the 256 Mb genome (diploid 2n = 34 chromosomes) of *Kalanchoe fedtschenkoi* for the first time. After the genome assembly, the 30,964 protein-coding genes were predicted and annotated. This was the first eudicot CAM lineage to be sequenced at that time. After the gene annotation has been done, this study by Yang identified two distinct whole genome duplication events in the *K. fedtschenkoi* based on the analysis of the synthenic patterns. The authors identified multiple key CAM genes to have undergone a recent genome duplication events, including PEPC, MDH, ALMT, NAD-ME, and NADP-ME families. This is in agreement with the view that gene duplication and functional diversification through temporal gene expression reprogramming could be a major source of the evolution of CAM in the *Kalanchoe fedtschenkoi*. However, this is different from the scenario in pineapple genome whereby the key CAM genes did not go through duplication. Instead, the pineapple CAM genes are the same ancestral copies that are present in non-CAM grass species but are more highly expressed in CAM tissue compared to non-photosynthetic tissues (Yang *et al.* 2017; Wai and VanBuren, 2018).

The transcriptome obtained through this same project by Yang et al were used for gene co-expression analysis, cluster analysis, and inferring the convergent evolution of key genes. The gene co-expression modules were firstly constructed from the timeseries 24-hr transcriptome data with the sampling frequency of every 2 hours. The total of 25 gene co-expression modules were obtained from the weighted correlated network. Among these 25 modules, there was one

module of 782 genes associated with the nighttime sample collection and one module of 1911 genes associated with the daytime sample collection. Notably, the genes encoding for β-Carbonic anhydrase (gene ID of Kaladp0018s0289), PEPC2 (gene ID of Kaladp0048s0578), PPCK (gene ID of Kaladp0037s0517), MDH (gene ID of Kaladp0022s0111) and ALMT6 (gene ID of Kaladp0062s0038) were clustered into the nighttime module whereas the PPDK-RP (gene ID of Kaladp0010s0106) was clustered into the daytime module (Yang *et al.* 2017).

In addition to the gene co-expression module analysis, Yang et al also performed cluster analysis. The key difference between the co-expression module and cluster analysis is that the co-expression module was based on the Weighted correlation network analysis (WGCNA) whereas the cluster analysis was done by performing polynomial regression on each gene expression function prior to choosing the number of clusters by minimizing the within group sum of square (Yang *et al.* 2017). The clustering analysis returns 11 clusters with a zinc-finger protein *CONSTANS-like* gene as a central hub for a cluster which contain *PEPC1* (Kaladp0095s0055) and *PPCK2* (Kaladp0604s0001), and evening-element binding *REVEILLE* transcription factors as hubs in cluster of *NADP-ME* genes (Kaladp0092s0166) (Yang *et al.*, 2017).

The study by Yang *et al* also attempted to identify the convergent evolution of CAM genes through the convergent diel expression profiles and the convergent in protein sequence. Initially, the convergent diel expression profiles were created by comparing the timeseries gene expression profiles of *Kalanchoe fedtschenkoi* (eudicot CAM), *Ananas comosus* (pineapple, monocot CAM) and *Arabidopsis thaliana* (eudicot C3). The authors use three conditions to define the convergent diel expression profiles between these two CAM species. The first condition was the expression profiles of *K. fedtschenkoi* being positively correlated with pineapple ($r > 0.8$) but not highly correlated with *A. thaliana* ($r < 0.5$). The second condition was the expression profile within the *K. fedtschenkoi* itself must have significant difference in expression level between midday-vs-midnight or dawn-vs-dusk. The final condition was the shift in 'timecourse' between *K. fedtschenkoi* and pineapple were less than 3 hours but the shift between *K. fedtschenkoi* and *A. thaliana* were more than 6 hours. This approach identified 54 genes with convergent diel expression profiles between *K. fedtschenkoi* and pineapple as opposed to the profiles of C3. These 54 genes include important genes with well-known functions such as phosphoenolpyruvate carboxylase kinase 1 (PPCK1), phototropin 2 (PHOT2) and Heat shock protein 70 (HSP70). In a parallel analysis, the convergence of protein sequence was studied through the construction of phylogenetic trees based on protein sequence of 13

plant species (8 C3 species, 2 C4 species, and 3 CAM species)[3]. Convergent changes in protein sequence were identified between *K. fedtschenkoi* and *P. equestris* for PEPC2 (gene ID of Kaladp0048s0578) and ELONGATED HYPOCOTYL5 (HY5, gene ID of Kaladp0060s0460) genes. The HY5 is a bZIP family transcription factor which was previously reported to entrain the circadian clock in C3 species through the blue light signaling pathway, but the equivalent role in CAM is yet to be confirmed (Hsu and Harmer 2014).

The evidence of convergence did not span homogeneously across all pathways. Although there are evidence of convergence in the carboxylation pathway, the analysis of the decarboxylation pathway did not detect any convergent in the diel gene expression pattern nor the convergence in protein sequence for decarboxylation genes of *K. fedtschenkoi* (Yang *et al.* 2017). Based on this line of evidence alone, the authors suggested that CAM evolution may have not involved the major changes in the decarboxylation genes. On a related note, there was no evidence of the convergence in the diel gene expression patterns of circadian genes between *K. fedtschenkoi* and pineapple despite HY5 being a circadian gene which showed the convergence in protein sequence between CAM species (Yang *et al.* 2017). One possible explanation is that there could be unknown circadian genes which are shared between *K. fedtschenkoi* and pineapple but whose orthologues are not part of the circadian network in the *A. thaliana*. Alternatively, it opens up a rather interesting question on the comparative evolution of eudicot CAM versus monocot CAM. What is the possibility of similar genes from the same functional pathways, be that the decarboxylation pathway or others, are expressed at the different time of the day and are controlled by different regulators but can still result in the fully functional CAM systems? What would be the most favourable evolutionary trajectories throughout the 60 independent origins of CAM?

Footnote

3 -The thirteen plant species include eight C3 species, two C4 species, and three CAM species

C3 = {*Amborella trichopoda, Arabidopsis thaliana, Brachypodium distachyon, Mimulus guttatus, Musa acuminata, Oryza sativa, Solanum lycopersicum, Vitis vinifera*}

C4 = {*Setaria italica, Sorghum bicolor*}

CAM = {*Ananas comosus, Kalanchoë fedtschenkoi, Phalaenopsis equestris*}

Large scale approaches (Section 4): Gene regulatory network inference

Gene Regulatory Networks (GRNs) encompass highly complex interactions of transcriptional regulators and their respective target genes. The availability of the microarray and/or transcriptome data has opened up the way for the construction of Gene Regulatory Networks (GRNs) by the *inference* methods in various biological systems. The term *inference* indicates the nature of the resultant GRNs which are based on the interactions of genes calculated through the model structures rather than the direct evidence of transcriptional regulator binding to the promoter regions of the genes of interest. The GRN inference has been widely used in animal system around year 2000 and subsequently being adopted by the plant community (Emmert-Streib *et al.* 2014). The application in plants have brought insight into the questions of abiotic-responses such as nitrogen-signaling in *Arabiodopsis* (Varala *et al.* 2018). Nonetheless, the application of GRN inference approaches had not been extended into CAM research. The availability of the timeseries transcriptome data from *K. fedtschenkoi* in 2017 has brought our attention to the application of GRN inference to address the GRN of CAM species. The co-expression network analysis suggested the number of potential transcriptional regulators of CAM to be as many as 1509 candidate genes from *Agave americana* (Yin *et al.* 2018). With this large number of potential regulators in CAM system, this is beyond the capacity of a small number of equations. Hence, the GRN inference approach has become a useful tool for addressing CAM questions.

The GENIE3 algorithm was the state-of-the-art gene regulatory network inference algorithm with the best performance in the DREAM4 challenge (Huynh-Thu *et al.* 2010; Mercatelli *et al.* 2020). The improved version of this algorithm, dynamical GENIE3 (dynGENIE3), was developed specifically to handle timeseries datasets (Huynh-Thu and Geurts 2018). The dynGENIE3 algorithm was validated extensively through the artificial data on DREAM4 challenge platform as well as biological benchmarks from three systems (*Saccharomyces cerevisiae*, *Drosophila melanogaster* and *Escherichia coli*) (Huynh-Thu and Geurts 2018). The dynGENIE3 algorithm uses the Ordinary Differential Equation (ODE) to model the expression level of any gene *j* to be dependent on the rate of gene expression minus mRNA degradation. Then, the transcription rate was assumed to be a potentially non-linear function f_j of the expression levels of any genes *p* (possibly including the gene *j* itself). In both GENIE3 and dynGENIE3an ensemble of regression trees was constructed to return the function f_j , which

are the outputs that contain the information of gene regulatory relationships. The regression trees split the input samples (i.e. timeseries gene expression data) with binary tests based on one input variable, trying to reduce as much as possible the variance of the output variable in the resulting subsets of samples. Subsequently, the Variable Important Measure (VIM) is computed from the reduction in variance at each branch of the trees and ranked in descending order to provide the Regulatory Link Ranking for each individual target gene. This Regulatory Link Ranking provides what would be of interest to experimental biologists, i.e. the ranking scores which hint the priority of the most likely candidate when a choice must be made to select which regulatory relationships would be knockout/knockdown before others. The large scale computational step was done in a non-bias manner and contain the score of all the possible combinations of regulatory relationship within the genome, that is (the total number of genes) x (the total number of genes - 1). With the genome size of *K. fedtschenkoi*, the total number of possible combinations of regulatory relationship within the genome would be 958,738,332 interactions which are impossible to be calculated by hands

Large scale approaches (Section 5): Nuclei isolation and ATAC-sequencing

Despite predictive power of the Gene Regulatory Network (GRN) approach, it relies heavily on transcriptome data. The developers of the dynGENIE3 algorithm acknowledged that complexity of biological network could limit the predictive power of computational model (Huynh-Thu and Geurts 2018). To complement the GRN predictions from computational methods, experimental data that capture the chromatin accessibility state will be required. This type of data can be obtained from early approaches such as DNaseI-seq, MNase-seq, FAIRE-seq , and ChIP-seq. Alternatively, the Assay for Transposase-Accessible Chromatin with high-throughput Sequencing (ATAC-seq) is the state of the art method with the fastest protocol yet requires the smallest number of pure nuclei compared to other protocols previously mentioned (Tsompana and Michael J Buck 2014). The ATAC-seq relies on the highly active transposase enzyme (Tn5) to insert the sequencing primers preferentially into the physically accessible regions of chromatin strands. Following with the next-generation sequencing, the sequencing reads can be aligned back to the genome to indicate the open regions across the genome.

The first step in the ATAC-seq protocol is to purify nuclei from the tissue samples. There are two methods for nuclei isolation: INTACT (Isolation of Nuclei TAgged in specific Cell Types) and sucrose-gradient sedimentation. The INTACT method requires the generation of transgenic plants containing biotinylated nuclei which can be extracted through affinity purification using the streptavidin-coated magnetic beads. The sucrose-gradient sedimentation method is a non-transgenic approach which relies on the gravity-based separation of nuclei from organelles and cellular debris according to the different density along the vertical axis of the sucrose-gradient centrifugation. In theory, the nuclei obtained from the INTACT method would be less contaminated than the nuclei obtained from the sucrose-gradient sedimentation method. However, the success of the INTACT depends on the feasibility of the transgenic protocol. For species with limitations in transgenic approach, the sucrose-gradient sedimentation method is sufficient for the ATAC-seq protocol which requires 50,000 nuclei for the subsequent transposase reaction. Due to the chloroplast and mitochondrial contamination, it is estimated that the number of sequencing reads per library mapped to nuclei is 50% and > 90% for sucrose-gradient sedimentation and INTACT method, respectively. To increase the number of sequencing reads being mapped to the nuclei, the number of reads per library can be increased at the sequencing step.

For CAM species, there is currently no protocol for generating the biotinylated-nuclei transgenic lines. The available option is to modify the transgenic protocol developed for the RNA inference method (Dever *et al.* 2015). However, this thesis explores the sucrose-gradient sedimentation method in combination with the spectral flow-cytometer in order to purify CAM nuclei for subsequent sequencing protocol (Lu *et al.* 2017; Bajic *et al.* 2018). Once nuclei isolation for CAM species is successful, the next step would be to aim for ATAC-seq. The sequencing reads would indicate the open chromatin regions. Given sufficient sequencing depth, the transcription factor footprinting could also be recovered. These pieces of information would complement the predictions from Gene Regulatory Network inference and verify the roles of candidate transcriptional regulators and the condition at which the promoter regions of the target genes are accessible. It is to be expected that the combination of Gene Regulatory Network and chromatin accessibility will advance understanding of the gene regulatory landscape of CAM in the near future.

General aims and hypotheses

This thesis aims to explore the question of regulatory control of CAM from the systems biology approach. The ultimate goal within and beyond this thesis is to identify the key regulators for CAM to function in a highly synchronised manner between multiple functional modules. It is apparent that there is temporal separation of key CAM genes at the level of enzymatic function. Since these temporal shifts are seen in enzymatic content and activation status as well as gene expression patterns, it suggests the complexity of the system is intertwined with transcriptional regulators, post-translational regulators and possibly epigenetics and small RNAs. With the author's appreciation for rigorous mathematical approaches, the second and third chapters (2 and 3) of the thesis are allocated to computational methods. However, although computational methods are powerful in the sense that they can tackle the tasks human cannot do by hand in a timely manner, these approaches could only provide an initial screening of the potential candidates or predictions that are yet to be verified by experimental manipulation. Hence, throughout this thesis, the computational tools were never used in isolation from the experimental data. The critical view of the author was that computational tools should be used, where possible, in conjunction with experimental data; computational outputs can be used to guide experimental design, while experimental results should be used to improve or validate the model or algorithm. The final results chapter (Chapter 4) of the thesis describes the initiation of an ATAC-sequencing project for CAM by purifying nuclei from non-transgenic *K. fedtschenkoi* for further downstream Tn5 transposase reactions and ATAC-seq analysis for the future (a promising piece of research beyond the PhD project, funded by the Cambridge Philosophical Society). The thesis contains three result chapters with their aims and hypotheses outlined in the following section.

Chapter 2: Ordinary Differential Equations representing CAM mesophyll metabolism

The aim of this chapter was to theoretically test whether the core component of existing systems dynamic model of CAM were sufficient to capture CAM behaviour, and to test the flexibility of parameterisation and optimisation steps.

Hypothesis 1: The ODE system is sufficient to capture the broad pattern of the CAM cycle but may be limited in terms of parameterisation and optimisation.

Hypothesis 2: The ODE system is limited by the higher level or more refined regulatory controls such as transcriptional controls and any associated metabolic or circadian feedback.

Chapter 3: Gene Regulatory Network from CAM timeseries transcriptome data

The aim of this chapter was to construct networks of gene regulation from timeseries transcriptome data using computational approaches for CAM for the first time.

Hypothesis 1: A Gene Regulatory Network can be constructed by dynGENIE3 algorithm using the available timeseries transcriptome data of *K. fedtschenkoi*, that is, the amount of data is suitable for the algorithm and the size of the computer cluster which could limit the computational power and storage memory.

Hypothesis 2: The output of a Gene Regulatory Network returns candidate Transcription Factors of key CAM genes for the following selected functional pathways: carboxylation, decarboxylation, circadian control and stomatal regulation.

Hypothesis 3: A number of these candidate Transcription Factors have transcription factor binding sites (TFBS) in the promoter regions of their corresponding target genes.

Hypothesis 4: There are shared transcription factors between multiple genes from the same pathway, and between genes from different pathways.

Hypothesis 5: Multiple transcription factors target a specific target gene to provide more than 1-to-1 control. The target genes expected to adopt this pattern include, but are not limited to, PEPC kinase (PPCK) and PPDK-RP.

Chapter 4: Nuclei isolation with sucrose-gradient and fluorescence-based spectral flow cytometer

The aim of this methodological chapter was to design and define the protocol for nuclei isolation for a CAM species, in preparation for ATAC sequencing.

Hypothesis 1: The combination of sucrose-gradient isolation and fluorescence-based spectral flow cytometer would yield purify nuclei of greater than 10% of the total number of particles in the nuclei extract.

Hypothesis 2: The combination of sucrose-gradient isolation and fluorescence-based spectral flow cytometer would facilitate the separation of chloroplasts out of the nuclei extract.

Hypothesis 3: The combination of sucrose-gradient isolation and fluorescence-based spectral flow cytometer would facilitate the separation of mitochondria out of the nuclei extracts.

**Additional background information of CAM is provided in Appendix B, which reproduces a peer-reviewed publication I authored.*

Chomthong M, Griffiths H. **2020**. Model approaches to advance crassulacean acid metabolism system integration. *The Plant Journal* **101**: 951–963.

Chapter 2

Ordinary differential equation modelling of Crassulacean acid metabolism to capture responses to perturbation conditions

Introduction:

Crassulacean acid metabolism (CAM) is a specialised form of photosynthesis documented in 6-7% of angiosperms with a strict temporal separation between the initial carboxylation enzyme, phosphoenolpyruvate carboxylase (PEPC), and the secondary carboxylation enzyme, Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase (Rubisco), which function during the dark and light period, respectively (Smith and Winter 1996). Whilst the nocturnal carbon dioxide (CO_2) fixation leading to vacuolar malic acid storage has been the most notable feature of CAM, dating back to leaf acidity observations in Roman times, with other specialised features of CAM also documented in terms of tissue succulence, nocturnal stomatal opening, root hydraulic responses to drought stress and the four Phase gas-exchange pattern (Osmond 1978; Lüttge 2004; Borland *et al.* 2009).

The combination of CAM metabolism and hydraulic characteristics have resulted in an improved water use efficiency (WUE) to be within the range of $6 – 30 \times 10^{-3}$ mol CO_2/mol H_2O, whereas the WUE of C3 and C4 are in the range of $0.6 – 1.3 \times 10^{-3}$ and $1.7–2.4 \times 10^{-3}$ mol CO_2/mol H_2O respectively (Black 1973; Lüttge 2004). This improved WUE has put CAM species of interest for deploying in marginal lands as an alternative form of biofuel (Borland *et al.* 2009; Owen *et al.* 2016). To this end, field-scale modelling has been developed to predict CAM productivity in response to environmental fluctuations (Hartzell *et al.* 2020).

In parallel to the global-scale productivity modelling, progress on CAM metabolic modelling has been made continuously. There are a number of modelling approaches addressing temporal orchestration and metabolite partitioning in CAM, which can be grouped into three categories. Firstly, flux balance analysis has captured the complete metabolic network under stoichiometric constraints and optimisation of the objective (Orth *et al.* 2010; Shameer *et al.* 2018; Töpfer *et al.* 2020). Secondly, a category includes mechanistic models which incorporate simplified mathematical representations of the CAM circadian rhythm (Blasius *et al.* 1998; Bartlett *et al.* 2014; Hartzell *et al.* 2018, 2020). Third, a systems dynamic representation of minimal CAM component biochemical and physiological pathways has allowed the fine-tuning of key parameters (Owen and Griffiths 2013). This latter approach has allowed the relative contribution to limitations such as stomatal sensitivity, carboxylation processes and vacuolar storage to predict the impact of succulence on gas exchange profiles (Owen and Griffiths, 2013).

In this study, the systems dynamic biochemical model of Owen and Griffiths (2013) was re-formulated into a mathematical rigorous form using ordinary differential equation (ODE) approaches. The ODE model was then used in simulating the expression of the CAM pathway in response to two studies representing contrasting experimental modes of manipulation: the imposition of atmospheric CO_2 transients (von Caemmerer and Griffiths 2009) and the knockdown expression of PEPC (Boxall *et al.* 2020). The ODE model, derived from the systems dynamic model (Owen and Griffiths, 2013), reduces the focus onto five key metabolic processes, which were shown to be suitable for manipulating CAM behaviour and verifying minimal CAM inputs. Having established a minimal model to capture outline behaviour, additional components allowing the molecular fine tuning of metabolism could be added to refine the model structure (Chomthong and Griffiths 2020). We hypothesised that the classical textbook description of the CAM pathway, with one key isozyme for each reaction, would be sufficient to capture the Phases of CAM under both standard conditions and following various manipulations.

Overall, the simulated results captured the CAM gas-exchange profile when compared with the raw data from these contrasting experimental manipulations. The updated system of ordinary differential equations (now developed as a more accessible format in Matlab) showed sensitivity to biochemical parameters and anatomical parameters, while addressing the significance of the interplay between two carboxylation enzymes: Rubisco and PEPC. Therefore, this improved version of the systems dynamic biochemical model has become a useful tool for the increasing numbers of studies perturbing CAM behaviour and provides an important *in silico* system for the CAM community to test environmental and molecular transformations (Yang *et al.* 2015; Hartwell *et al.* 2016).

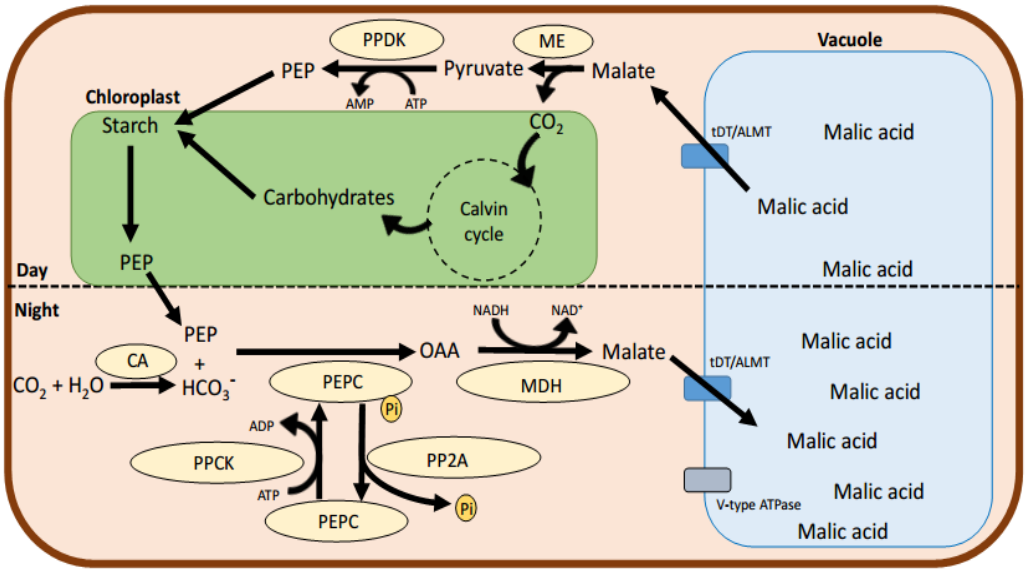


Figure 1 Crassulacean acid metabolism (CAM) pathway in *Kalanchoe*
During the night period, primary carboxylation reaction is mediated by PEPC while the final product in the form of malate is stored in the vacuole. In the subsequent light period, malate undergoes decarboxylation reaction and release CO_2 for secondary carboxylation by Rubisco to enter the Calvin cycle**.** Yellow components and blue components represent key enzymes and transporters respectively. CA Carbonic anhydrase, PEPC Phosphoenolpyruvate carboxylase, PPCK Phosphoenolpyruvate carboxylase kinase, PP2A Protein phosphatase 2A, MDH Malate dehydrogenase, ME Malic enzyme, PPDK Pyruvate phosphate dikinase, tDT Tonoplast dicarboxylate transporter, ALMT Aluminium-activated malate transporter and Pi Inorganic phosphate.

Material and methods:

1. Key experimental references

Two publications that framed the scope of our simulation were the CO_2 perturbation study on *Kalanchoe daigremontiana* by von Caemmerer and Griffiths in 2009, and the genetic perturbation study on *Kalanchoe laxiflora* by Boxall *et al.* in 2020. The first study set out to explore the stomatal responsiveness by manipulation of the environment within a leaf chamber, whilst the second study set out to determine the significance of phosphoenolpyruvate carboxylase (PEPC) using antisense RNA interference methods. The key findings from these two publications highlighted CAM plasticity under different types of manipulation (physiological and molecular) whilst framed by species-specific gas-exchange patterns.

2. Ordinary differential equation (ODE) modelling structure

A system of ordinary differential equations (ODE) was developed from the original systems dynamic model of Owen and Griffiths (2013) so as to capture the mesophyll photosynthetic process of CAM plants mechanistically, based around empirical modules framing stomatal behaviour and key enzymatic processes which are used to entrain the timing of the four CAM four-Phases (Osmond, 1978). The parameter descriptions are summarised in Table 1 (Matlab scripts are provided in Appendix A).

The system of five state variables includes

Equation 1:

dy_1/dt = gs*(Cca-Conversion*y_1)

-gm*(y_1*Conversion-y_2/Chlorenchyma Volume/Henry's law constant

Equation 2:

dy_2/dt = gm*(y_1*Conversion-y_2/ Chlorenchyma Volume/ Henry's law constant)

+Decarboxylation activation term*{Vdmax * y_3/ (Kd+y_3)} +Vm

- Rubisco activation term*(Vcmax * y_2/ Chlorenchyma Volume) / (Kc

+ y_2/ Chlorenchyma Volume) – PEPC reaction;

Equation 3:

dy_3/dt = PEPC reaction

- Decarboxylation activation term*{Vdmax * y_3/ (Kd+y_3)}

-Malic acid influx rate*(1- y_4/Xvmax)*y_3

+Malic acid efflux rate*(y_4/Xvmax)

Where

PEPC reaction = PEPC activation term*(Vpmax * y_2/ Chlorenchyma Volume) /

{Kp*(1+y_3/Ki) + y_2/ Chlorenchyma Volume}

Equation 4:

dy_4/dt = Malic acid influx rate*(1-y_4/Xvmax)*y_3

-Malic acid efflux rate*(y_4/Xvmax)

Equation 5:

dy_5/dt = Rubisco activation term*Vcmax*y_2/Chlorenchyma Volume

/ (Kc+y_2/Chlorenchyma Volume)-Vm

Where

y_1 represents intercellular CO_2, y_2 represents intracellular CO_2

y_3 represents cytosolic malate pool, y_4 represents vacuolar malate pool

y_5 represents chloroplastic carbohydrate pool

3. Parameter setting

Within the five differential equations, there are the total number of 19 parameters. The maximum likelihood estimation method has a maximum limit of 5 learning parameters based on this model structure, hence the priority was given to the manipulation of key biochemical parameters and malate transport rates, with the remaining parameters were fixed according to published values (for details, see Table 1). Maximum likelihood estimation was performed against the representative data published by Boxall et al 2020. Parameters Vmax and K represent the maximum rate of enzyme activity and effective K_m constant for its substrate, respectively; these were written in a general form of density-dependent feedback of Michaelis-Menten kinetics (Owen and Griffiths 2013).

Table 1 Parameter descriptions and references

Parameter	*K. daigremontiana* with constant atmospheric CO_2	*K. laxiflora* wild-type
Biochemical parameters		
Maximum rate of PEPC activity (V_{pmax}) Unit: µmol m^{-2} s^{-1}	80 Minimised RMSE	5.0 Maximum likelihood estimation
Michaelis constant for PEPC (Kp) Unit: µM	12.0 (Nott and Osmond 1982)	12.0 (Nott and Osmond, 1982)
Maximum rate of Rubisco activity (V_{cmax}) Unit: µmol m^{-2} s^{-1}	25 Minimised RMSE	10.82 Maximum likelihood estimation
Michaelis constant for Rubisco (K_c) Unit: µM	10.8 (Griffiths *et al.,* 2008)	10.8 (Griffiths *et al.,* 2008)
Michaelis-Menten constant for malic acid-induced inhibition of activated PEPC (K_{ia}) Unit: µM	8000 (Anne M Borland and Griffiths 1997b)	8000 (Borland and Griffiths, 1997)
Maximum rate of malic acid decarboxylation (V_{dmax}) Unit: µmol m^{-2} s^{-1}	9 (Borland and Griffiths, 1997, Maxwell *et al.,* 1999, Griffiths *et al.,* 2002)	5.5 Maximum likelihood estimation
Michaelis-Menten constant for Malic enzyme (Kd) Unit: µM	Kd = 11 (Artus and Edwards 1985)	Kd = 11 (Artus and Edwards 1985)

Mitochondrial respiration rate (V_m) Unit: µmol m^{-2} s^{-1}	0.6 (Owen and Griffiths 2013)	0.6 (Owen and Griffiths 2013)
Malic acid efflux rate Unit: µmol m^{-2} s^{-1}	40 (Owen and Griffiths 2013)	33.19 Maximum likelihood estimation
Malic acid influx rate Unit: µmol m^{-2} s^{-1}	2 (Owen and Griffiths 2013)	7.34 Maximum likelihood estimation
Anatomical parameters		
Mesophyll Conductance (g_m) Unit: mol m^{-2} s^{-1} bar^{-1}	0.058 (Griffiths *et al.,* 2008)	0.058 (Griffiths *et al.,* 2008)
Maximum vacuolar capacity (X_{vmax}) Unit: µmol of malate m^{-2}	150000 (Borland and Griffiths 1997)	150000 (Borland and Griffiths 1997)
Succulence (S) Unit: kg m^{-2}	2.8 (Griffiths *et al.,* 2008) (Maxwell *et al.* 1997)	2.8 (Griffiths *et al.,* 2008) (Maxwell *et al.* 1997)
Stomatal aperture parameter (SA)*	Time-series function (Von Caemmerer and Griffiths 2009)	Time-series function (Boxall et al 2020)
Physical parameters		
Henry's law constant (mol L^{-1} bar^{-1})	0.03445	0.03445
Conversion for converting CO_2 (µmol m^{-2}) to CO_2 (µbar)	0.025 (Owen and Griffiths 2013)	0.025 (Owen and Griffiths 2013)

Ratio of conductance to water vapour : conductance to CO_2	1.6	1.6
Atmospheric CO_2 (Cca) Unit: µbar	400 (Von Caemmerer and Griffiths 2009)	400 (Boxall et al 2020)

*****Stomatal aperture parameter (SA) was parameterised with experimental data from each experimental set-up using MATLAB curve-fitting tool with exponential functions for Phase I,II,III, and IV separately (Supplementary data).

4. Establishing settings for Initial conditions

The ODE system could be simulated across multiple 24-hr cycles to reflect the repeated diel behaviour of the biological system. However, simulation outputs at the corresponding time-of-day from different diel cycles can be different despite the same setting of time-dependent parameters. This is due to the different values of the state-variables at the start of each diel cycle influencing the simulation outputs of each diel cycle.

These simulation results have emphasised the significance of the state-variables at the start of every diel cycle as well as the choice of initial conditions at the start of the simulation. This has allowed for 2 model development approaches. The first approach is to rely on the available reference source of the initial conditions, whereas the second approach is to search for alternative sets of initial conditions that can return repeated diel simulation outputs. The primary settings for the initial condition were referenced to the system dynamic model outputs (Owen 2013).

Under the first approach, biochemical parameters can be estimated from the maximum likelihood method. This is advantageous because it allows model users to account for uncertainty in biochemical measurement and variability between species, varieties and growing condition. However, the values of the state-variables at the end of the simulation cycle (t = 24 hr) were not the same as the values of the state-variables at the beginning of the simulation (t = 0 hr) which results in 2 consecutive diel simulations returning different outputs at the corresponding time-of-day. Therefore, this approach would restrict the use of the ODE model to only 1 diel cycle. This approach was adopted for the *Kalanchoe daigremontiana* simulation.

Under the second approach, an alternative set of initial conditions is selected from the values of state-variables after the simulation has entered the repeated diel cycle behaviour (> 100 diel simulation cycles). This methodology would not be subjected to the uncertainty of the initial condition with reference to the system dynamic model, but it is under the influence of the choice of parameters for the first cycle which also determine whether the maximum likelihood estimation would converge (Mantel and Myers 1971). This second approach was adopted for *Kalanchoe laxiflora* simulation.

5. Assessing the model performance

Normalised root mean square error (RMSE) was calculated between simulations of CO_2 assimilation and actual experimental data (Von Caemmerer and Griffiths, 2009; Boxall *et al.* 2020) according to the following formula.

$$\mathrm{RMSE} = \sqrt{(\sum_{i=1}^{n}(w_i - y_i)^2)/n}$$

Where

y_i is the observed value for the i^{th} observation,

w_i is the predicted value,

n is the total number of observation

Results:

1. Simulation of *Kalanchoe daigremontiana* under an unperturbed condition

The original gas exchange data from the von Caemmerer and Griffiths (2009) study were kindly supplied by Prof S von Caemmerer, and were initially plotted as a control dataset without the additional manipulations in CO_2 concentration that had been applied (Figure 2). The simulation results captured the classical CO_2 influx peak during Phase I (night time) of the succulent *Kalanchoe daigremontiana* (Figure 2). The initial dip at the start of the dark period represents the adjustment to the pre-set parameters, described in Section 4 above. Towards the end of the night-time, the second CO_2 influx peak was produced from the simulation as a result of PEPC activity with the increase in stomatal conductance between 11 hr and 12 hr, a feature that had been taken into the parameterisation step to represent expected pattern of CAM Phase II. In this scenario, Phase II occurred at night coinciding with the increased of stomatal conductance of the experimental data. The marked decline of the assimilation rate at the transition into Phase III was due to the stomatal closure, absence of PEPC activity and the low Rubisco activity level. Subsequently, during 12 hr to 18 hr, the ODE simulated CO_2 influx was at its minimum within the range of -0.5 and 0.5 μmol m^{-2} s^{-1}.

The daytime reverse flux of CO_2 from intercellular spaces into the atmosphere during Phase III was seen in both the raw data and captured by the model (Figure 2). Because the decarboxylation rate was higher than internal CO_2 re-uptake by Rubisco, the accumulated intercellular CO_2 was higher than atmospheric CO_2 hence constituting a reverse flux though incompletely-closed stomata. However, as Rubisco became fully activated at the middle of the day (t = 18 hr), the secondary CO_2 fixation was still mostly supplied by the internal CO_2 behind the closed stomata.

After the plant entered Phase IV at t = 20 hr (Figure 2), coinciding with the stomatal re-opening behaviour, intercellular CO_2 (Ci) level was below atmospheric level allowing CO_2 to flow inwards for the direct fixation by Rubisco. This atmospheric CO_2 assimilation during Phase IV can be observed in some CAM species under well-watered condition. It is worth emphasising that the Rubisco activation state was set to reached its maximum activity at the middle of the day (t = 18 hr) and subsequently declined towards the end of the light period. Nonetheless, the CO_2 assimilation was still sustained by the declining Rubisco activity provided that the activity was sufficiently high to keep Ci lower than atmospheric CO_2 concentration.

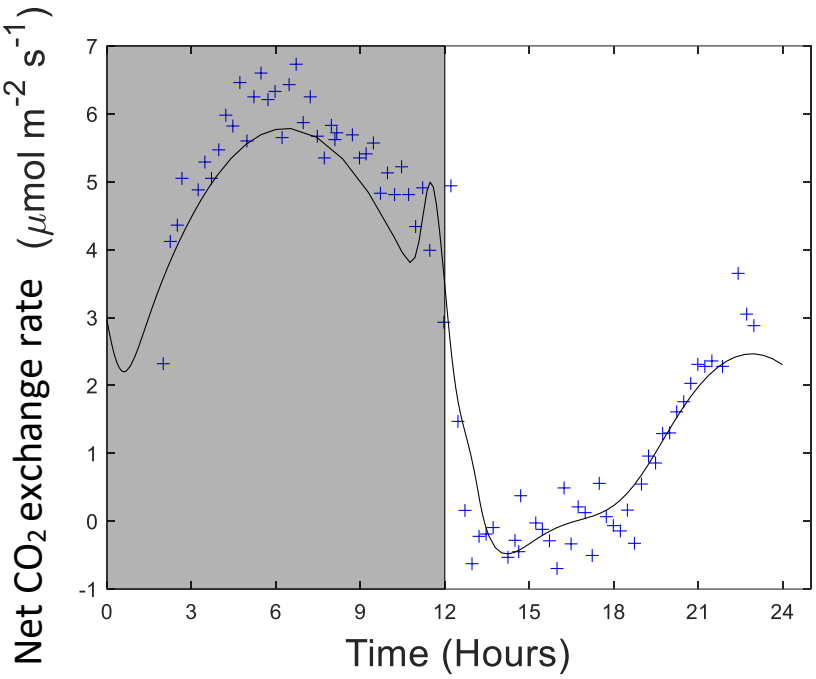


Figure 2: Performance of the ODE simulation gas-exchange profiles compared to experimental data from *Kalanchoe daigremontiana*. The black line represents ODE modelling outputs with optimised parameter setting (supplementary data); data points (blue crosses) represent experimental data across 1 cycle of 12-hr light/12-hr dark. Experimental data were obtained from Von Caemmerer and Griffiths 2009. The apparent Phase II occurred at night as a result of an increased stomatal conductance (obtained from experimental data of Von Caemmerer and Griffiths 2009) at the corresponding time-point.

Simulation of *Kalanchoe daigremontiana* under CO_2 transient perturbation conditions

Subsequently, the short-term manipulations in atmospheric CO_2 concentration was introduced to test the model simulation on *Kalanchoe daigremontiana* data, to mimic the experimental study by Von Caemmerer and Griffiths (2009) which compared stomatal responsiveness of *Kalanchoe* species by using atmospheric CO_2 transients (400 µbar to 95 µbar) over 30 minutes to compare the response and recovery of gas exchange in terms of assimilation and stomatal conductance. The transients were applied twice during Phase I, once during Phase III and once during Phase IV of the CAM cycle.

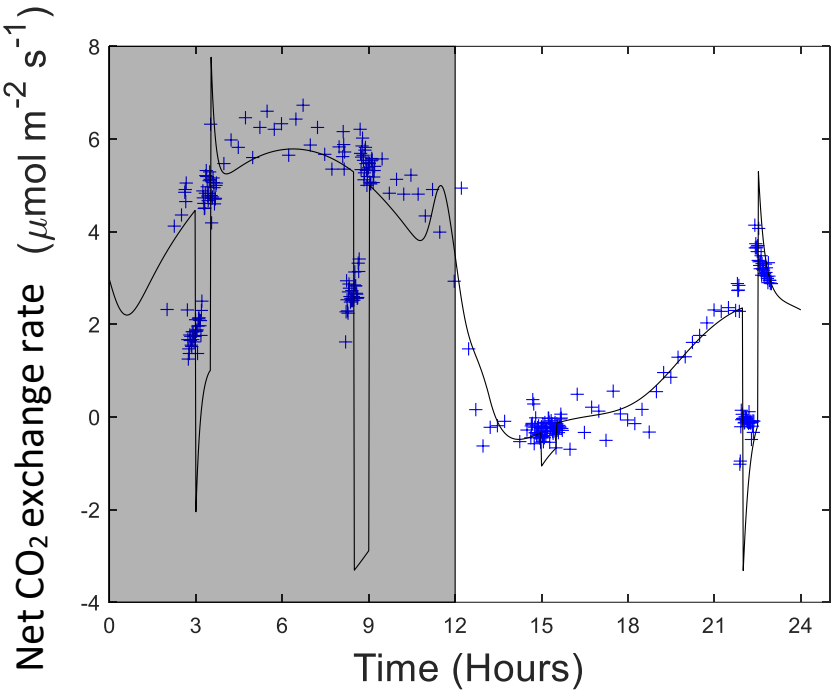


Figure 3: Performance of the ODE simulation gas-exchange profiles compared to experimental data from *Kalanchoe daigremontiana* under CO_2 perturbation conditions. The black line represents ODE modelling outputs with optimised parameter settings (supplementary data); data points (blue crosses) represent experimental gas exchange data across 1 cycle of 12-hr light/12-hr dark. The atmospheric CO_2 was set at 400 µbar with four 30-minute transient drop to 95 µbar. Experimental data were obtained from Von Caemmerer and Griffiths 2009. Data were collected at every minute during the 30-minute CO_2 transient perturbation and the following 30 minutes.

The CO_2-transient simulation during each of the four phases captured the general profile of background gas exchange data for *K. daigremontiana*. However, an over-sensitive response was seen in the 30-minute periods corresponding to the low CO_2 transients, where the model predicted both a more extreme reduction in assimilation rate, and the overshoot of the recovery at the first and last transient periods, before re-establishing values close to the measured data (Figure 3). In general, the model captures stomatal sensitivity and impact on CO_2 assimilation

rate more consistently for daytime gas exchange (Phase III and IV), and suggests that CO_2 uptake would be more responsive to a transient reduction in ambient CO_2 concentration during the dark. This response may be explained by size of the simulation time-step as well as the time-lag for leaf chamber equilibrium adjustment, which would not have captured the predicted magnitude of transient responses (Weiss *et al.* 2009).

3. Simulation of *Kalanchoe laxiflora* with a genetically perturbation condition

The model was then applied to an additional data set in which PEPC activity had been reduced by genetic modification in *K. laxiflora.* In figure 4, a 24-hour cycle of the original diel CO_2 assimilation data from Boxall et al., (2020) is repeated to provide 5 successive replicate day-night cycles. Values for wild-type (Figure 4A) and PEPC transgenics (Figure 4B) are shown and compared (Figure 4C), with the model simulations shown as continuous lines. In figure 4A, the major CO_2 assimilation processes of wild-type occurred during the night time which agreed to the simulation outputs. The simulation thus successfully captured the biochemistry of malate inhibitory effect on PEPC enzyme and the associated Phase I carboxylation rate in wild-type (Figure 4). The simulation over-estimated gas exchange activity by day for the wild-type plants, and suggested that net CO_2 uptake might continue into Phases III and IV, with a positive offset of 0.75 µmol m^{-2} s^{-1} as compared to the experimental data showing CO_2 release by day (Figure 4A). The non-zero assimilation rate during the daytime of the simulation was a result of Rubisco activity and the intercellular CO_2 (Ci) value being lower than atmospheric CO_2 as a carried on effect from end of the night period.

Simulation of transgenic RNAi lines with no detectable PEPC activity (*rPPC1-B*) showed a lower nocturnal CO_2 assimilation rate than wildtype throughout Phase I, agreeing with the experimental data (Boxall *et al.*, 2020). These simulated gas-exchange pattern of the *rPPC1-B* knockdown line reflected the major assimilation phase during the day-time. The lack of PEPC activity reported in experiment and implemented in the model was sufficient to reduce the assimilation peak characteristic seen in wild-type down to near-zero night-time assimilation under the simulation of the *rPPC1-B* line. The stomatal opening behaviour during the day-time had allowed direct atmospheric CO_2 fixation by Rubisco. Nonetheless, the optimised parameter setting to capture initial assimilation during the day of 7 µmol m^{-2} s^{-1}, had required Rubisco to have higher activity than wild-type at the beginning of the day.

The alteration in carboxylation partitioning between PEPC and Rubisco in the PEPC knockdown line (*rPPC1-B*) as compared to the wild-type can influence stomatal behaviour through intercellular CO_2 (Ci) signalling. This effect was incorporated into the model through stomatal conductance parameterisation. The diurnal opening of stomata for the PEPC knockdown line (*rPPC1-B*) was associated with the Rubisco demand, hence producing direct CO_2 assimilation during the light period. The Rubisco activation status was boosted to the maximum daytime value at the 30-minute transition from dark into light period as a metabolic response to the night-time carbon 'famine' state as documented for CAM species (Maxwell *et al.* 1999; Griffiths *et al.* 2002). Nonetheless, the discrepancies between the simulation and experimental data at the second half of the light period could be explained by more rapid Rubisco inactivation *in vivo* due to the effect of ADP/ATP ratio and redox potential on Rubisco inhibition, the saturation of electron transport rate, or malate transport rate. All of these suggestions would allow for finer Rubisco parameter tuning, if experimental data for these responses were available.

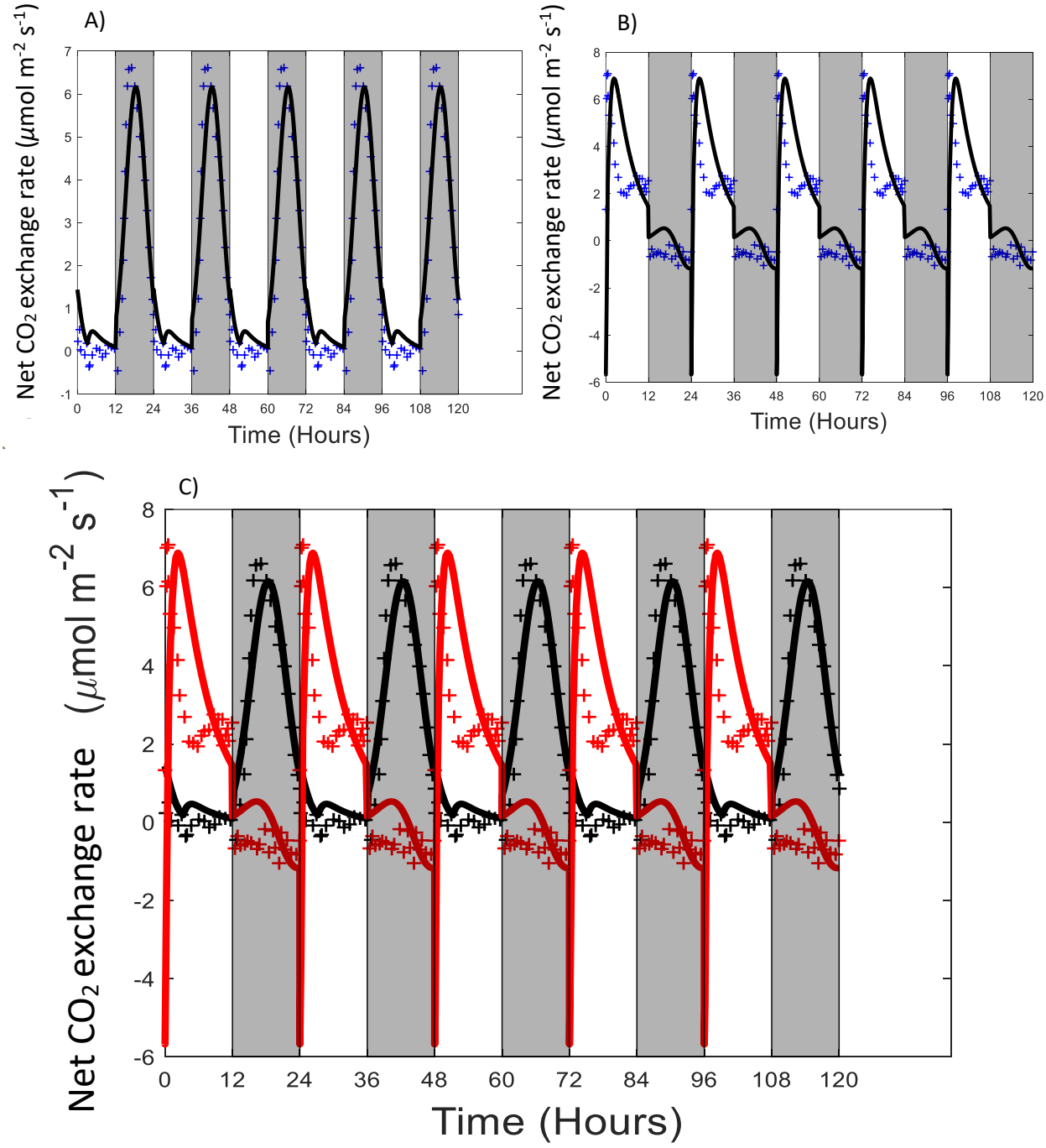


Figure 4: Performance of the ODE simulation gas-exchange profiles compared to experimental data from *Kalanchoe laxiflora*. (A) Wild-type line, **(B)** PEPC1 knockdown line (*rPPC1-B*). The black line represents ODE modelling outputs with optimised parameter setting (supplementary data); data points (blue crosses) represent experimental data extended into 5 cycles of 12-hr light/12-hr dark. Experimental data were obtained from Boxall *et al*. 2020. **(C)** Overlay plot of the Wild-type line (black) and PEPC1 knockdown line rPPC1-B (red). Solid lines represent the ODE simulation results. Plus markers (+) represent the experimental data from Boxall *et al*. 2020.

A notable feature of the *rPPC1-B* simulation was the discrepancies in the maximum assimilation rate and the slope of the diurnal phase between the simulation and the representative *rPPC1-B* plant (Figure 5). The simulation set reached the maximum at t = 2.303 hr with assimilation rate of 6.892 µmol m^{-2} s^{-1} whereas the representative plant reached the maximum earlier at t = 0.570 hr with assimilation rate of 7.096 µmol m^{-2} s^{-1}. In terms of the slope of the diurnal phase, the magnitude of the slope of the simulation set was 0.621 µmol m^{-2} $s^{-1}hr^{-1}$ whereas that of the representative plant was 1.367 µmol m^{-2} s^{-1} hr^{-1}.

To attempt to explain these discrepancies from the biological perspective, the stomatal conductance patterns were revisited. The corresponding stomatal conductance of the representative *rPPC1-B* plant reached the maximum at t = 0.303 hr which was closer to the maximum assimilation timing of the representative plants but preceded the maximum assimilation timing of both the representative plant and the simulation set (Figure 5). This led to further explanation that the *'delay'* in the simulation set may be explained by any parameters that influence the substomatal cavity CO_2 level in such a way that CO_2 remains low for a longer period of time in the simulation scenario than in the representative plant. The anatomical parameters and parameters related to the diurnal CO_2 supply and drawdown were systematically adjusted to investigate the effect they may have on the slope of the diurnal phase CO_2 assimilation (Figure 6). Within the scope of the parameter optimisation in this study, the results suggested that the discrepancies between the simulation set and the representative plant cannot be resolved by parameterising the parameters in this ODE system. Further exploration on how sensitive the model is to each parameter (i.e. the parameter sensitivity analysis) is to be presented in the subsequent section.

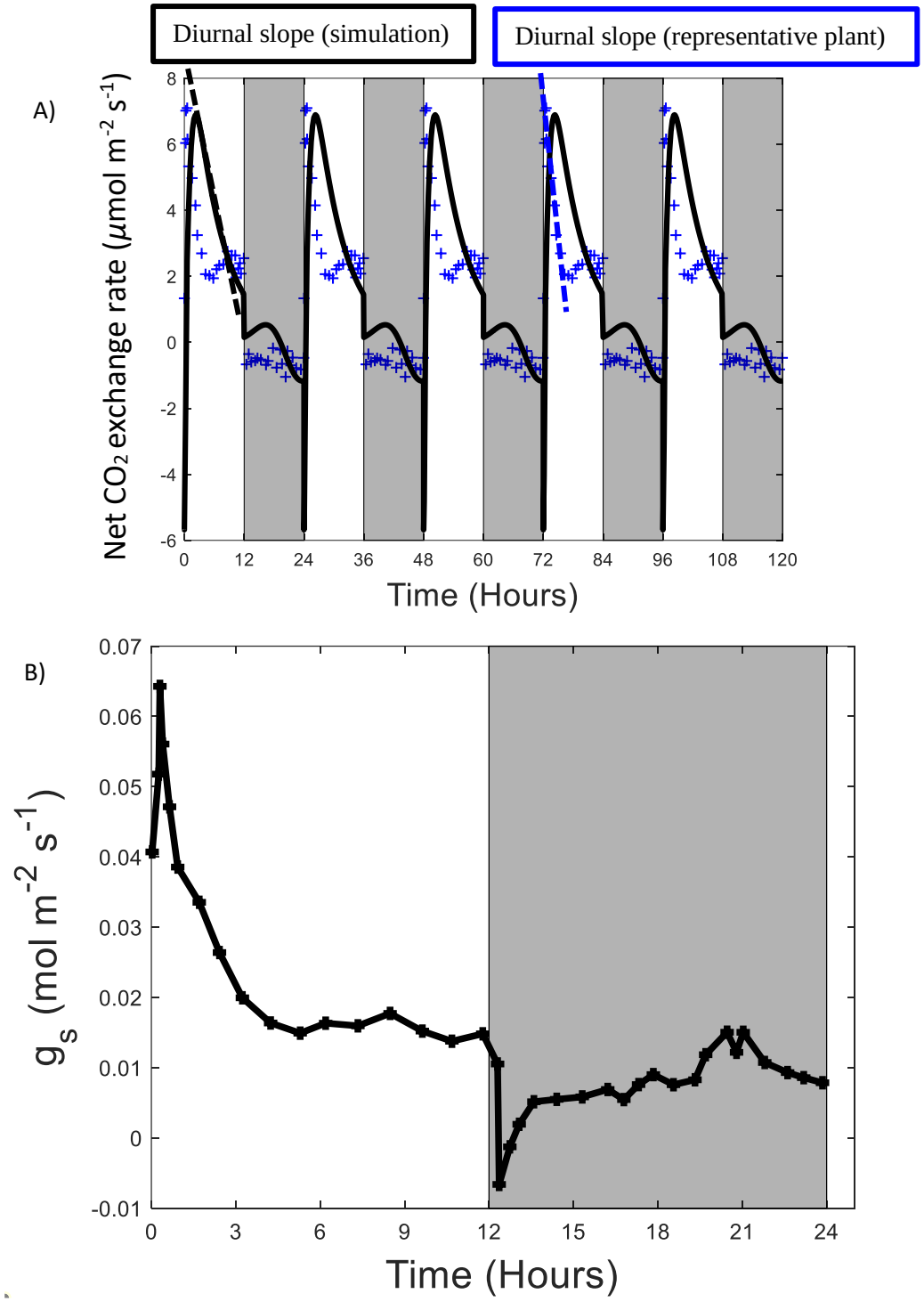


Figure 5 Diurnal slope of the simulation and the representative *rPPC1-B* plant and the associated stomatal conductance patterns. A) PEPC1 knockdown line (*rPPC1-B*). Black line represents ODE modelling outputs with optimised parameter setting (supplementary data); data points (blue crosses) represent experimental data extended into 5 cycles of 12-hr light/12-hr dark. Dashed lines (black and blue) represent the slopes of ODE output and experimental data, respectively. **B)** Stomatal conductance (g_s) from the representative *rPPC1-B* plant. This experimental g_s pattern was used in the ODE setting.

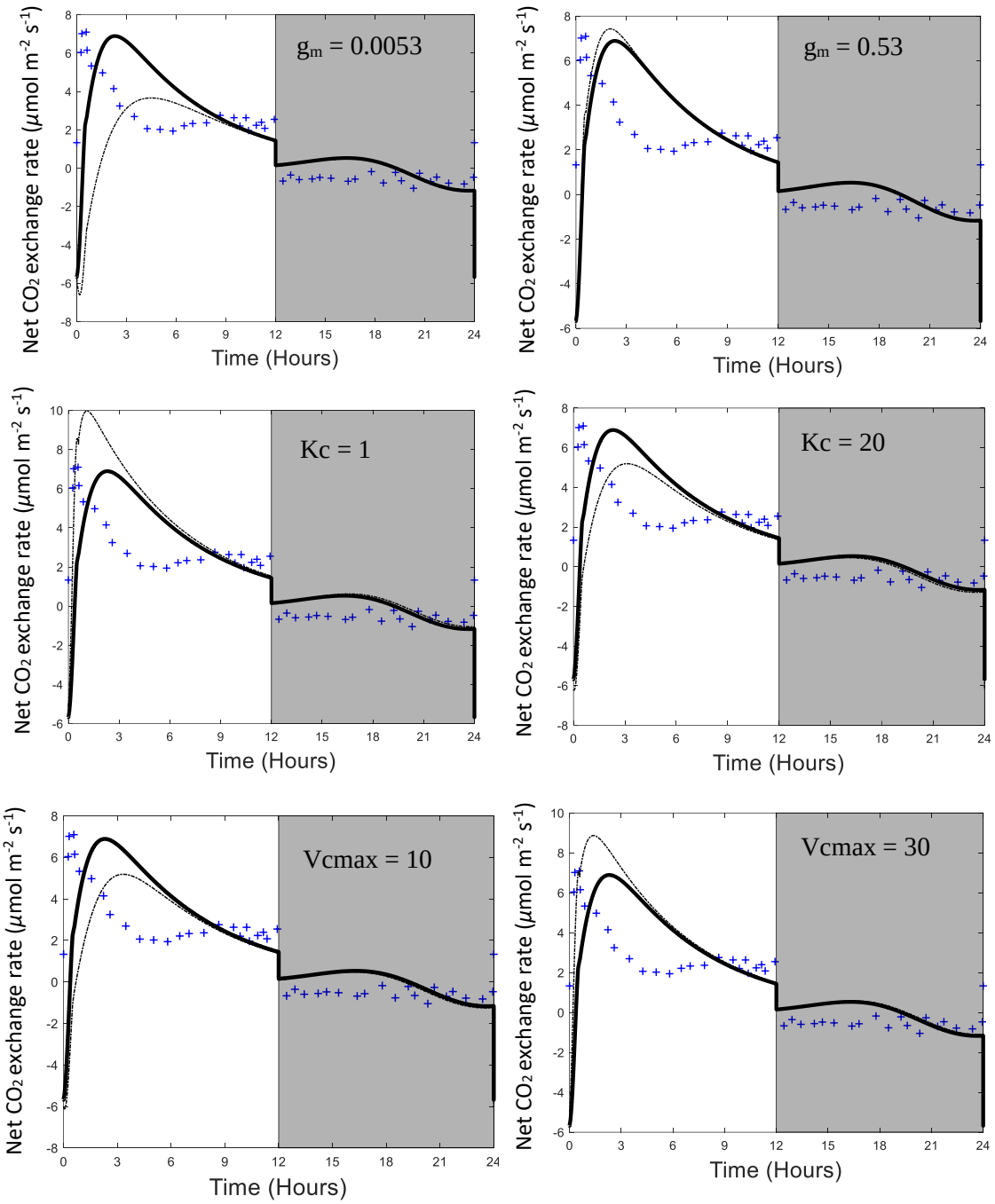


Figure 6 The effect of anatomical and Rubisco parameters on the simulation of PEPC1 knockdown line (*rPPC1-B*). Black solid line represents ODE modelling outputs with optimised parameter setting (supplementary data); data points (blue crosses) represent experimental data; black dashed line represents the ODE modelling outputs with the adjusted parameter setting specified in each plot. The default parameter setting of each parameter is as following; g_m = 0.053 mol m^{-2} s^{-1} bar^{-1}, Kc = 10.8 µM, and Vcmax = 15 µmol m^{-2} s^{-1}.

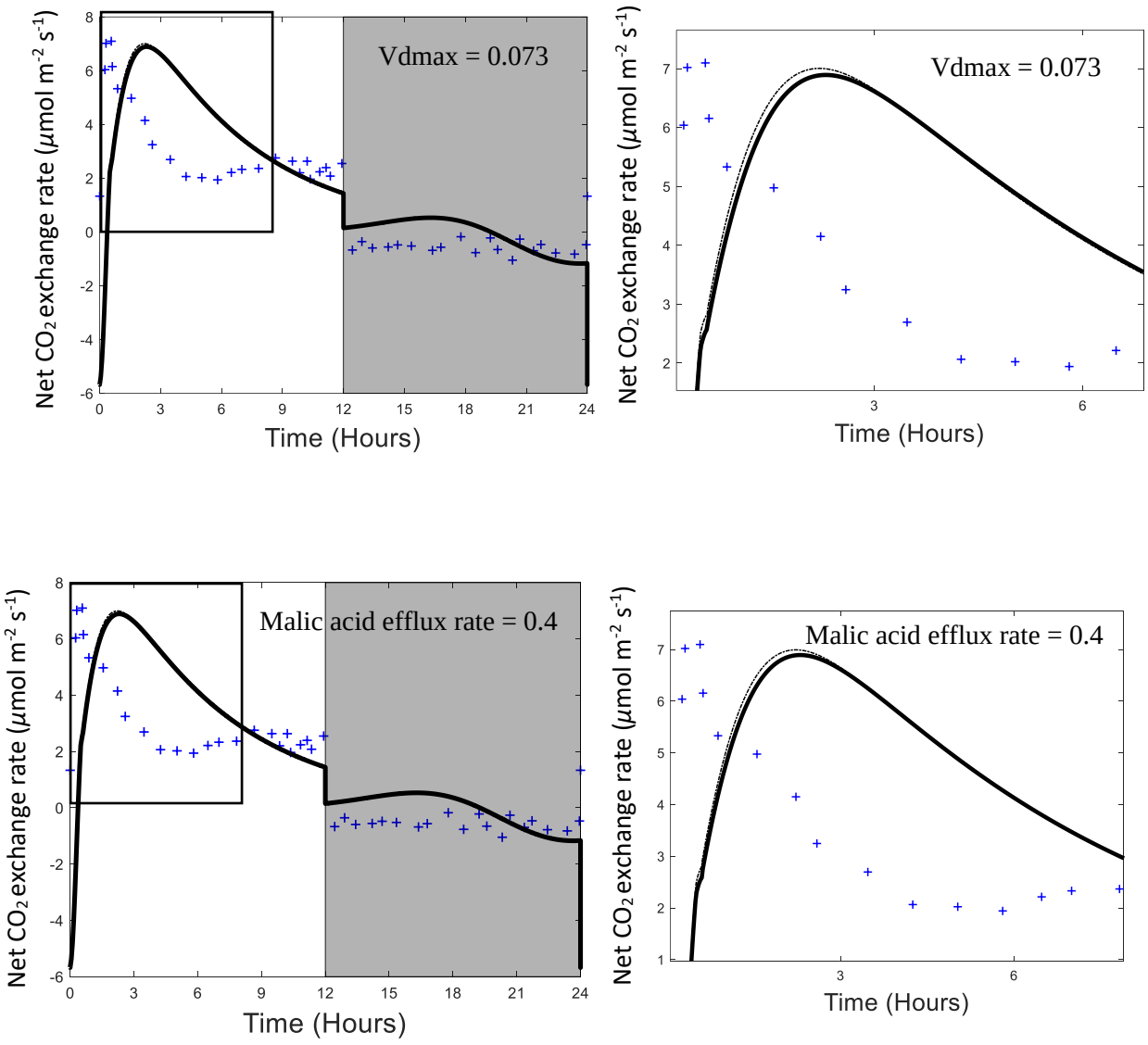


Figure 7 The effect of the decarboxylation and malate efflux parameters on the simulation of PEPC1 knockdown line (*rPPC1-B*). Black solid line represents ODE modelling outputs with optimised parameter setting (supplementary data); blue marks represent experimental data; the black dashed line represents the ODE modelling outputs with the adjusted parameter setting specified in each plot. The default parameter setting of each parameter is as following; Vdmax = 7.27 µmol m^{-2} s^{-1}, and Malic acid efflux rate= 40 µmol m^{-2} s^{-1}. Panels on the right column are the zoom-in plots of the panels on the left column.

An additional explanation for these discrepancies also relates to the nature of the primary data provided by Dr Hartwell and colleagues: the gas exchange profile of the *rPPC1-B* mutant (Figure 4B), showed a severely truncated degree of daytime CO_2 uptake, more consistent with the mid-drought period when measured with a whole plant gas exchange cuvette in the original data, (see Boxall *et al.* 2020, Figure 4D). Although the LICOR 6400 gas exchange measurements were made on attached leaves (Boxall *et al.* 2020), it seems likely that the reduced daytime CO_2 assimilation seen in the transgenic line is indicative of drought stress. If one compares the gas exchange profiles for data from the droughted experiment (compare day 1-6 with day 12- 18), there is a dramatic reduction in the extent of daytime gas exchange for *rPPC1-B* mutant as drought intensifies (Boxall *et al.* 2020, Figure 4D).

Parameter sensitivity

The sensitivity of the predicted CO_2 assimilation to ODE model parameters is illustrated in Figure 8. All four Phases of CAM show sensitivity to biochemical parameters (Carboxylation rate, Decarboxylation rate, Mitochondrial respiration rate and Malate transport rate) as well as anatomical parameters (Succulence, Vacuolar capacity and Mesophyll conductance). Importantly, the gas-exchange simulation outputs are sensitive to carboxylation enzyme activities. This emphasise the future application of this ODE model structure which allows for an empirical stomatal conductance parameterisation, yet the output is still responsive to other key parameters especially the PEPC and Rubisco activities.

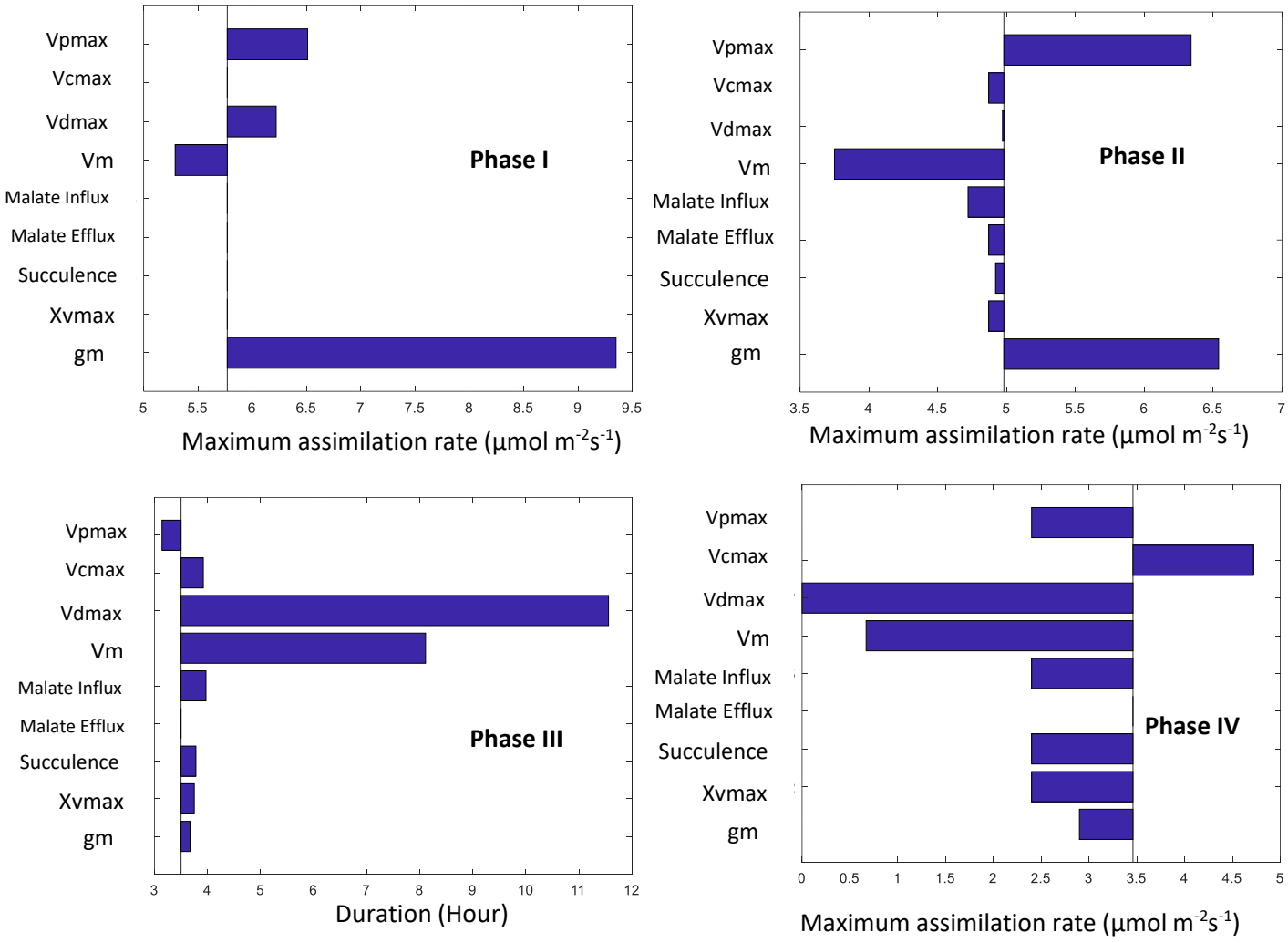


Figure 8: Parameter sensitivity of the ODE simulation gas-exchange profiles. The horizontal bars represent the changes in gas-exchange characteristics when the base-line simulation was substituted with 10-times original parameter setting of nine parameters labelled on the Y-axis as compared to the original gas-exchange characteristic indicated with the vertical base-lines of 5.77 µmol $m^{-2}s^{-1}$, 4.98 µmol $m^{-2}s^{-1}$, 3.50 hour and 3.46 µmol $m^{-2}s^{-1}$ for Phase I, Phase II, Phase III and Phase IV respectively. The base line simulation was *Kalanchoe daigremontiana* under a constant atmospheric CO_2 condition. The sensitivity parameters include: Maximum rate of PEPC activity (Vpmax), Maximum rate of Rubisco activity (Vcmax), Maximum rate of decarboxylation activity (Vdmax), Mitochondrial respiration rate (Vm), Malate influx rate, Malate efflux rate, Succulence, Vacuolar capacity (Xvmax) and Mesophyll conductance (gm). The stomatal conductance (gs) were set according to the measured value from each of the experiments.

The parameter sensitivity showed both expected and unexpected trends. The expected trends include the positive responses of night time assimilation (Phase I and II) to the increased PEPC activity, the positive responses of the daytime assimilation (Phase III and IV) to the increased Rubisco activity, and the positive response of Phase III to the increased decarboxylation activity. However, the increased PEPC activity also shortened the length of Phase III and reduced the Phase IV maximum assimilation. A possible explanation is that higher PEPC activity led to a greater inward CO_2 flux thus leaving CO_2 available for Rubisco at the transition of Phase II to III, thus delaying the onset of Phase III. On the contrary, the negative influence of PEPC activity on Phase IV assimilation can only be explained through the adjusted balance between intercellular CO_2, intracellular CO_2, and malate pool through equation 2 and 3 (Methods section 2). A similar explanation also applies to the effect of increased Rubisco activity on the decreased maximum assimilation during Phase II. In contrast, the negative response of Phase IV to the increased decarboxylation rate can be explained by the prolonged Phase III with backward diffusion which extended into and shortened Phase IV. Figure 8 indicated the lack of Phase IV with zero assimilation rate. This suggested that the decarboxylation rate was significantly higher than the Rubisco fixation rate which led to the CO_2 loss through open stomata.

Secondly, the effect of mitochondrial respiration (Vm) on assimilation rate can be interpreted in relation to the saturation level of the carboxylation enzymes. If the carboxylation enzymes are operating near their saturation level, increasing Vm would prevent more CO_2 from entering the leaves, thus resulting in the decreased in assimilation rate from the atmosphere. However, the respired CO_2 provided internal resource for the carbon fixation process, thus resulting in the higher total CO_2 available during the prolonged Phase III.

Thirdly, malate influx, succulence and vacuolar capacity shared the same trends (Figure 8). Increased values of these parameters allowed more malate to be stored during the night and sustained the prolonged Phase III. Because malate continues to be released during Phase IV, the build-up in CO_2 prevented the inward flow of atmospheric CO_2 thus led to the apparent lower CO_2 assimilation from the atmosphere. However, the greater amount of storage does not ensure greater maximum assimilation rate during Phase I because it is reflected by the area under the graph rather than the maximum point. On the other hand, the responses to malate efflux indicate that the original parameter setting was at its maximum limit. Thus, increasing the malate efflux further had negligible effects on the simulation outputs.

Finally, increased mesophyll conductance is expected to facilitate the CO_2 flow. This was the case for Phase I, II and III but not Phase IV. When mesophyll conductance was increased, the maximum assimilation rate during Phase IV reached lower values 1hr later than the original simulation. This could be explained by the fact that mesophyll conductance also facilitates the backward diffusion during Phase III. Therefore, the system took longer time to restore CO_2 back to the inward flow direction and Phase IV started after Rubisco has passed its maximum activity rate, resulting in the lower Phase IV maximum assimilation even though the mesophyll conductance was higher.

The effect of the relative strength between the two carboxylation enzymes on the gas-exchange simulation is summarised in Figure 9. The simulation with varying carboxylation strength (Black solid lines) were compared to baseline simulation (Blue dashed line) showing that simulation outputs are responsive to the carboxylation activity in such a way that nocturnal assimilation rate and diurnal assimilation rate are positively dependent on the PEPC activity and Rubisco activity, respectively. Firstly, the increase in PEPC strength by 50% at each step in the horizontal direction affects the simulation output in terms of increasing the night-time assimilation rate until the simulation reached t = 12hr which is the transition point from PEPC to Rubisco activities. Secondly, the decrease in Rubisco strength by 50% at each step in the vertical direction affects the simulation output in terms of decreasing the day-time assimilation rate during Phase IV when stomata are re-opened. Finally, the diagonal line from top left to bottom right reflects the upregulation of CAM and down regulation of Rubisco activities which results in the sequentially increase of Phase I fixation and decrease of Phase IV. An important feature emerging from the ODE simulation is that the Rubisco activities during day-time can be the cause of CO_2 loss from the backward diffusion process if the accumulated Ci during Phase III was higher than atmospheric CO_2 (Figure 8: subplot with Vcmax = 12.5). Hence, the simulation suggested that the down-regulation of Rubisco throughout the diurnal cycle is not beneficial for the overall photosynthetic yield, instead the maximum enzyme activities with strict temporal control are required for an efficient operating system.

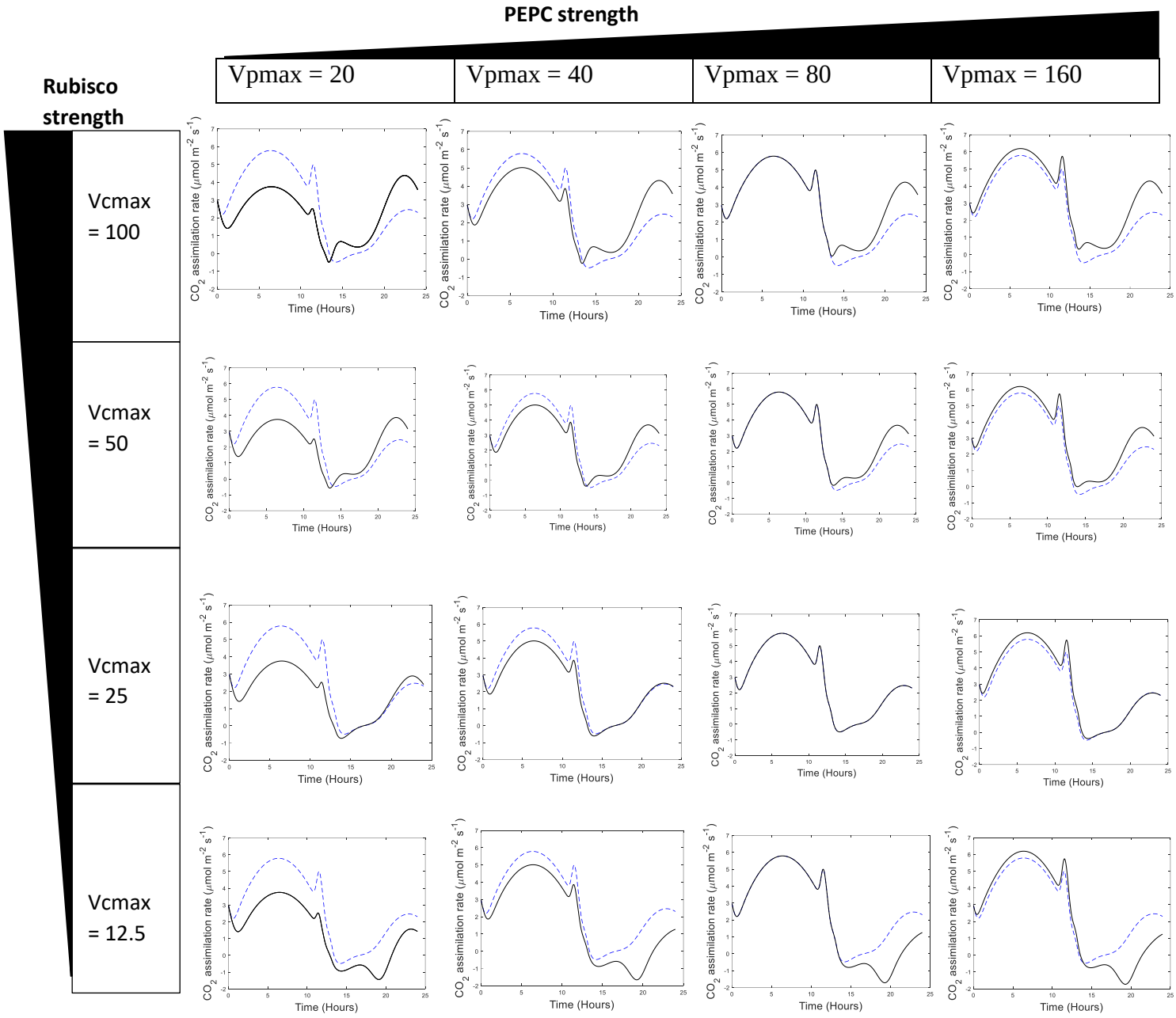


Figure 9: Gas-exchange pattern simulation along the gradient of Rubisco and PEPC strength. The enzyme strength is expressed in terms of Maximum catalytic activities which are under biological control at the level of gene expression level, protein abundance and/or post-translational modification controls. The X-axis represents the gradient of PEPC strength while the Y-Axis represents the gradient of Rubisco strength. Standard simulation of *Kalanchoe daigremontiana* experiment (Vpmax = 80 and Vcmax = 25) is plotted as reference line in each subfigure (Blue dashed line). The simulation results with a different combinations of Rubisco and PEPC strengths represented in each subfigure (Black solid line). Stomatal conductance has been set to represent the *Kalanchoe daigremontiana* experimental data, in order to investigate the sensitivity of the simulation to non-stomatal parameters. The panel in row 3, column 3 showing the lines aligned completely because they are under the same conditions.

Discussion:

The goal of this chapter is to transform the system dynamics model of CAM photosynthesis into a more compact and rigorous mathematical format of Ordinary Differential Equation (ODE). Another objective of the ODE format was to minimise the number of parameters which cannot be derived from existing experimental data. The key findings of this chapter are listed below:

1. Simulation of the ODE captures classical patterns of CAM gas-exchange patterns.
2. Simulation of the ODE captures CAM responses to internal and external perturbations.
3. Parameterisation with stomatal conductance data does not fix the model to the stomatal behaviours alone. Instead, the models are shown to be responsive to key parameters including biochemical parameters (Carboxylation rate, Decarboxylation rate, Mitochondrial respiration rate and Malate transport rate) and anatomical parameters (Succulence, Vacuolar capacity and Mesophyll conductance).

This section will discuss in greater detail the properties that are conserved after the systems dynamic model was transformed into ODE format, followed by the significance and implications of parameterisation and compartmentalisation seen in the current version of the model.

Regarding the model transformation, it is worth emphasising that the apparent concentrations of any chemical entity at a specific time including that of enzymes are the results of the rate of production minus the rate of degradation at any point in time. This is the fundamental mathematical concept of the Ordinary Differential Equation (ODE) model. The time-dependent nature and the ability of the simulation output at t_{n-1} to influence the simulation output at t_n are also retained after systems dynamic model was written in the ODE format. In addition, Owen's systems dynamic model (Owen and Griffiths, 2013) has previously identified the timing of enzyme activation to be critical for the CAM carboxylation patterns. This property of the system was also transferred into the ODE model.

In terms of the model parameterisation, firstly, it is crucial to incorporate stomatal conductance data into the model to capture contrasting CAM 4 Phases of species with contrasting degree of succulence (Neales 1975; Griffiths *et al.* 2008; von Caemmerer and Griffiths 2009; Boxall *et al.* 2020). The model construction that included the 4-Phase stomatal conductance pattern differed from the method that assigned a fixed oscillatory term such as the first half of the

sinusoidal wave form to the stomatal conductance variable. The rigid oscillatory stomatal behaviour with a fixed oscillatory term may be suitable for primary model derivation but has limited application for capturing the plastic behaviour of CAM (Dodd *et al.* 2002). By having stomatal conductance as one of the 'carefully' parameterised inputs , raised a concern to the author at the time of model construction that it may lead to overfitting or making the model results being too 'rigid' (i.e. only responsive to the stomatal conductance but not responsive to other variables). However, it has been successfully shown here that the dynamic system that has been parameterised with experimental stomatal conductance is still responsive to biochemical parameters (Figure 8 and 9). These responsive behaviours indicated that the stomatal conductance (gs) parameterisation did not result in over-fitting by allocating too much weight onto stomatal conductance (gs) parameter alone. In other words, the ODE model structure allows for an empirically fine-tuning parameterisation but the output is yet still responsive to expected key biochemical parameters which included but not limited to the PEPC activity and Rubisco activity. This property is crucial but often overlooked when discussing the performance and the success of a model construction. Henceforth, it is both favourable and advisable to any field biologist to collect the stomatal conductance (gs) data when performing gas-exchange measurements to feed into further model development, as well as to parameterise the model system to be specific to their specific set of biological samples.

Secondly, the relevance of Rubisco cannot be neglected even when working under the context of CAM. In C3 species, it is well documented that Rubisco is activated during the day through the inhibitor removal by Rubisco activase followed by the carbamylation process, while the activity of the Rubisco activase is in turned under the influence of ADP/ATP ratio and redox potential (Zhang and Portis 1999; Parry *et al.* 2008). In *K. daigremontiana*, the carbamylation state increased during the morning and reached its maximum before midday (Phase III) preceding the maximum Rubisco activity during Phase IV (Maxwell *et al.* 1999). A separate study in *K. daigremontiana* also showed that protein level of the enzyme Rubisco activase reached maximum at midday (Phase III) (Griffiths *et al.* 2002). These lines of evidence support the idea that the regulation of Rubisco activity by Rubisco activase and carbamylation process is also operating in CAM species. In addition, the conserved sequence and structural prediction of Rubisco across plant phylogeny resulted in expected Rubisco sensitivity to the chemical microenvironment even in CAM species (Griffiths *et al.* 2002). It was speculated that the Rubisco regulation maybe modified by the environmental conditions that limit nocturnal PEPC activity in such a way that the Rubisco becomes active earlier in the diurnal phase following a

night when the CO_2 fixation by PEPC was restricted (Griffiths *et al*. 2002). From our ODE model, the simulation of transgenic RNAi lines with no detectable PEPC activity (*rPPC1-B*) was reflected in the requirement for higher Rubisco activity at the onset of light period, rather than the temporal shift in Rubisco activation function, to sustain the diurnal CO_2 assimilation pattern. The furthest that the ODE model can inform without stretching the evidence is that the activity of Rubisco at the onset of light period influences the CO_2 assimilation pattern throughout the diurnal phases. The ODE model and the existing empirical evidence pointed towards the same conclusion that the *precise* activation of Rubisco is critical for CAM. Hence, we encourage the plant sciences community to consider Rubisco function when studying or thinking about CAM.

Thirdly, although the parametrisation step has allowed for a more precise simulation outputs than fixing the systems of equations with an oscillatory term, there are limits to the maximum number of parameters that can be optimised simultaneously. For the parameters that can be measured, this problem can be mitigated by performing direct measurements on the species and experimental condition of interest (Wedding *et al.* 1976; von Caemmerer *et al.* 1994). However, there are still gaps in the knowledge regarding the identity and activity of malate transporters. The current understanding about malate transport in CAM points towards two putative transporters, tonoplast dicarboxylate transporter (tDT) and aluminium activated malate transporter (ALMT), however, the *in vivo* activities in CAM have not been estimated (Hafke *et al.* 2003; Emmerlich *et al.* 2003; Holtum *et al.* 2005; Meyer *et al.* 2011; Frei *et al.* 2018). This provides another venue of parameter fine-tuning once the transporter identities and their responsiveness to physiological pH have been revealed.

In terms of the model compartmentalisation, chemical reactions were grouped into modules. The term ‘module’ in this chapter refers to the fundamental unit of the model defined by the broad biological function (i.e. stomata, primary carboxylation, malate storage, decarboxylation, secondary carboxylation), thus each module incorporates more than 1 chemical reaction (Owen and Griffiths 2013). It is to be expected that different modules should be synchronized or at the very least interact in some ways to prevent problems from futile cycles (Borland *et al.* 2016). A good example of the prevention of futile cycle in CAM is at the temporal separation of nocturnal primary carboxylation and diurnal decarboxylation. However, when zooming into the detail of the carboxylation and decarboxylation processes, the complex pathway involve trafficking of organic compounds of varying length through phospholipid bilayer membranes. In other words, the futile cycling prevention is the collective effort of

various components including transporter proteins, enzymes, enzyme activators, enzyme inhibitors, enzyme degradation pathways, and many signaling molecules. Although various components are required to function in a synchronised manner, it is a trade-off whether to incorporate more components and their associated error rates into any model or to analyse a more compact model more thoroughly.

Finally, this compact five state-variable ordinary differential equation system relies on the minimal number of key CAM enzymes with the representation of one gene isoform for each reaction. This is the proof of the concept that the current understanding of CAM pathway can be represented by this textbook version even when modelling the perturbation studies. Notably, this does not contradict to the *Kalanchoe* transcriptome analysis which reported multiple isoforms of each enzymes, since the expression level of only one gene tended to dominate the system (Yang *et al.* 2017). Nevertheless, more model components can be expanded into the carbohydrate metabolic pool which are tightly linked to the regeneration of phosphoenolpyruvate (Borland *et al.* 2016). The carbohydrate metabolism module in this ODE model is not expanded into its minute detail because the objective of the model construction here was to minimise the number of parameters which cannot be constrained from existing experimental data. Here, the carbohydrate metabolism was reduced into an assumption that the PEP regeneration is not a limiting factor for the following. Despite being sufficient in this study, this aspect could be the missing component that resulted in the simulation overestimating the night-time CO_2 flux in the *rPPC1-B* line (Boxall *et al.* 2020).

Conclusions:

This study has provided a foundation for applying the systems dynamic modelling to capture experimental perturbation studies on the model CAM species, *Kalanchoe*. Thus, this leaves an open venue for future development of this simulation platform to match with a growing number of *Kalanchoe* genetic perturbation studies. The reconciliation between this compact system of differential equations with additional components representing enzymes or transporters of interest emerging from metabolic network flux balance analysis would help accelerate the process of building a more complete picture of CAM metabolic function. The chapter that follows moves on to consider the possibility of transcriptional regulations on CAM function.

Acknowledgements:

I would like to thank Professor Susanne von Caemmerer and Dr James Hartwell for sharing *Kalanchoe* gas-exchange data which has provided the basis for the comparison between model simulation outputs and the biological data.

Chapter 3

Gene regulatory network inference identifies key transcriptional regulators of CAM in *Kalanchoe fedtschenkoi*

Abstract:

The completed 24-hour timeseries transcriptomic study of *Kalanchoe fedtschenkoi* provided an opportunity to investigate the molecular building blocks of Crassulacean Acid Metabolism (CAM) and suggest evidence for molecular convergence from an evolutionary genomics perspective. This dataset offers a rich resource for large-scale computational studies including, but not limited to, Gene Regulatory Network (GRN) inferences. There is a longstanding hypothesis in the research field of Crassulacean Acid Metabolism (CAM) that the nocturnal primary carbon fixation together with the *orchestration* of the associated carbohydrate metabolic pathways have been achieved through the rewiring of the regulatory components onto the orthologues of C3 genes, rather than neofunctionalization to create CAM genes. To investigate this matter further in CAM, we developed an algorithmic pipeline for the combination of Gene Regulatory Network (GRN) inferences and Transcription Factor Binding Site (TFBS) searches. The combination of these two algorithms was the most rigorous computational approach for identifying the transcriptional regulatory network between transcription factor genes (source nodes) and the CAM-functional genes (target nodes) given this type of transcriptomic data. In this study, for the first time for a CAM system, we identified potential regulatory candidates of gene expression in an unbiased manner when processing the whole transcriptome with the size of 30,964 protein-coding genes without *a priori* assumptions. The dynGENIE3 algorithm has been developed by Huynh-Thu and colleagues to tackle complex gene regulatory network with options of performing supervised or unsupervised machine learning (Huynh-Thu and Geurts 2018). The output from the DynGENIE3 algorithm yielded a ranked list of transcriptional regulatory candidates for 118 key CAM genes including phosphoenolpyruvate carboxylase kinase (*ppck*), and the pyrophosphate dikinase regulatory protein gene (*ppdk-rp*). Both of these genes are known to demonstrate an inverted timing relative to their orthologues in C3 species, and to have a prominent role in the circadian control of carboxylation and regeneration of intermediate carbon compound, respectively. Subsequently, the dynGENIE3 outputs were passed onto Transcription Factor Binding Site (TFBS) analysis, using the well-established computational tool FIMO (Grant *et al.* 2011). Statistically significant relationships between transcriptional regulatory candidates and their respective targets were identified from the combination of the two algorithms to provide a complete list of transcriptional regulatory candidates for each target gene of interest. This study was focussed on four key subnetworks that are orchestrated to govern the function of CAM: namely carboxylation, decarboxylation, circadian and stomatal subnetworks. From this

analysis we have identified a list of 21 common transcriptional regulatory candidates from 12 transcription factor families which reach the target genes across all four subnetworks. The identities of genes of interest could be specified to the unique gene ID for both the transcription factors and their target genes. These 12 transcription factor families were AP2/ERF-ERF, BBR-BPC, bHLH, C2C2-Dof, C2C2-GATA, C2H2, GRAS, HB-KNOX, LOB, MADS-MIKC, RWP-RK, and TCP. The notable CAM genes included, but were not limited to, genes coding for PPCK, PPDK-RP and malate transporters, as well as regulatory components of the circadian function (HY5, LUX, ELF3, and ELF4) and stomatal function (OST1, SLAC1, PHOT2). In addition to providing the most complete gene regulatory network analysis of CAM, this framework is a show-case for the reciprocity between experimental and computational techniques in accelerating the identification of fundamental mechanisms underpinning classical biological questions.

Introduction:

Temporal separation between the primary carbon fixation by phosphoenolpyruvate carboxylase (PEPC) at night, and the secondary carbon fixation by ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) by day, is the unique feature of Crassulacean acid metabolism (CAM). This specialised form of photosynthesis is found in 6-7% of flowering plant families and has evolved independently at least 60 times (Silvera *et al.* 2010; Edwards 2019). So far, it has been suggested that this strict temporal activity in CAM was achieved through the rewiring of regulatory networks which already exist in the C3 pathway, hence resulting in re-scheduling of transcriptional timing of key metabolic genes (Wai *et al.* 2017; Yin *et al.* 2018; Heyduk *et al.* 2019; Chen *et al.* 2020). However, the structure of the rewired network of CAM has remained elusive. Given the availability of time-series transcriptome data, it is now possible to construct gene regulatory networks using data-driven approaches (Emmert-Streib *et al.* 2014; Mochida *et al.* 2018).

A systems approach through Gene Regulatory Network (GRN) inference can help in deciphering the CAM puzzle. Firstly, phosphoenolpyruvate carboxylase kinase (PPCK, PEPC kinase) functions directly in promoting PEPC activity at night and the *ppck* gene expression pattern is shifted from day to night during the CAM induction process. This *ppck* encodes a Ser/Thr kinase enzyme which phosphorylates PEPC and reduces its sensitivity to malate inhibition, hence allowing for its primary carboxylation activity at night (Hartwell *et al.* 1996; Hartwell *et al.* 1999; Nimmo 2003). Although the transcript abundance of *ppck* gene shows a distinct nocturnal accumulation pattern, the transcriptional regulator of *ppck* is not known to date. Therefore, the transcriptional regulatory links on *ppck* gene are of interest for the CAM community, as are the regulatory processes leading to phosphoenolpyruvate (PEP) regeneration, controlled by the pyrophosphate dikinase regulatory protein gene (*ppdk-rp*). Secondly, carbohydrate metabolism must be synchronised to the CAM carboxylation process to provide sufficient phosphoenolpyruvate as substrates throughout the night (Borland *et al.* 2016). Thirdly, the mechanisms that underlie the nocturnal stomatal opening behaviour of CAM are still unclear. On one hand, the inverted timing of CAM stomatal behaviour could be secondary responses to the internal CO_2 concentration- as drawdown at night, or increase during decarboxylation processes of CAM which result in up to 1% CO_2 within internal air spaces (Cockburn *et al.* 1979). On the other hand, stomata could be controlled directly by rewired circadian components. Finally, a circadian network could potentially controls CAM

behaviour but is itself responsive to perturbation of carboxylation processes (Borland and Griffiths, 1997; Dever *et al.* 2014; Wai and VanBuren, 2018). Therefore, unlocking the gene regulatory network of CAM is of great interest not only in terms of determining the overall network structure, but also to shed light on the regulatory links that permit the unique inverse-timing behaviour inherent to elements of the CAM cycle.

Clustering analysis of the time-series transcriptome data from CAM model species, *Kalanchoe fedtschenkoi*, provided the foundation for an initial network analysis (Yang *et al.* 2017). The clustering analysis identified 11 temporal clusters and distinguished them from genes without a fluctuation in diel expression patterns. Although clustering has advantages for showing the temporal grouping of genes, it does not directly reflect the regulatory relationship between transcription factors and their targets. Therefore, additional analyses such as gene regulatory network inference, hereafter GRN inference, are still required to decipher the transcriptional regulatory relationships. The difference between clustering analysis and GRN inference lies in the algorithm logic: co-expression clustering groups genes with shared co-expression features without directional indication, whereas GRN inference calculates the directional relationship score between any gene pair. In this study, we used the time-series gene expression data from *Kalanchoe fedtschenkoi* (Yang *et al.* 2017) to construct GRN inference networks. The specific focus was on the phosphoenolpyruvate carboxylase kinase gene (*ppck*) which has such a prominent role in regulating the nocturnal carboxylase activity of PEPC, as well as the pyrophosphate dikinase regulatory protein gene (*ppdk-rp*), which controls regeneration of the PEPC substrate, PEP. Both exhibit a marked diel expression fluctuation profile, but regulatory factors were still unknown.

The use of GRN inference is still a growing branch of computational biology. Various algorithms have shown satisfactory performances in the DREAM4/5 comparative challenges, although false positives are common to all (Greenfield *et al.* 2010; Marbach *et al.* 2012; Huynh-Thu and Geurts 2018). Therefore, it is important to utilise multiple algorithms to confirm the identity of the strong candidate regulators and provide a comprehensive list of previously unidentified CAM transcriptional regulators. Here, we used dynGENIE3, an algorithm based on random forest decision tree which has been developed specifically to tackle timeseries dataset, to obtain the overall network structure and the regulatory links for genes of interest (Huynh-Thu *et al.* 2010). Ultimately, the regulatory links from dynGENIE3 algorithms are passed through FIMO (Find Individual Motif Occurrences tool) to confirm the identity of strong potential regulators, whilst discarding false positives (Grant *et al.* 2011).

As a result of GRN inference, firstly we identified members of BES1, bHLH, TCP, AP2/ERF-ERF, C2H2 and MYB transcription factor families as strong candidate regulators on *ppck* gene (Kaladp0037s0517). Secondly, we identified members of TCP, bHLH, AP2/ERF-ERF, C2H2, and C2H2-GATA transcription factor families as strong candidate regulators on *ppdk-rp* gene (Kaladp0010s0106). Thirdly, we identified members of GARP-G2-like, MYB-related and AP2/ERF-ERF transcription factor families as strong candidate regulators on HY5 gene (Kaladp0060s0460), and members of RWP-RK and AP2/ERF-ERF transcription factor families as strong candidate regulators on LUX (Kaladp0033s0047). We also demonstrated that 21 transcriptional regulatory candidates from 12 transcription factor families are shared between all four subnetworks. Overall, this study has provided a step towards building the complete regulatory landscape of a CAM species. The benefit of applying the GRN inference approach to the *Kalanchoe fedtschenkoi* transcriptome is that this species already has transformable protocols that would allow direct functional validation through RNA interference perturbation studies and would allow the exploration of more specific controls on CAM circadian components and stomatal networks (Hartwell *et al*. 2016).

Material and methods:

Data source:

The 24-hour timeseries transcriptome data with the sampling rate of every 2 hours were collected and published by Yang *et al* in 2017 and subsequently deposited in the public databases NCBI SRA with the BioSample accession codes SAMN07453940 - SAMN07453987 (Yang *et al.*, 2017). This specific set of input data for the dynGENIE3 model construction of *K. fedtschenkoi* in this study were obtained directly from the *Kalanchoe* genome project team leader. Gene ID was obtained from the phytozome database version 13 (Goodstein *et al.* 2012).

Data pre-processing:

Timeseries transcriptome data were converted into a list of arrays. Each array represent the data collected from an experiment which can be described as a 2 dimensional matrix whereby rows are the time points and columns are the genes. Each value in the matrix represents the gene expression value for the individual gene at a given time point. Genes that contain 0 gene expression value for all time points were removed from the list of arrays.

The dynamic GENIE3:

The dynamic GENIE3 (dynGENIE3) is a model that infers gene regulatory networks from time series expression data. It works by creating a function which models the expression data of a gene in the dataset by taking the input from time series expression data of all genes using random forest method, and identifying which gene has the highest importance in predicting the expression value of a given gene. This algorithm returns the ranking of scores of all candidate regulators for each and every target genes across the whole genome.

Random Forest is a method for creating a regression function based on the principle of decision tree, a flowchart structure that has several nodes, each node is a true-or-false test splitting into true or false branches. It starts by creating a bootstrap dataset from the learning data by randomly selecting expression data of the gene in a different time point, and creating an array with identical shape to the learning data. Next, the decision tree is created from the bootstrap data. In the tree growing process, random forest allows each node to identify the best split from

the set of K genes which are selected randomly at each divide. For each gene, the bootstrapping and decision tree creation process are repeated n times. Each decision tree may predict different results from the given input, but the final result is the prediction averaged from all the trees.

Hyperparameters: nTrees = 1000, K = sqrt, Method = RF

The Random Forest method allows the importance of a variable in predicting the output to be measured, which can be used in assigning edge scores for the Gene regulatory network (GRN). In the random forest ensemble of a given gene, the reduction of variance caused by each test node in a decision tree is computed. In one decision tree, the relevance of a gene (variable) in predicting the gene expression value is the sum of all variance reduction of the nodes where the gene is used to split. This process is repeated to assign variable importance scores in all ensemble trees. The edge score of a gene is then averaged from all the trees (Huynh-Thu and Geurts 2018). The edge scores for permutations of every gene pair are ranked. After looking at the score distribution, it was decided to use the threshold of edge score $\geq 1e^{-4}$ **(**unitless). The choice of the threshold value is always subjective and reflects how many edges down the rank from the most significant score would be of interest within a particular study.

Transcription factor binding site (TFBS) search:

The promoter regions of all target genes were extracted from the complete genome sequence using Linux command (courtesy of Dr Citu and Dr Yadav from National Institute of Plant Genome Research, New Delhi, India, supplementary material). These promoter regions were defined as upstream/downstream (+/-) across the *ori* point of each gene. Subsequently, the promoter regions were passed into the FIMO (Find Individual Motif Occurrences) to check for the presence of corresponding *cis*-elements of candidate regulators for a particular gene. The presence of the *cis*-elements of candidate regulators within the promoter region of prospective target genes with a significant score of p-value $\leq 1e^{-4}$ were kept and reported as the *highly likely* candidate transcriptional regulators.

Defining subnetworks:

The entire 30,964 protein-coding genes were annotated from the original transcriptome and sequencing study (Yang *et al.* 2017). In this study, key CAM genes were selected from these 30,964 genes and grouped into 4 subnetworks according to the key metabolic processes of CAM over the 24-hour period. These four subnetworks included:

A) Carboxylation subnetwork containing 39 target genes (Table 1)
B) Decarboxylation subnetwork containing 17 target genes (Table 2)
C) Circadian subnetwork containing 33 target genes (Table 3)
D) Stomatal subnetwork containing 20 target genes (Table 4)

Different ways of re-grouping subnetworks would not affect the regulatory ranking score for any target gene because the scores were calculated across the entire genome regardless of the subnetwork in which they function. The subnetworks were defined to facilitate the data handling at a later stage such as the search for common regulators within the same functional unit of CAM.

Importantly, more subnetworks can be defined later. The gene regulatory network (GRN) construction included the entire 30,964 protein-coding genes, thus it can return the maximum number of 958,738,332 regulatory pairs. The subnetwork only comes into play when a modeller would like to extract parts of the network for further analysis or data visualisation. Hence, any other subnetworks are always to be encouraged since there is no need to repeat the heavy computing-power step of random forest construction.

Table 1: Target genes of carboxylation subnetwork

Transcript ID	Gene name
Kaladp0018s0287.1	β-CA
Kaladp0018s0289.1	β-CA
Kaladp0024s0122.1	β-CA
Kaladp0034s0051.1	β-CA
Kaladp0081s0140.1	β-CA
Kaladp0081s0143.1	β-CA
Kaladp0538s0011.1	β-CA
Kaladp0037s0517.1	PPCK1
Kaladp0050s0014.1	PPCK

Transcript ID	Gene name
Kaladp0082s0192.1	PPCK
Kaladp0604s0001.1	PPCK2
Kaladp0011s0355.1	PEPC
Kaladp0011s1355.1	PEPC
Kaladp0048s0578.1	PEPC2
Kaladp0062s0055.1	PEPC
Kaladp0095s0055.1	PEPC1
Kaladp0001s0257.1	MDH
Kaladp0022s0111.1	MDH
Kaladp0048s0189.1	MDH
Kaladp0058s0569.1	MDH
Kaladp0095s0052.1	MDH
Kaladp0095s0564.1	MDH
Kaladp0101s0211.1	MDH
Kaladp0082s0194.1	MDH
Kaladp0093s0088.1	MDH
Kaladp1038s0012.1	MDH
Kaladp0068s0169.1	MDH
Kaladp0101s0012.1	MDH
Kaladp0024s0194.1	ALMT
Kaladp0048s0850.1	ALMT
Kaladp0050s0298.1	ALMT
Kaladp0062s0038.1	ALMT6
Kaladp0073s0021.1	ALMT6
Kaladp0007s0011.1	ALMT
Kaladp0011s0027.1	ALMT
Kaladp0011s0028.1	ALMT
Kaladp0087s0091.1	ALMT
Kaladp0091s0013.1	ALMT
Kaladp0093s0149.1	ALMT

*Carboxylation subnetwork containing 39 target genes

Table 2: Target genes of decarboxylation subnetwork

Transcript ID	Gene name
Kaladp0042s0251.1	TDT
Kaladp0010s0106.1	PPDK-RP
Kaladp0060s0363.1	PPDK-RP
Kaladp0039s0092.1	PPDK
Kaladp0076s0229.1	PPDK
Kaladp0023s0088.1	PEPCK
Kaladp0040s0194.1	PEPCK
Kaladp1116s0004.1	PEPCK
Kaladp0024s0016.1	NADP-ME
Kaladp0045s0427.1	NADP-ME
Kaladp0092s0166.1	NADP-ME
Kaladp0001s0130.1	NAD-ME
Kaladp0015s0134.1	NAD-ME
Kaladp0033s0124.1	NAD-ME
Kaladp0037s0467.1	NAD-ME
Kaladp0063s0037.1	NAD-ME
Kaladp0472s0027.1	NAD-ME

*Decarboxylation subnetwork containing 17 target genes

Table 3: Target genes of circadian subnetwork

Transcript ID	Gene name
Kaladp0011s0927.1	COP1
Kaladp0071s0308.1	CRY1
Kaladp0082s0193.1	CRY2
Kaladp0039s0732.1	ELF3
Kaladp0036s0214.1	FKF1
Kaladp0034s0172.1	PHYA
Kaladp0039s0298.1	PHYB
Kaladp0496s0018.2	CCA1
Kaladp0032s0054.1	CHE
Kaladp0040s0489.1	GI
Kaladp0033s0047.1	LUX
Kaladp0057s0097.1	PIF3
Kaladp0058s0661.1	PRR3
Kaladp0032s0115.1	PRR5
Kaladp0101s0041.1	PRR7
Kaladp0032s0115.1	PRR9
Kaladp0040s0446.2	TOC1
Kaladp0809s0098.1	ZTL
Kaladp0607s0046.1	LNK1
Kaladp0099s0129.1	LNK2
Kaladp0574s0015.1	RVE1
Kaladp0055s0349.1	RVE6
Kaladp0577s0020.1	RVE8
Kaladp0016s0180.1	CKB4
Kaladp0045s0206.1	ELF4
Kaladp0089s0025.1	FIO1
Kaladp0060s0460.1	HY5
Kaladp0076s0198.1	JMJD5
Kaladp0048s0797.1	LWD1
Kaladp0056s0075.1	PRMT5
Kaladp0040s0680.1	SKIP
Kaladp0071s0383.1	STIPL1
Kaladp0040s0530.1	TEJ

***** Circadian subnetwork containing 33 target genes

Table 4: Target genes of stomatal subnetwork

Transcript ID	Gene name
Kaladp0011s0443.1	ABI1
Kaladp0048s0509.1	ABI2
Kaladp0008s0304.1	AHA2
Kaladp0055s0506.1	AKT1
Kaladp0062s0038.1	ALMT9
Kaladp0043s0196.1	BAK1
Kaladp0062s0090.1	BLUS1
Kaladp0040s0351.1	CPK23
Kaladp0042s0341.1	CPK3
Kaladp0055s0096.1	CPK6
Kaladp0073s0100.1	HT1
Kaladp0008s0789.1	KAT1
Kaladp0840s0007.1	KAT2
Kaladp0016s0289.1	OST1
Kaladp0098s0188.1	OST2/AHA1
Kaladp0071s0248.2	PHOT1
Kaladp0033s0113.1	PHOT2
Kaladp0008s0082.1	PYL9
Kaladp0091s0013.1	QUAC1/ALMT12
Kaladp0050s0214.1	SLAC1

* Stomatal subnetwork containing 20 target genes

Results:

Regulatory relationships inferred from the computational method were derived as **directional edges** with an edge score ≥ $1e^{-4}$ from **dynGENIE3 method**, which was **combined with the evidence of transcription factor binding sites (TFBS)** within the promoter region of the target gene. These regulatory relationships were grouped into the 4 biological processes related to CAM function, namely carboxylation, decarboxylation, circadian and stomatal function.

In this result section, the data will be presented in the same consecutive order for each subnetwork. Firstly, key genes within each subnetwork were identified from the literature and initially presented as a directed graph showing directional edges which were initiated from transcriptional factor families and pointed towards particular key target genes (Figure 1). Subsequently, the data were presented as diagrams to show multiple target genes that could potentially be regulated by the same transcription factor. The final set of figures highlights the most detailed information arising from this study which define the exact locations of potential transcription factor binding sites within the promoter region of key target genes.

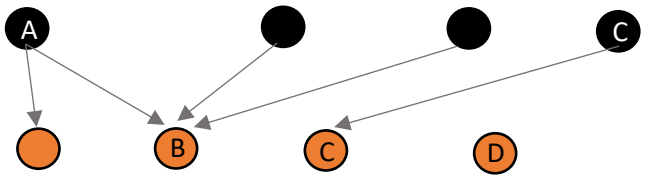


Figure 1: Example network diagram The source nodes (black) are linked to the target nodes (orange) through directional edges (grey arrows). The relationship can be one-to-many (A), many-to-one (B), one-to-one (C), or not exist (D). Multiple arrows connecting the same source node to the same target node can be seen if the node represents a group of genes (i.e. gene family) rather than a single gene with its unique gene ID.

Carboxylation subnetwork:

Carboxylation genes in *Kalanchoe fedtschenkoi* are comprised of 39 genes from 5 gene families, namely β-carbonic anhydrase (β-CA:7 genes), Phosphoenolpyruvate carboxylase kinase (PPCK: 4 genes), Phosphoenolpyruvate carboxylase (PEPC:5 genes), Malate dehydrogenase (MDH;12 genes), and Tonoplast aluminium-activated transporter (ALMT:11 genes). Among these carboxylation genes, the Phosphoenolpyruvate carboxylase kinase1 (*ppck1*) with the gene ID of Kaladp0037s0517 and Phosphoenolpyruvate carboxylase1 (*pepc1*) with the gene ID of Kaladp0095s0055 have been annotated as the functional members of their gene families according to the diel gene expression profiles and the highest mRNA and protein abundance compared to those of other gene members within same gene families (Yang *et al.* 2017; Abraham *et al.* 2020).

Our results were initially presented as directed graph showing directional edges (i.e. arrows) which started from the transcriptional factor families and pointed towards these key target genes. For the purpose of visualisation, the transcription factors and target genes were grouped into their gene families. As a result, multiple arrows could be connecting between the same source node (i.e. where the arrow starts) and the target node (i.e. where the arrow ends). The source nodes are transcription factor families and the target nodes are target genes from the carboxylation pathways and vacuolar malate transporters.

The carboxylation subnetwork contained 248 directional edges from 150 transcription factor source nodes for the 39 carboxylation gene target nodes. Figure 2 shows the relationships between transcription factor source nodes (black) and their carboxylation target nodes (blue) of the top 100 edges ranked by edge score. These 150 transcription factor nodes belonged to 33 different transcription factor families from the total number of 69 transcription factor families identified in *K. fedtschenkoi* (Zheng *et al.* 2016). The data were also analysed to show individual transcription factor genes with the highest number of different target genes within the carboxylation subnetwork (Figure 2). The maximum number of unique target genes sharing the same transcription factor was 4 target genes per 1 transcription factor gene. In this subnetwork, there were nine transcription factor genes with this characteristic: Kaladp0019s0146 (ERF family), Kaladp0028s0030 (Dof family), Kaladp0037s0181 (ERF family), Kaladp0039s0249 (ERF family), Kaladp0042s0372 (HD-ZIP family), Kaladp0053s0637 (bHLH family), Kaladp0059s0271 (GRAS family), Kaladp0076s0056 (ERF family) and Kaladp0081s0001 (C2H2 family) (Figure 3).

Analysing the regulatory relationship on the *ppck1* gene (Kaladp0037s0517) revealed 8 transcription factor binding sites (Figure 3). With the additional information from transcription factor binding site (TFBS) predictions, three members of bHLH family and one member of BES1 family showed double transcription factor binding site (TFBS) positions, whereas a member of TCP, ERF,C2H2 and MYB transcription factor family have one TFBS position in the promoter region of the *ppck1* gene (p-value $\leq 1e^{-4}$) (Figure 4).

With the focus on the *pepc1* gene, there were also 8 transcription factor genes forming regulatory links onto the *pepc1* gene (Kaladp0095s0055) (Figure 4). Firstly, two members of NAC transcription factor family (Kaladp0067s0128 and Kaladp0093s0119) had triple TFBS positions. Secondly, the third member of NAC member (Kaladp0035s0004) formed double TFBS positions within the promoter region of the *pepc1* gene. Finally, the fourth member of NAC family (Kaladp0075s0012), two members of MYB-related family (Kaladp0043s0101 and Kaladp0016s0316), a member of MADS-MIKC (Kaladp0016s0148), and a member of ERF family (Kaladp0076s0056) each had a single TFBS position within the promoter region of the *pepc1* gene (p-value $\leq 1e^{-4}$) (Figure 5).

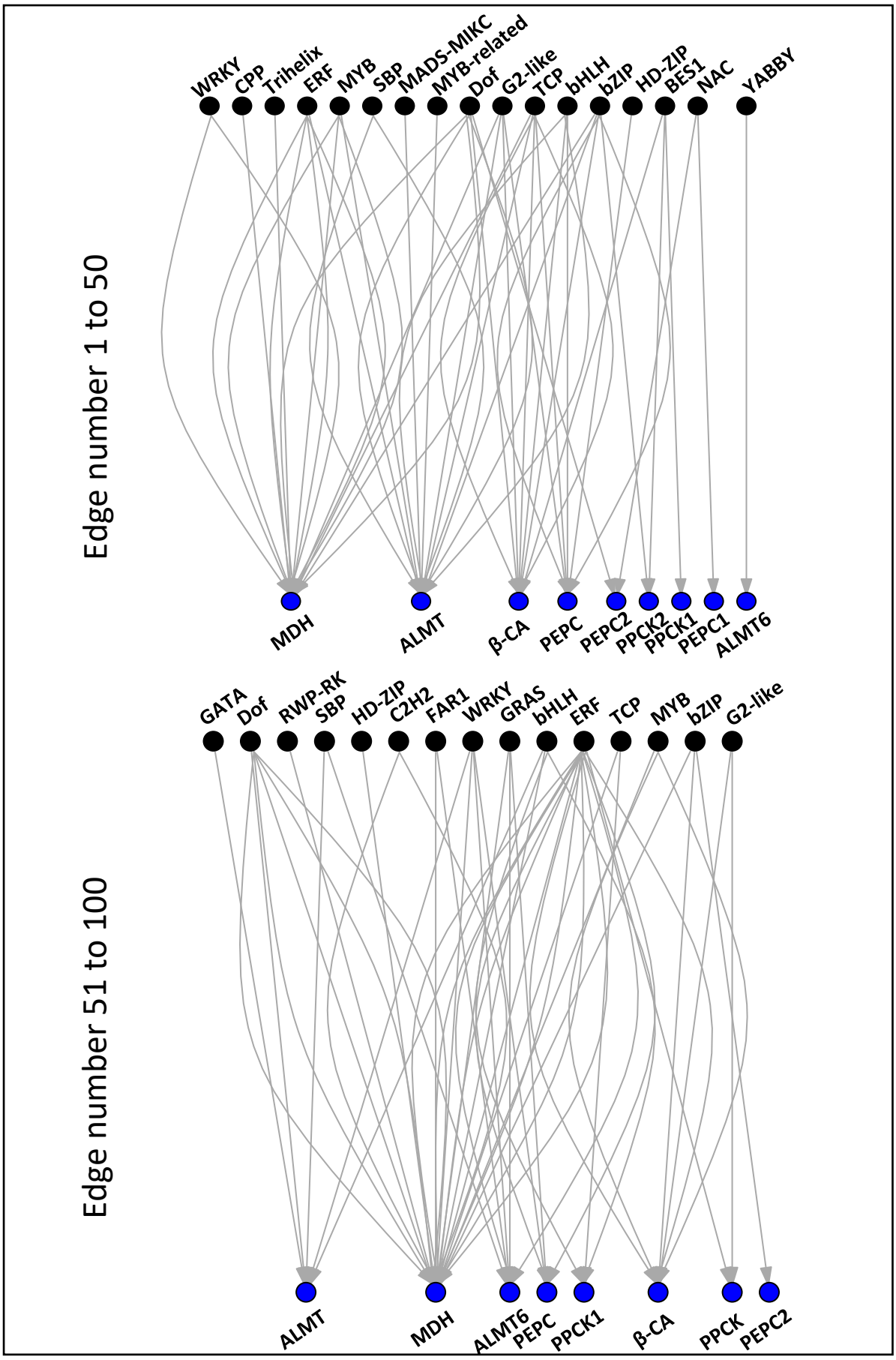


Figure 2: Carboxylation subnetwork with transcription factors as source nodes

Figure 2: Carboxylation subnetwork with transcription factors as source nodes showing the top 100 edges from transcription factor source nodes (black) to carboxylation target nodes (blue). Transcription factor nodes and carboxylation target nodes were grouped by transcription factor families and carboxylation gene families respectively (Edge score $\geq 1e^{-4}$ and p-value $\leq 1e^{-4}$).

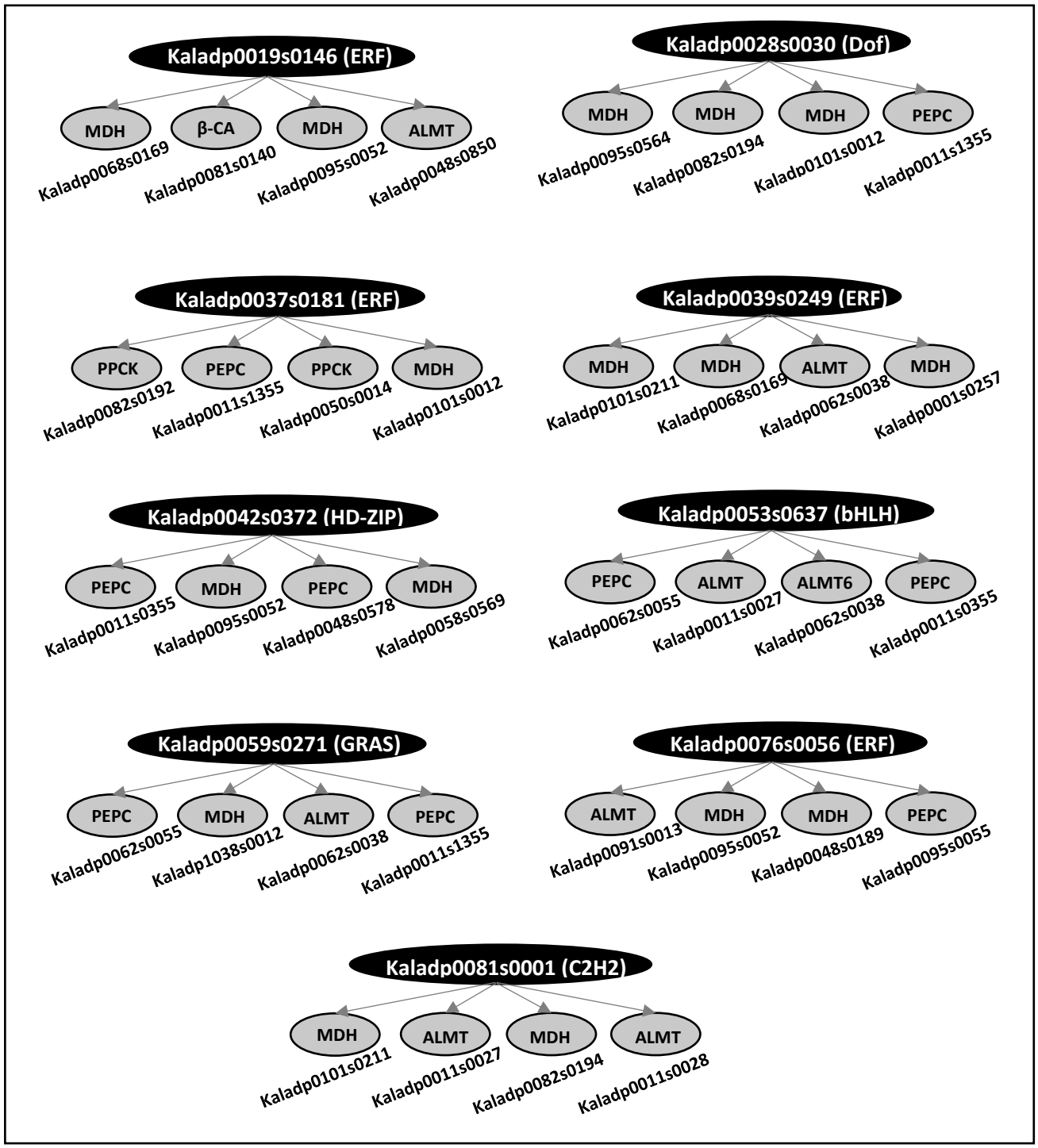


Figure 3: Multiple carboxylation genes sharing the same transcription factor source nodes Transcription factors with the maximum number of different target genes within the carboxylation subnetwork (4 target genes:1 transcription factor gene) were selected to plot in this figure. Black nodes indicate the transcription factor gene ID and family, whereas grey nodes indicate the target gene ID and gene family (Edge score $\geq 1e^{-4}$ and p-value $\leq 1e^{-4}$)

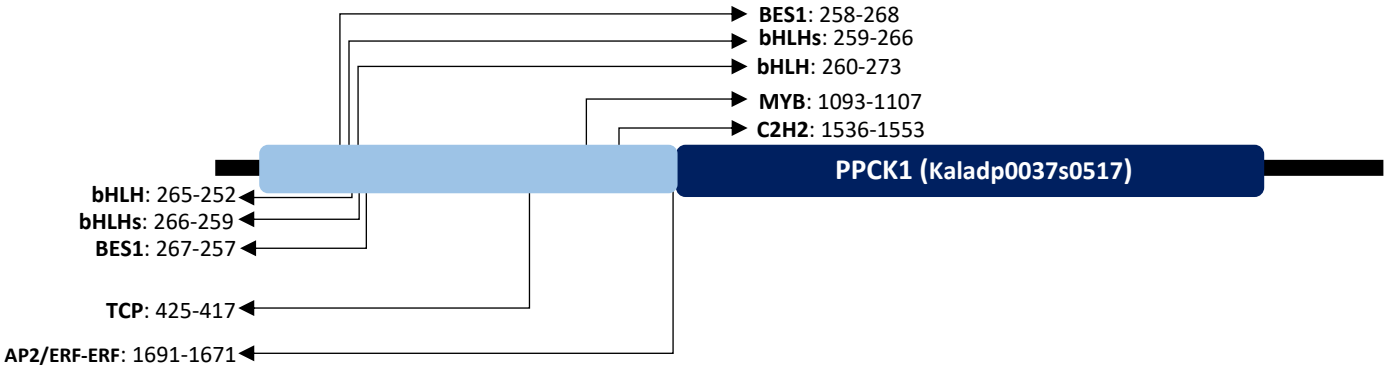


Table Regulatory links on *ppck1* gene

Transcription factor gene ID	Transcription factor family	Target gene ID	Target gene function	start	stop	strand	score	p-value
Kaladp0011s0808	BES1	Kaladp0037s0517	PPCK1	258	268	+	5.5625	5.02E-05
Kaladp0011s0808	BES1	Kaladp0037s0517	PPCK1	257	267	-	3.35938	7.87E-05
Kaladp0012s0014	bHLH	Kaladp0037s0517	PPCK1	259	266	+	14.0593	1.60E-05
Kaladp0012s0014	bHLH	Kaladp0037s0517	PPCK1	259	266	-	14.0593	1.60E-05
Kaladp0131s0006	TCP	Kaladp0037s0517	PPCK1	417	425	-	10.2316	9.19E-05
Kaladp0007s0017	bHLH	Kaladp0037s0517	PPCK1	259	266	+	14.7816	9.78E-06
Kaladp0007s0017	bHLH	Kaladp0037s0517	PPCK1	259	266	-	13.9425	1.96E-05
Kaladp0008s0733	AP2/ERF-ERF	Kaladp0037s0517	PPCK1	1671	1691	-	10.4375	2.07E-05
Kaladp0092s0233	C2H2	Kaladp0037s0517	PPCK1	1536	1553	+	8.96875	6.88E-05
Kaladp0059s0339	MYB	Kaladp0037s0517	PPCK1	1093	1107	+	13.5625	1.08E-05
Kaladp0063s0028	bHLH	Kaladp0037s0517	PPCK1	260	273	+	13.9688	1.27E-05
Kaladp0063s0028	bHLH	Kaladp0037s0517	PPCK1	252	265	-	12.0156	3.00E-05

Figure 4: Transcription factors as regulatory candidates of *ppck1* gene Edges with supporting evidence from transcription factor binding site (TFBS) prediction within the promoter regions were classified as candidate regulatory relationships. Multiple TFBS with their associated binding positions were illustrated in the diagram with their statistical scores listed in the table. The 'bHLHs' labels denotes the exact same TFBS position of two members of bHLH family (Edge score $\geq 1e^{-4}$, p-value $\leq 1e^{-4}$, Promoter region = 1,700 bp with 1,500 bp upstream and 200 bp downstream of the target gene starting position).

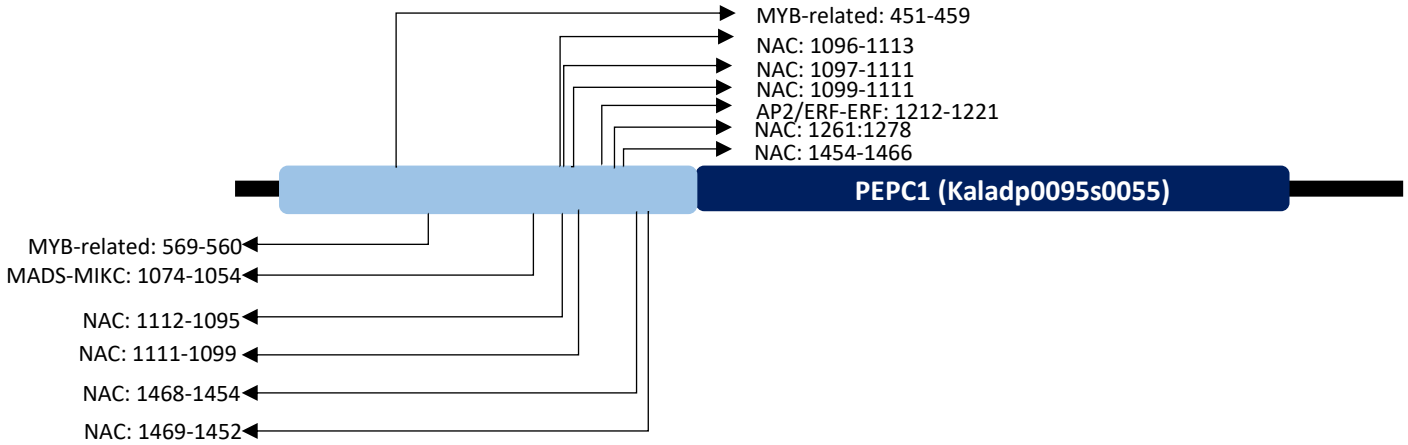


Table Regulatory links on *pepc1* gene

Transcription factor gene ID	Transcription factor family	Target gene ID	Target gene function	start	stop	strand	score	p-value
Kaladp0067s0128	NAC	Kaladp0095s0055	PEPC1	1099	1111	+	12.5469	2.96E-05
Kaladp0067s0128	NAC	Kaladp0095s0055	PEPC1	1099	1111	-	11.7812	6.20E-05
Kaladp0067s0128	NAC	Kaladp0095s0055	PEPC1	1454	1466	+	11.3438	8.81E-05
Kaladp0043s0101	MYB-related	Kaladp0095s0055	PEPC1	560	569	-	11.3368	9.41E-05
Kaladp0035s0004	NAC	Kaladp0095s0055	PEPC1	1097	1111	+	14.2031	9.15E-06
Kaladp0035s0004	NAC	Kaladp0095s0055	PEPC1	1454	1468	-	13.2812	1.69E-05
Kaladp0016s0148	MADS-MIKC	Kaladp0095s0055	PEPC1	1054	1074	-	10.75	7.67E-05
Kaladp0075s0012	NAC	Kaladp0095s0055	PEPC1	1095	1112	-	13.7188	9.91E-06
Kaladp0093s0119	NAC	Kaladp0095s0055	PEPC1	1096	1113	+	14.6562	5.51E-06
Kaladp0093s0119	NAC	Kaladp0095s0055	PEPC1	1452	1469	-	14.6094	5.70E-06
Kaladp0093s0119	NAC	Kaladp0095s0055	PEPC1	1261	1278	+	9.875	9.80E-05
Kaladp0076s0056	AP2/ERF-ERF	Kaladp0095s0055	PEPC1	1212	1221	+	11.9895	2.54E-05
Kaladp0016s0316	MYB-related	Kaladp0095s0055	PEPC1	451	459	+	12.2174	6.15E-05

Figure 5: Transcription factors as regulatory candidates of *pepc1* gene Edges with supporting evidence from transcription factor binding site (TFBS) prediction within the promoter regions were classified as candidate regulatory relationships. Multiple TFBS with their associated binding positions were illustrated in the diagram with their statistical scores listed in the table. (Edge score ≥ $1e^{-4}$, p-value ≤ $1e^{-4}$, Promoter region = 1,700 bp with 1,500 bp upstream and 200 bp downstream of the target gene starting position)

Decarboxylation subnetwork:

Genes associated with the decarboxylation subnetwork in *Kalanchoe fedtschenkoi* comprised 17 genes from 6 gene families, namely Tonoplast dicarboxylate transporter (TDT:1 gene), Pyruvate, orthophosphate dikinase (PPDK:2 genes), PPDK regulatory protein (PPDK-RP:2 genes), Phosphoenolpyruvate carboxykinase (PEPCK:3 genes), NAD-dependent malic enzyme (NAD-ME:6 genes), and NADP-malic enzyme (NADP-ME:3 genes). Based on the taxonomic distribution of carbon metabolism enzymes among CAM plants, CAM species can decarboxylate malate via two routes; NAD(P)-ME or PEPCK. Our species of interest *K. fedtschenkoi*, has been classified in the inactive PEPCK group. Among the two Pyruvate orthophosphate dikinase (PPDK) genes, a paralogous gene (gene ID of Kaladp0076s0229) has the maximum FPKM expression level twice that of other paralogues, hence this Kaladp0076s0229 gene was selected to be representative of the PPDK step. In the case of two paralogous PPDK regulatory protein (PPDK-RP) genes, the maximum expression level of both genes did not differ for more than two-times and both showed raised cyclical expression patterns which were clustered into the same co-expression module, and hence the focus was on both genes in this study (Yang *et al.* 2017).

The decarboxylation subnetwork contained 114 directional edges from 94 transcription factor source nodes associated with 17 decarboxylation gene target nodes (Figure 6). These 94 transcription factor nodes belonged to 28 different transcription factor families out of the total number of 69 transcription factor families identified in *K. fedtschenkoi* (Zheng *et al.* 2016). The maximum number of unique target genes sharing the same transcription factor was 5 target genes per 1 transcription factor gene. In this subnetwork, Kaladp0071s0407 (HB-KNOX family) was the only transcription factor gene with this characteristic. The second highest number of target genes sharing the same transcription factor gene was 4 target genes per transcription factor gene; Kaladp0016s0148 (MADS-MIKC family) and Kaladp0037s0181 (ERF family) (Figure 7).

Within this decarboxylation subnetwork, firstly, there were 8 transcription factors that formed regulatory links onto the *ppdk-rp* gene with the gene ID of Kaladp0010s0106. With the additional information from transcription factor binding site (TFBS) predictions, a member of C2H2 family showed six TFBS positions, whereas other transcription factors from TCP, ERF, C2H2 and MYB families had one or two TFBS position in the promoter region of this *ppdk-rp* gene (p-value $\leq 1e^{-4}$) (Figure 8). Secondly, with the focus on the other *ppdk-rp* gene with

the gene ID of Kaladp0060s0363, there are 6 transcription factors forming regulatory links onto this *ppdk-rp* gene. A member of C2C2-Dof, MYB-related and bZIP family showed double transcription factor binding site (TFBS) positions, whereas a member of SBP and Trihelix family showed single TFBS position within the promoter region of the *ppdk-rp* gene with the gene ID of Kaladp0060s0363 (Figure 9).

With a focus on the *ppdk* gene, there were 7 transcription factor genes forming regulatory links onto the *ppdk* gene (Kaladp0076s0229). First, a member of AP2/ERF-ERF transcription factor family (Kaladp0037s0181) had 6 TFBS positions. Second, the other member of AP2/ERF-ERF transcription factor family (Kaladp0039s0495) had triplet TFBS positions. Finally, a member from Trihelix, bHLH, MYB-related, NAC and B3-ARF transcription factor family each showed a single TFBS position within the promoter region of the *ppdk* gene (p-value $\leq 1e^{-4}$) (Figure 10).

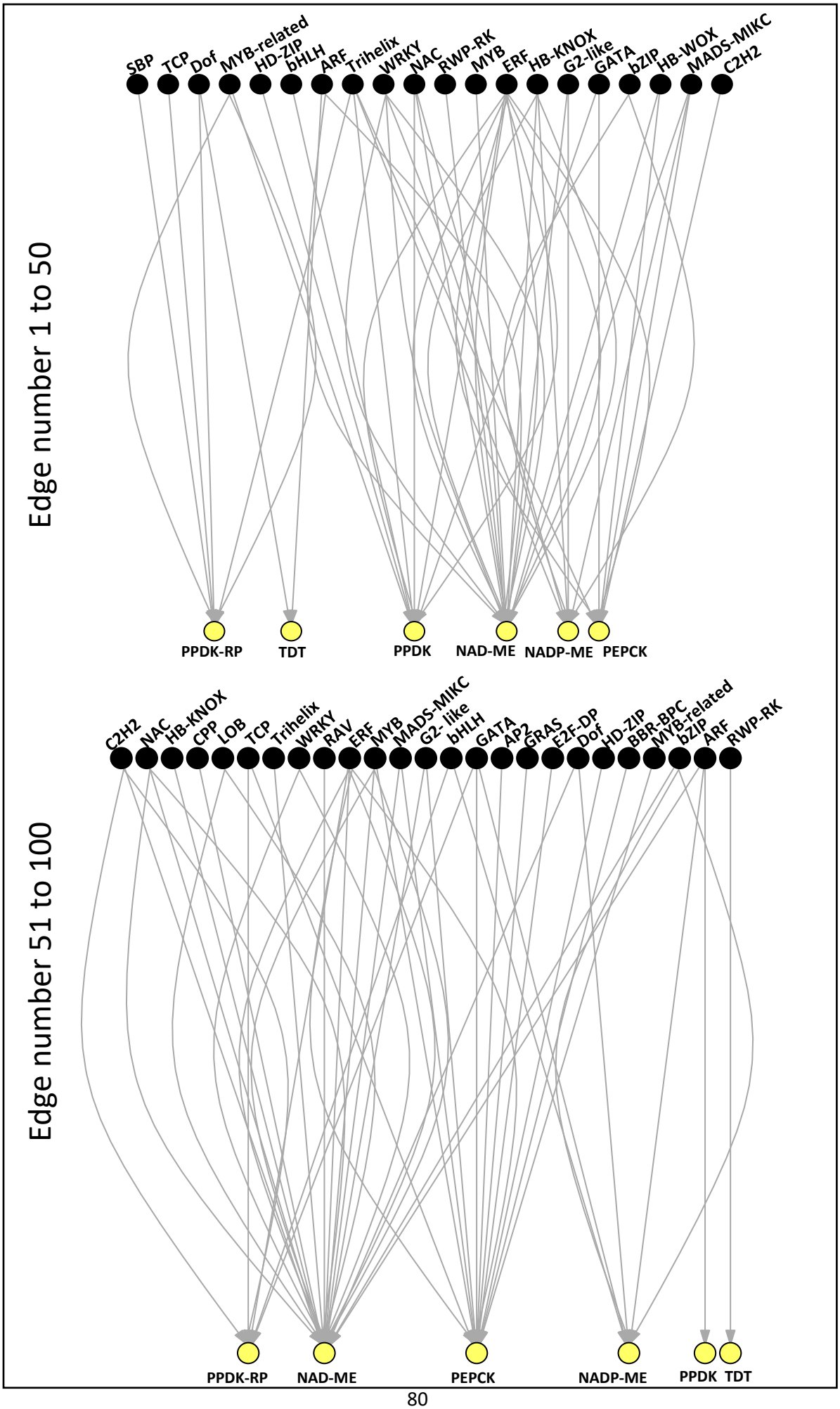

Edge number 1 to 50
SBP
TCP
Dof
MYB-related
HD-ZIP
bHLH
ARF
Trihelix
WRKY
NAC
RWP-RK
MYB
ERF
HB-KNOX
G2-like
GATA
bZIP
HB-WOX
MADS-MIKC
C2H2
PPDK-RP
TDT
PPDK
NAD-ME
NADP-ME
PEPCK
Edge number 51 to 100
C2H2
NAC
HB-KNOX
CPP
LOB
TCP
Trihelix
WRKY
RAV
ERF
MYB
MADS-MIKC
G2- like
bHLH
GATA
AP2
GRAS
E2F-DP
Dof
HD-ZIP
BBR-BPC
MYB-related
bZIP
ARF
RWP-RK
PPDK-RP
NAD-ME
PEPCK
NADP-ME
PPDK
TDT

Figure 6: Decarboxylation subnetwork with transcription factors as source nodes showing the top 100 edges from transcription factor source nodes (black) to decarboxylation target nodes (yellow). Transcription factor nodes and decarboxylation target nodes were grouped by transcription factor families and decarboxylation gene families respectively(Edge score $\geq 1e^{-4}$ and p-value $\leq 1e^{-4}$).

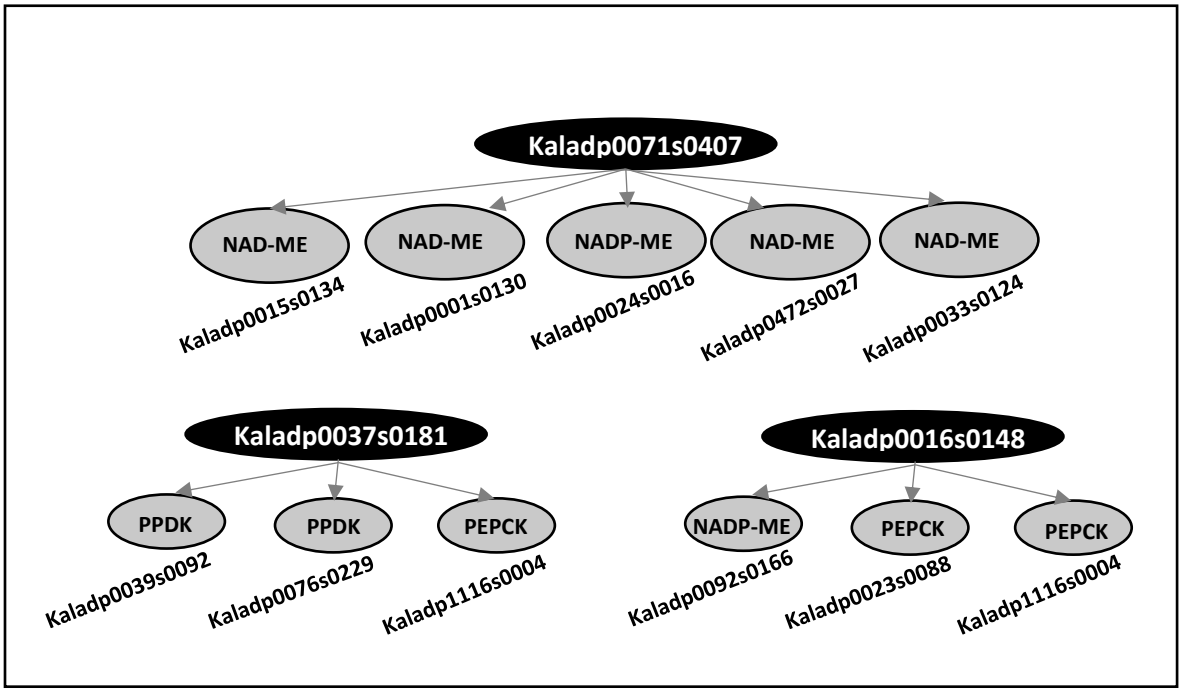
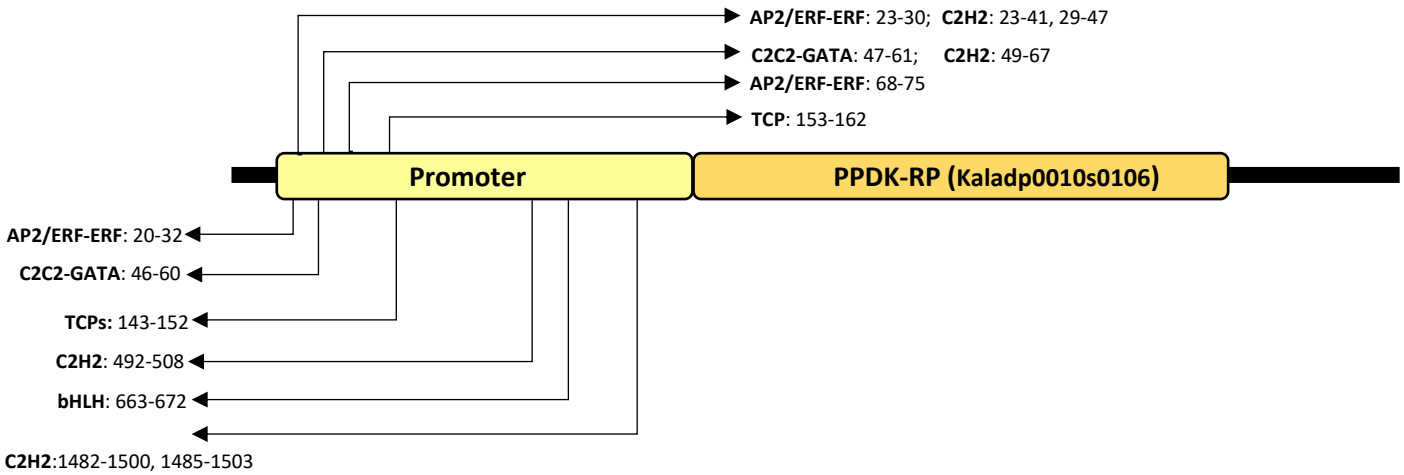
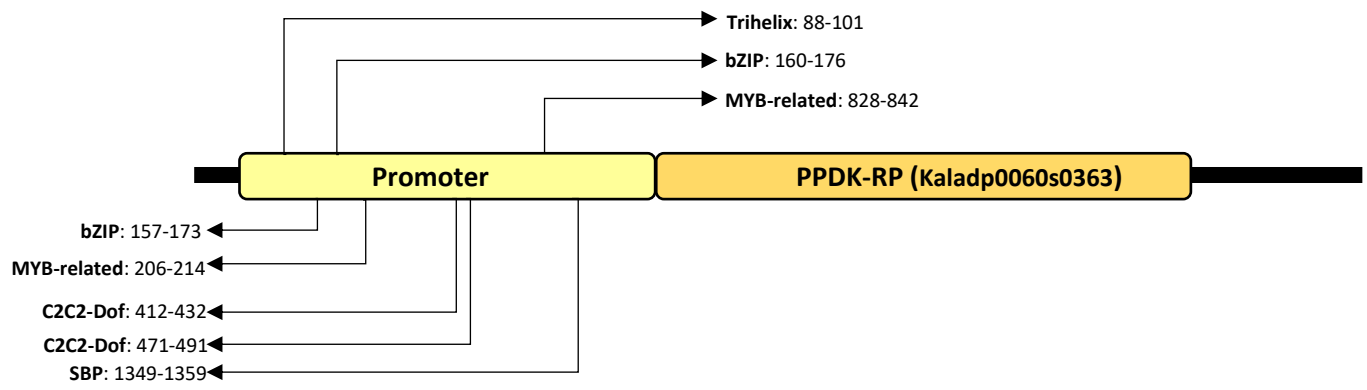


Figure 7: Multiple decarboxylation genes sharing the same transcription factor source nodes Transcription factors with the maximum number of different target genes within the decarboxylation subnetwork (5 target genes:1 transcription factor gene) and the second maximum number of different target genes within the decarboxylation subnetwork (3 target genes:1 transcription factor gene) were selected to plot in this figure. Black nodes indicate the transcription factor gene ID and family, whereas grey nodes indicate the target gene ID and gene family (Edge score $\geq 1e^{-4}$ and p-value $\leq 1e^{-4}$).



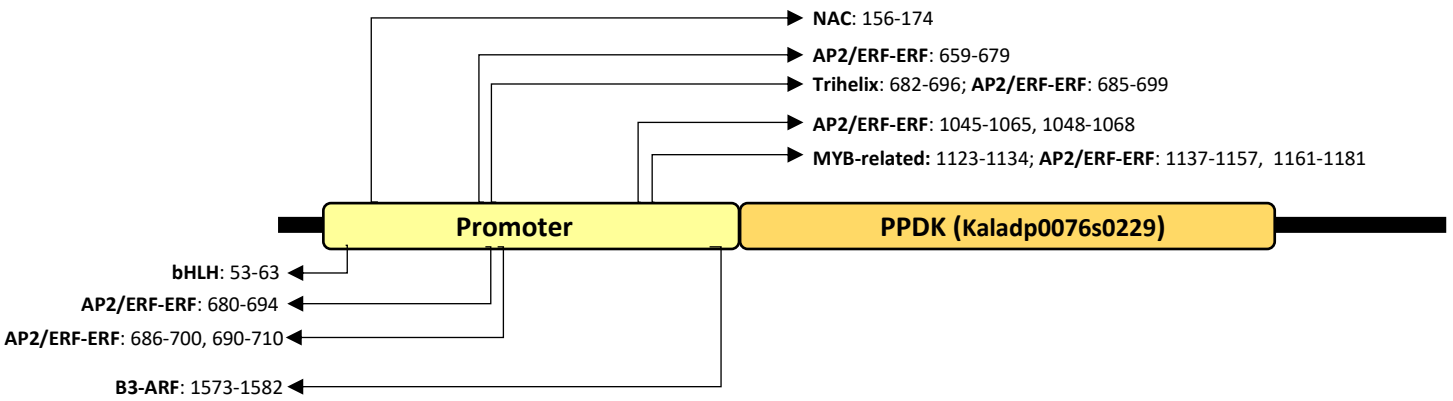
Transcription factor gene ID	Transcription factor family	Target gene ID	Target gene function	start	stop	strand	score	p-value
Kaladp0630s0020	TCP	Kaladp0010s0106	PPDK-RP	143	152	-	13.1789	1.43E-05
Kaladp0057s0097	bHLH	Kaladp0010s0106	PPDK-RP	663	672	-	10.3594	8.34E-05
Kaladp0066s0035	AP2/ERF-ERF	Kaladp0010s0106	PPDK-RP	23	30	+	13.8571	1.05E-05
Kaladp0066s0035	AP2/ERF-ERF	Kaladp0010s0106	PPDK-RP	68	75	+	11.7524	4.82E-05
Kaladp0081s0001	C2H2	Kaladp0010s0106	PPDK-RP	26	44	+	18.4917	2.44E-07
Kaladp0081s0001	C2H2	Kaladp0010s0106	PPDK-RP	23	41	+	15.125	2.14E-06
Kaladp0081s0001	C2H2	Kaladp0010s0106	PPDK-RP	1482	1500	-	8.35833	8.67E-05
Kaladp0081s0001	C2H2	Kaladp0010s0106	PPDK-RP	29	47	+	8.28333	8.99E-05
Kaladp0081s0001	C2H2	Kaladp0010s0106	PPDK-RP	49	67	+	8.25	9.14E-05
Kaladp0081s0001	C2H2	Kaladp0010s0106	PPDK-RP	1485	1503	-	8.125	9.71E-05
Kaladp0015s0060	C2C2-GATA	Kaladp0010s0106	PPDK-RP	46	60	-	13.0312	2.06E-05
Kaladp0015s0060	C2C2-GATA	Kaladp0010s0106	PPDK-RP	47	61	+	10.875	8.27E-05
Kaladp0060s0097	C2H2	Kaladp0010s0106	PPDK-RP	492	508	-	8.1875	9.20E-05
Kaladp0093s0044	TCP	Kaladp0010s0106	PPDK-RP	143	152	-	14.8421	4.55E-06
Kaladp0093s0044	TCP	Kaladp0010s0106	PPDK-RP	153	162	+	11.0105	5.30E-05
Kaladp0089s0020	AP2/ERF-ERF	Kaladp0010s0106	PPDK-RP	20	32	-	12.2812	2.60E-05

Figure 8: Transcription factors as regulatory candidates of *PPDK-RP* (Kaladp0010s0106) Edges with supporting evidence from transcription factor binding site (TFBS) prediction within the promoter regions were classified as candidate regulatory relationships. Multiple TFBS with their associated binding positions were illustrated in the diagram with their statistical scores listed in the table. (Edge score ≥ $1e^{-4}$, p-value ≤ $1e^{-4}$, Promoter region = 1,700 bp with 1,500 bp upstream and 200 bp downstream of the target gene starting position).



Transcription factor gene ID	Transcription factor family	Target gene ID	Target gene function	start	stop	strand	score	p-value
Kaladp0053s0101	C2C2-Dof	Kaladp0060s0363	PPDK-RP	412	432	-	11.375	6.25E-05
Kaladp0053s0101	C2C2-Dof	Kaladp0060s0363	PPDK-RP	471	491	-	11.0312	7.96E-05
Kaladp0055s0349	MYB-related	Kaladp0060s0363	PPDK-RP	206	214	-	13.3905	4.43E-05
Kaladp1295s0017	MYB-related	Kaladp0060s0363	PPDK-RP	828	842	+	10.7656	9.57E-05
Kaladp0016s0065	SBP	Kaladp0060s0363	PPDK-RP	1349	1359	-	10.4531	8.66E-05
Kaladp0059s0110	Trihelix	Kaladp0060s0363	PPDK-RP	88	101	+	10.6562	6.29E-05
Kaladp0076s0065	bZIP	Kaladp0060s0363	PPDK-RP	157	173	-	23.6094	1.95E-11
Kaladp0076s0065	bZIP	Kaladp0060s0363	PPDK-RP	160	176	+	17.9531	6.72E-07

Figure 9: Transcription factors as regulatory candidates of *PPDK-RP* (Kaladp0060s0363) Edges with supporting evidence from transcription factor binding site (TFBS) prediction within the promoter regions were classified as candidate regulatory relationships. Multiple TFBS with their associated binding positions were illustrated in the diagram with their statistical scores listed in the table. (Edge score $\geq 1e^{-4}$, p-value $\leq 1e^{-4}$, Promoter region = 1,700 bp with 1,500 bp upstream and 200 bp downstream of the target gene starting position).



Transcription factor gene ID	Transcription factor family	Target gene ID	Target gene function	start	stop	strand	score	p-value
Kaladp0011s0356	Trihelix	Kaladp0076s0229	PPDK	682	696	+	15.5938	3.10E-06
Kaladp0747s0001	bHLH	Kaladp0076s0229	PPDK	53	63	-	12.7703	3.48E-05
Kaladp0577s0020	MYB-related	Kaladp0076s0229	PPDK	1123	1134	+	11.0312	9.57E-05
Kaladp0037s0181	AP2/ERF-ERF	Kaladp0076s0229	PPDK	1048	1068	+	15.5942	1.46E-06
Kaladp0037s0181	AP2/ERF-ERF	Kaladp0076s0229	PPDK	1161	1181	+	9.26087	3.45E-05
Kaladp0037s0181	AP2/ERF-ERF	Kaladp0076s0229	PPDK	1137	1157	+	9.17391	3.58E-05
Kaladp0037s0181	AP2/ERF-ERF	Kaladp0076s0229	PPDK	690	710	-	7.86957	6.24E-05
Kaladp0037s0181	AP2/ERF-ERF	Kaladp0076s0229	PPDK	659	679	+	7.3913	7.60E-05
Kaladp0037s0181	AP2/ERF-ERF	Kaladp0076s0229	PPDK	1045	1065	+	6.98551	8.96E-05
Kaladp0039s0495	AP2/ERF-ERF	Kaladp0076s0229	PPDK	685	699	+	16.3438	1.37E-06
Kaladp0039s0495	AP2/ERF-ERF	Kaladp0076s0229	PPDK	686	700	-	12.0156	1.36E-05
Kaladp0039s0495	AP2/ERF-ERF	Kaladp0076s0229	PPDK	680	694	-	7.45312	9.45E-05
Kaladp0056s0140	NAC	Kaladp0076s0229	PPDK	156	174	+	11.7812	3.59E-05
Kaladp0100s0015	B3-ARF	Kaladp0076s0229	PPDK	1573	1582	-	12.875	2.39E-05

Figure 10: Transcription factors as regulatory candidates of *PPDK* Edges with supporting evidence from transcription factor binding site (TFBS) prediction within the promoter regions were classified as candidate regulatory relationships. Multiple TFBS with their associated binding positions were illustrated in the diagram with their statistical scores listed in the table. (Edge score ≥ $1e^{-4}$, p-value ≤ $1e^{-4}$, Promoter region = 1,700 bp with 1,500 bp upstream and 200 bp downstream of the target gene starting position).

Circadian subnetwork:

Core circadian genes of *Kalanchoe fedtschenkoi* have been identified from the set of orthologous circadian genes in *Arabidopsis thaliana* (Moseley et al., 2018). For the purpose of this study, the circadian genes of *K. fedtschenkoi* were defined according to the study by Yang *et al*, rather than a full functional validation of candidate circadian genes. Therefore, this set of circadian genes could include non-functional circadian gene orthologues and/or miss out some other CAM-specific circadian genes, if any exist. Based on this definition, the circadian genes comprised of 32 genes from 33 gene families, namely Constitutive photomorphogenic 1 (COP1), Cryptochrome 1 (CRY1), Cryptochrome 2 (CRY2), Early flowering 3 (ELF3), Flavin-binding, Kelch repeat, F box 1 (FKF1), Phytochrome A (PHYA), Phytochrome B (PHYB), Circadian clock associated 1 (CCA1), CCA1 hiking expedition (CHE), Gigantea (GI), Lux arrhythmo (LUX), Phytochrome interactive factor 3 (PIF3), Pseudo-response regulator 3 (PRR3), Pseudo-response regulator 5 (PRR5), Pseudo-response regulator 7 (PRR7), Pseudo-response regulator 9 (PRR9), Timing of Cab expression 1 (TOC1), Zeitlupe (ZTL), Night light-inducible and clock regulated 1 (LNK1), Night light-inducible and clock regulated 2 (LNK2), Reveille 1 (RVE1), Reveille 6 (RVE6), Reveille 8 (RVE8), Casein kinase II beta subunit 4 (CKB4), Early flowering 4 (ELF4), Fiona 1 (FIO1), Elongated hypocotyl 5 (HY5), Light insensitive period 1 (JMJD5), Light-regulated WD 1 (LWD1), Protein arginine methyltransferase 5 (PRMT5), SNW/SKI-interacting protein (SKIP), Spliceosomal timekeeper locus 1 (STIPL1), and Poly(ADP-ribose) glycohydrolase 1 (TEJ). Among these circadian genes, the PHYA, PHYB, PIF3 and HY5 have been shown to play critical roles in C4 photosynthesis, whilst ELF3,ELF4 and LUX displayed a concerted phase shift of 4 hr ahead of their *Arabidopsis* orthologues and pass 3 strong selection criteria for circadian expression pattern characteristics, namely: amplitude, period and oscillation (Ezer *et al.* 2017; Yang *et al.* 2017).

The circadian subnetwork contained 210 directional edges from 138 transcription factor source nodes for the 30 circadian gene target nodes (Figure 11). These 138 transcription factor nodes belonged to 33 different transcription factor families out of the total number of 69 transcription factor families identified in *K. fedtschenkoi* (Zheng *et al.* 2016). In this subnetwork, the maximum number of target genes sharing the same transcription factor was 6 targets per 1 transcription factor. The transcription factor Kaladp0022s0176 (CPP family) formed regulatory links to 6 target genes, including PIF3. The second highest number of target genes per one

transcription factor is 5 targets per 1 transcription factor, namely Kaladp0037s0181 (AP2/ERF-ERF family) with HY5 as one of the five target genes (Figure 12).

With the focus on the transcription factor binding site (TFBS) data, there were 13 transcription factors forming regulatory links onto the *PHYA* gene, 7 transcription factors forming regulatory links onto the *PHYB* gene, 11 transcription factors forming regulatory links onto the *PIF3* gene, 4 transcription factors forming regulatory links onto the *HY5* gene, 5 transcription factors forming regulatory links onto the *ELF3*, 8 transcription factors forming regulatory links onto the *ELF4*, and 2 transcription factors forming regulatory links onto the *LUX* (Figure 13-19).

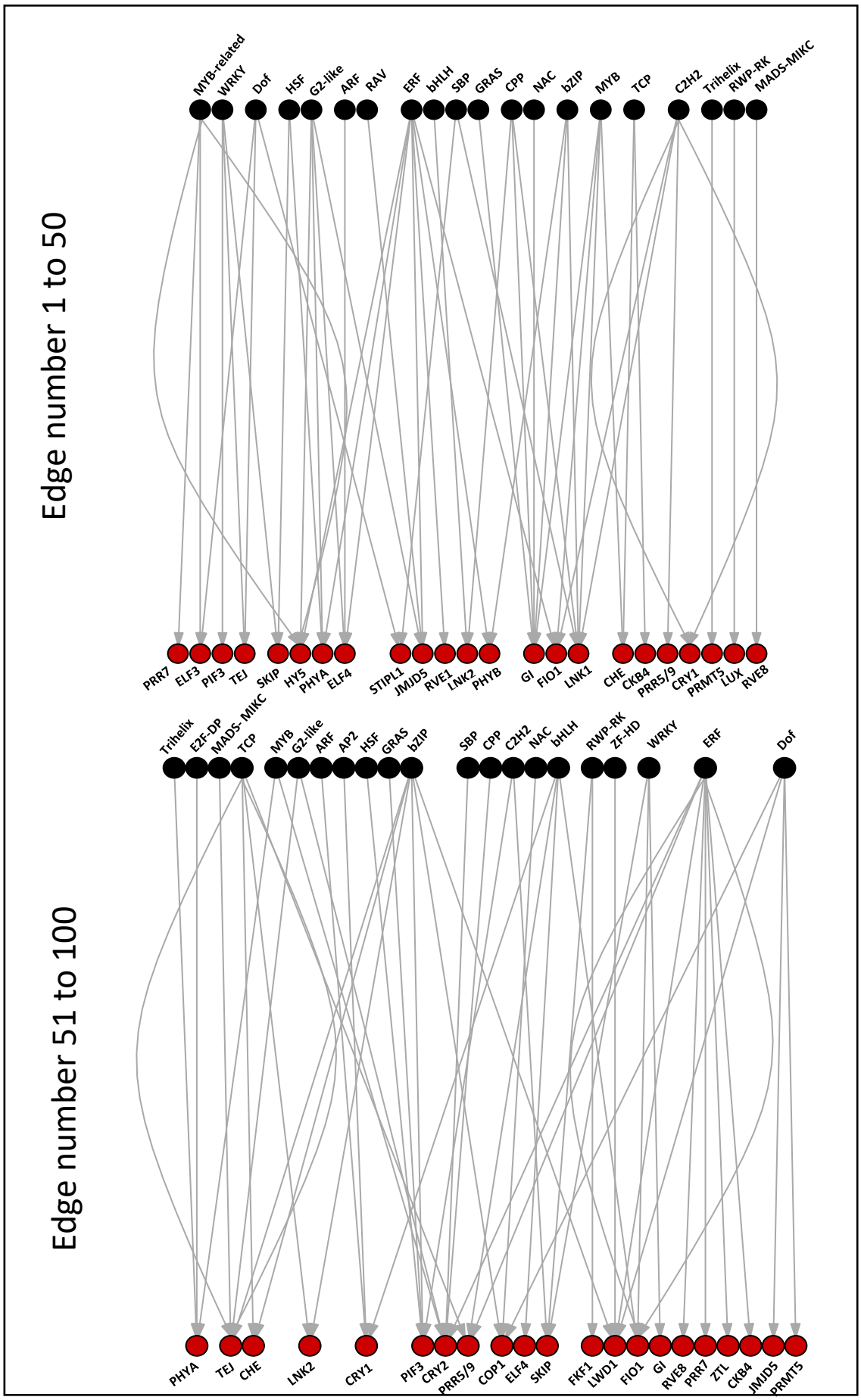

Edge number 1 to 50
MYB-related
WRKY
Dof
HSF
G2-like
ARF
RAV
ERF
bHLH
SBP
GRAS
CPP
NAC
bZIP
MYB
TCP
C2H2
Trihelix
RWP-RK
MADS-MIKC
PRR7
ELF3
PIF3
TEJ
SKIP
HY5
PHYA
ELF4
STIPL1
JMJD5
RVE1
LNK2
PHYB
GI
FIO1
LNK1
CHE
CKB4
PRR5/9
CRY1
PRMT5
LUX
RVE8
Edge number 51 to 100
Trihelix
E2F-DP
MADS- MIKC
TCP
MYB
G2-like
ARF
AP2
HSF
GRAS
bZIP
SBP
CPP
C2H2
NAC
bHLH
RWP-RK
ZF-HD
WRKY
ERF
Dof
PHYA
TEJ
CHE
LNK2
CRY1
PIF3
CRY2
PRR5/9
COP1
ELF4
SKIP
FKF1
LWD1
FIO1
GI
RVE8
PRR7
ZTL
CKB4
JMJD5
PRMT5

Figure 11: Circadian subnetwork with transcription factors as source nodes showing the top 100 edges from transcription factor source nodes (black) to circadian target nodes (red). Transcription factor nodes and circadian target nodes were grouped by transcription factor families and circadian gene families respectively (Edge score $\geq 1e^{-4}$ and p-value $\leq 1e^{-4}$).

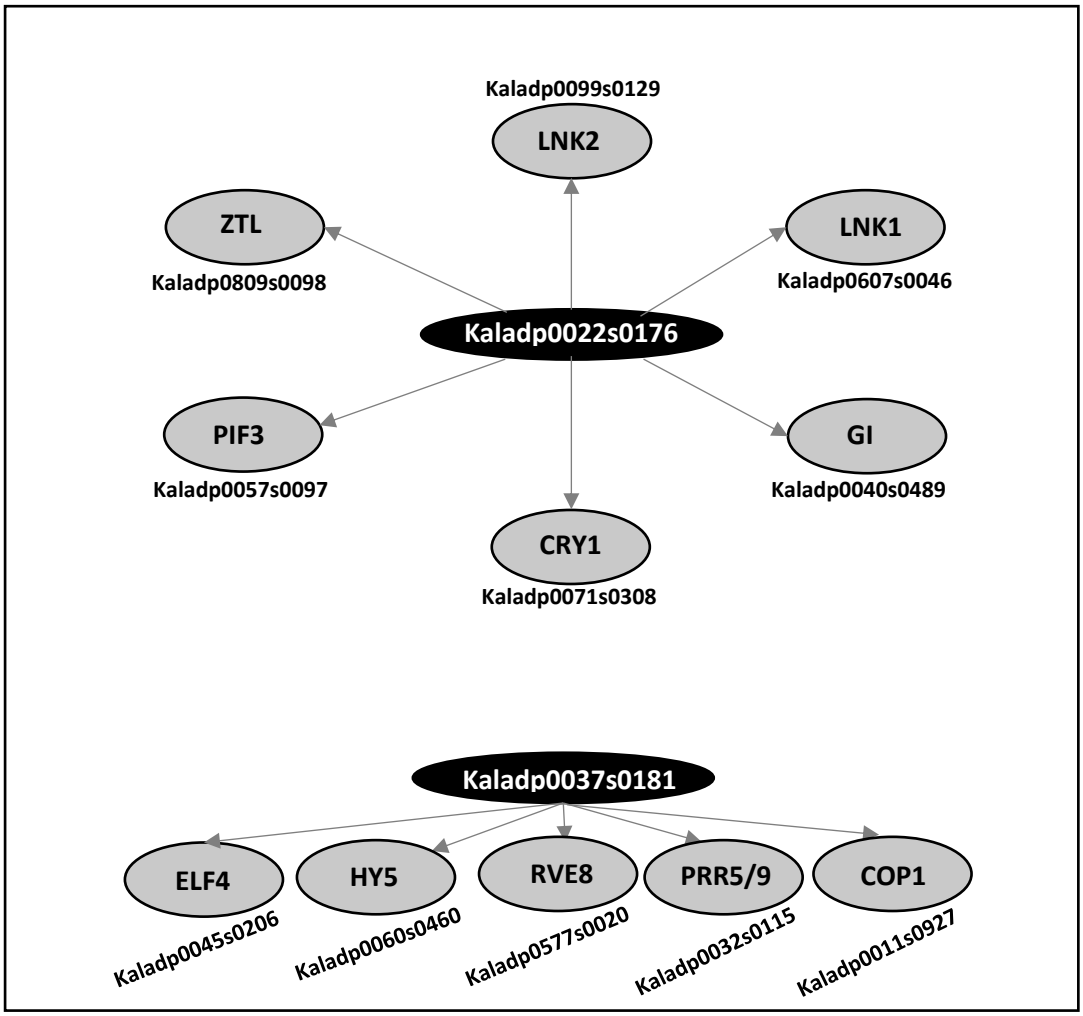
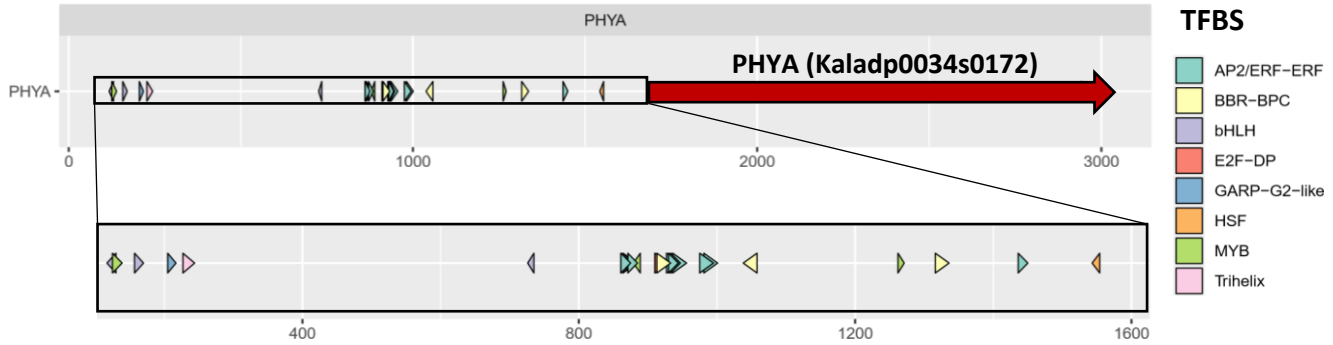


Figure 12: Multiple circadian genes sharing the same transcription factor source nodes Transcription factors with the maximum number of different target genes within the circadian subnetwork (6 target genes:1 transcription factor gene) and the second maximum number of different target genes within the circadian subnetwork (5 target genes:1 transcription factor gene) were selected to plot in this figure. Black nodes indicate the transcription factor gene ID and family, whereas grey nodes indicate the target gene ID and gene family(Edge score $\geq 1e^{-4}$ and p-value $\leq 1e^{-4}$).

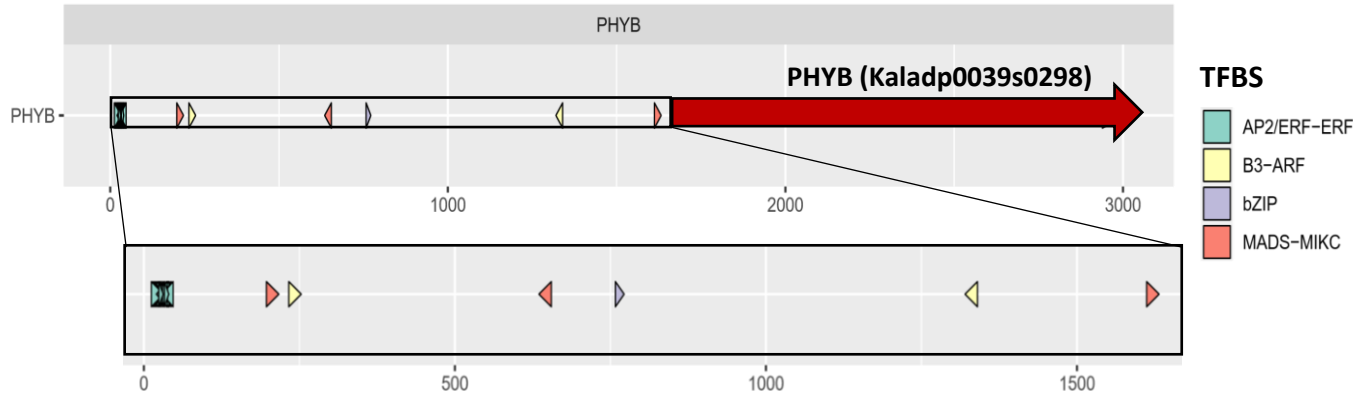


Transcription factor gene ID	Transcription factor family	Target gene ID	Target gene function	start	stop	strand	score	p-value
Kaladp0032s0071	HSF	Kaladp0034s0172	PHYA	1543	1554	-	11.9219	3.09E-05
Kaladp0019s0146	AP2/ERF-ERF	Kaladp0034s0172	PHYA	933	953	+	16.5652	7.19E-07
Kaladp0019s0146	AP2/ERF-ERF	Kaladp0034s0172	PHYA	930	950	+	16.4783	7.54E-07
Kaladp0019s0146	AP2/ERF-ERF	Kaladp0034s0172	PHYA	927	947	+	13.6522	3.30E-06
Kaladp0019s0146	AP2/ERF-ERF	Kaladp0034s0172	PHYA	864	884	+	13.5072	3.55E-06
Kaladp0019s0146	AP2/ERF-ERF	Kaladp0034s0172	PHYA	981	1001	+	11.0145	1.13E-05
Kaladp0019s0146	AP2/ERF-ERF	Kaladp0034s0172	PHYA	861	881	+	7.63768	4.57E-05
Kaladp0019s0146	AP2/ERF-ERF	Kaladp0034s0172	PHYA	975	995	+	7.57971	4.68E-05
Kaladp0019s0146	AP2/ERF-ERF	Kaladp0034s0172	PHYA	936	956	+	5.66667	9.48E-05
Kaladp0073s0079	GARP-G2-like	Kaladp0034s0172	PHYA	205	217	+	10.7344	8.26E-05
Kaladp0674s0108	E2F-DP	Kaladp0034s0172	PHYA	910	930	+	3.34783	2.83E-05
Kaladp0008s0226	Trihelix	Kaladp0034s0172	PHYA	227	244	+	11.4058	6.27E-05
Kaladp0008s0352	MYB	Kaladp0034s0172	PHYA	1262	1271	+	11.8105	4.72E-05
Kaladp0046s0332	MYB	Kaladp0034s0172	PHYA	879	889	-	15.9875	2.16E-06
Kaladp0057s0097	bHLH	Kaladp0034s0172	PHYA	726	735	-	11.4531	5.70E-05
Kaladp0039s0495	AP2/ERF-ERF	Kaladp0034s0172	PHYA	933	947	+	13.7812	5.69E-06
Kaladp0039s0495	AP2/ERF-ERF	Kaladp0034s0172	PHYA	930	944	+	12.625	1.02E-05

Figure 13: Transcription factors as regulatory candidates of *PHYA* (continue next page)

Transcription factor gene ID	Transcription factor family	Target gene ID	Target gene function	start	stop	strand	score	p-value
Kaladp0039s0495	AP2/ERF-ERF	Kaladp0034s0172	PHYA	864	878	+	9.71875	3.81E-05
Kaladp0039s0495	AP2/ERF-ERF	Kaladp0034s0172	PHYA	861	875	+	8.75	5.68E-05
Kaladp0039s0495	AP2/ERF-ERF	Kaladp0034s0172	PHYA	871	885	+	7.96875	7.74E-05
Kaladp0039s0495	AP2/ERF-ERF	Kaladp0034s0172	PHYA	1436	1450	+	7.875	8.03E-05
Kaladp0001s0060	BBR-BPC	Kaladp0034s0172	PHYA	1316	1336	+	-0.68919	4.77E-05
Kaladp0001s0060	BBR-BPC	Kaladp0034s0172	PHYA	913	933	+	-1.24324	5.58E-05
Kaladp0001s0060	BBR-BPC	Kaladp0034s0172	PHYA	1038	1058	-	-1.94595	6.79E-05
Kaladp0045s0236	bHLH	Kaladp0034s0172	PHYA	157	170	+	11.3125	7.22E-05
Kaladp0045s0236	bHLH	Kaladp0034s0172	PHYA	117	130	-	9.65625	9.62E-05
Kaladp0016s0134	MYB	Kaladp0034s0172	PHYA	125	139	+	11.4219	5.27E-05
Kaladp0087s0163	AP2/ERF-ERF	Kaladp0034s0172	PHYA	861	875	+	9.59375	7.96E-05

Figure 13: Transcription factors as regulatory candidates of *PHYA* Edges with supporting evidence from transcription factor binding site (TFBS) prediction within the promoter regions were classified as candidate regulatory relationships. Multiple TFBS with their associated binding positions were illustrated in the diagram with their statistical scores listed in the table (Edge score $\geq 1e^{-4}$, p-value $\leq 1e^{-4}$, Promoter region = 1,700 bp with 1,500 bp upstream and 200 bp downstream of the target gene starting position).

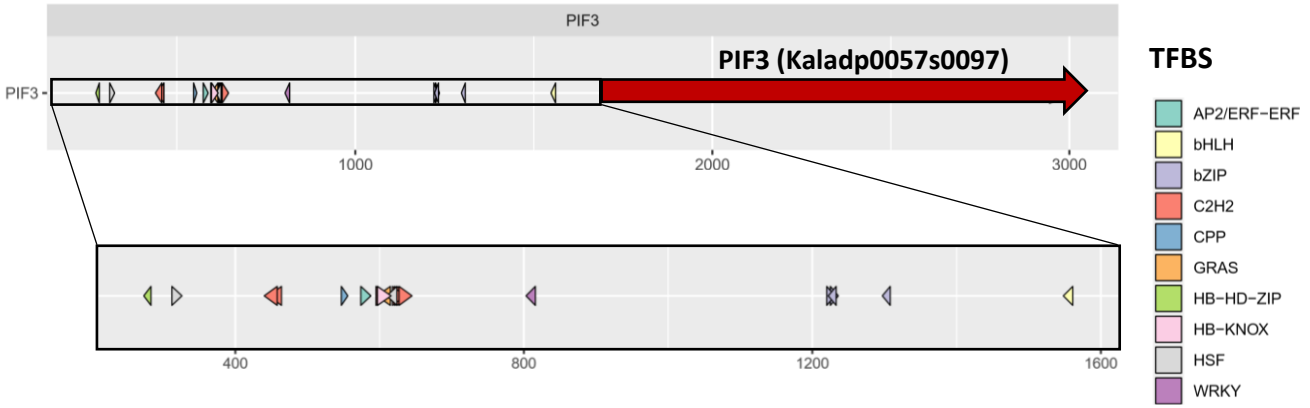


Transcription factor gene ID	Transcription factor family	Target gene ID	Target gene function	start	stop	strand	score	p-value
Kaladp0921s0014	bZIP	Kaladp0039s0298	PHYB	758	772	+	9.6875	4.81E-05
Kaladp0066s0035	AP2/ERF-ERF	Kaladp0039s0298	PHYB	27	34	+	15.8	2.25E-06
Kaladp0066s0035	AP2/ERF-ERF	Kaladp0039s0298	PHYB	30	37	+	15.8	2.25E-06
Kaladp0066s0035	AP2/ERF-ERF	Kaladp0039s0298	PHYB	33	40	+	11.7905	4.42E-05
Kaladp0066s0035	AP2/ERF-ERF	Kaladp0039s0298	PHYB	24	31	+	11.2667	6.24E-05
Kaladp0066s0035	AP2/ERF-ERF	Kaladp0039s0298	PHYB	26	33	-	10.5143	8.58E-05
Kaladp0066s0035	AP2/ERF-ERF	Kaladp0039s0298	PHYB	29	36	-	10.5143	8.58E-05
Kaladp0066s0035	AP2/ERF-ERF	Kaladp0039s0298	PHYB	32	39	-	10.5143	8.58E-05
Kaladp0048s0420	AP2/ERF-ERF	Kaladp0039s0298	PHYB	18	38	+	27.1562	1.10E-10
Kaladp0048s0420	AP2/ERF-ERF	Kaladp0039s0298	PHYB	21	41	+	26.3125	2.84E-10
Kaladp0048s0420	AP2/ERF-ERF	Kaladp0039s0298	PHYB	15	35	+	23.875	3.36E-09
Kaladp0048s0420	AP2/ERF-ERF	Kaladp0039s0298	PHYB	12	32	+	18.1875	3.45E-07
Kaladp0048s0420	AP2/ERF-ERF	Kaladp0039s0298	PHYB	24	44	+	14.2344	4.42E-06
Kaladp0081s0316	AP2/ERF-ERF	Kaladp0039s0298	PHYB	26	44	-	25.0156	1.18E-09
Kaladp0081s0316	AP2/ERF-ERF	Kaladp0039s0298	PHYB	23	41	-	24.25	2.60E-09
Kaladp0081s0316	AP2/ERF-ERF	Kaladp0039s0298	PHYB	29	47	-	17.875	4.73E-07

Figure 14: Transcription factors as regulatory candidates of *PHYB* (continue next page)

Transcription factor gene ID	Transcription factor family	Target gene ID	Target gene function	start	stop	strand	score	p-value
Kaladp0081s0316	AP2/ERF-ERF	Kaladp0039s0298	PHYB	20	38	-	11.9844	1.39E-05
Kaladp0101s0229	AP2/ERF-ERF	Kaladp0039s0298	PHYB	27	34	+	17.7158	1.13E-06
Kaladp0101s0229	AP2/ERF-ERF	Kaladp0039s0298	PHYB	30	37	+	17.7158	1.13E-06
Kaladp0101s0229	AP2/ERF-ERF	Kaladp0039s0298	PHYB	33	40	+	13.8632	1.36E-05
Kaladp0101s0229	AP2/ERF-ERF	Kaladp0039s0298	PHYB	24	31	+	12.1263	2.98E-05
Kaladp0008s0410	B3-ARF	Kaladp0039s0298	PHYB	233	253	+	7.4375	3.38E-06
Kaladp0008s0410	B3-ARF	Kaladp0039s0298	PHYB	1320	1340	-	-0.9125	4.23E-05
Kaladp0016s0148	MADS-MIKC	Kaladp0039s0298	PHYB	1612	1632	+	13.7353	1.17E-05
Kaladp0016s0148	MADS-MIKC	Kaladp0039s0298	PHYB	197	217	+	11.8676	4.06E-05
Kaladp0016s0148	MADS-MIKC	Kaladp0039s0298	PHYB	635	655	-	11.1765	6.07E-05

Figure 14: Transcription factors as regulatory candidates of *PHYB* Edges with supporting evidence from transcription factor binding site (TFBS) prediction within the promoter regions were classified as candidate regulatory relationships. Multiple TFBS with their associated binding positions were illustrated in the diagram with their statistical scores listed in the table (Edge score $\geq 1e^{-4}$, p-value $\leq 1e^{-4}$, Promoter region = 1,700 bp with 1,500 bp upstream and 200 bp downstream of the target gene starting position).

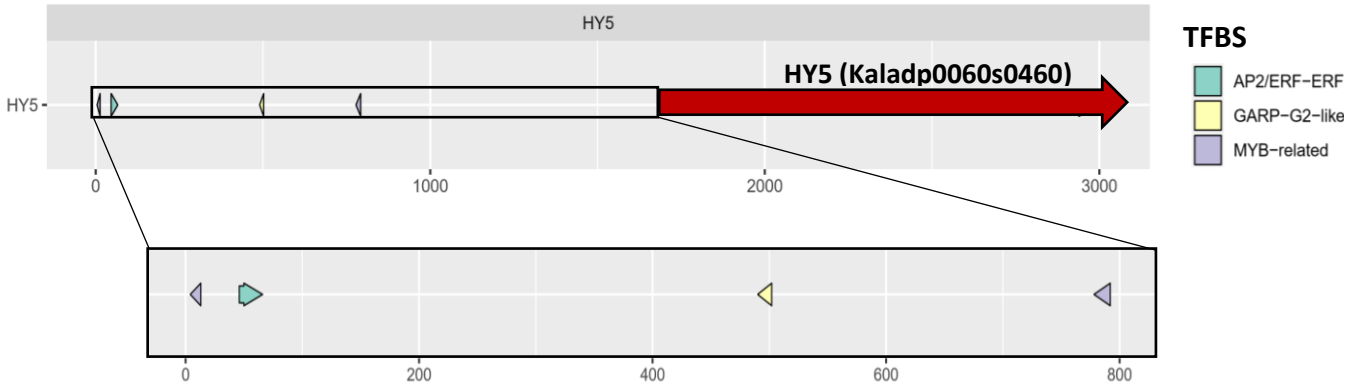


Transcription factor gene ID	Transcription factor family	Target gene ID	Target gene function	start	stop	strand	score	p-value
Kaladp0053s0301	WRKY	Kaladp0057s0097	PIF3	804	816	-	7.85938	9.20E-05
Kaladp0059s0271	GRAS	Kaladp0057s0097	PIF3	596	615	-	10.9714	6.53E-05
Kaladp0050s0021	bZIP	Kaladp0057s0097	PIF3	1225	1236	+	11.7263	5.36E-05
Kaladp0050s0021	bZIP	Kaladp0057s0097	PIF3	1220	1231	+	11.6105	5.80E-05
Kaladp0050s0021	bZIP	Kaladp0057s0097	PIF3	1222	1233	-	11.6105	5.80E-05
Kaladp0050s0021	bZIP	Kaladp0057s0097	PIF3	1297	1308	-	11.3158	6.96E-05
Kaladp0081s0001	C2H2	Kaladp0057s0097	PIF3	621	639	+	18.3833	2.63E-07
Kaladp0081s0001	C2H2	Kaladp0057s0097	PIF3	446	464	-	14.1917	3.75E-06
Kaladp0081s0001	C2H2	Kaladp0057s0097	PIF3	624	642	+	14.0833	3.99E-06
Kaladp0081s0001	C2H2	Kaladp0057s0097	PIF3	618	636	+	12.075	1.26E-05
Kaladp0081s0001	C2H2	Kaladp0057s0097	PIF3	440	458	-	9.70833	4.41E-05
Kaladp0081s0001	C2H2	Kaladp0057s0097	PIF3	627	645	+	9.61667	4.62E-05

Figure 15: Transcription factors as regulatory candidates of *PIF3* (continue next page)

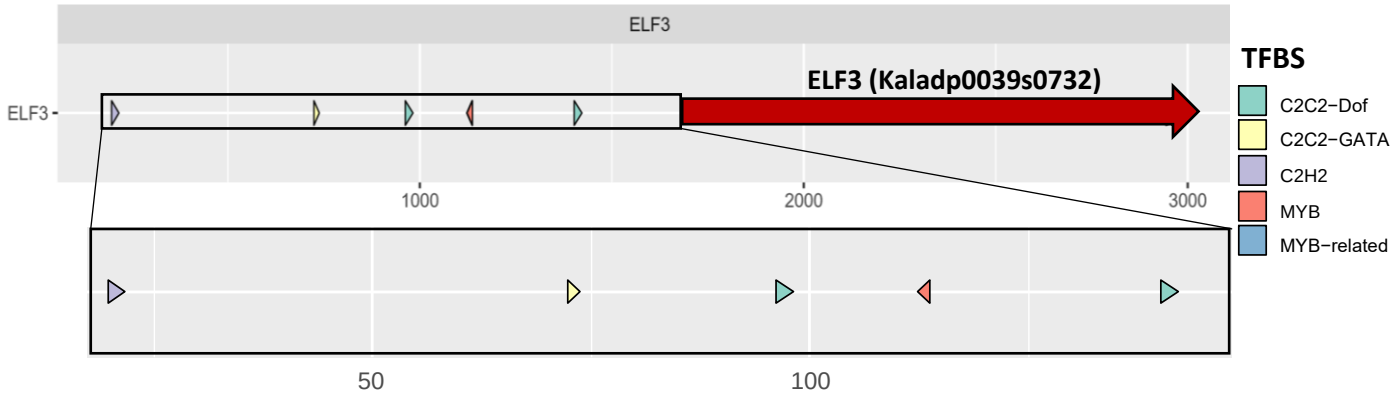
Transcription factor gene ID	Transcription factor family	Target gene ID	Target gene function	start	stop	strand	score	p-value
Kaladp0081s0001	C2H2	Kaladp0057s0097	PIF3	595	613	+	9.15833	5.83E-05
Kaladp0081s0001	C2H2	Kaladp0057s0097	PIF3	597	615	+	8.59167	7.73E-05
Kaladp0042s0072	HSF	Kaladp0057s0097	PIF3	312	326	+	11.6622	1.26E-05
Kaladp0042s0072	HSF	Kaladp0057s0097	PIF3	610	624	-	6.27027	8.70E-05
Kaladp0071s0407	HB-KNOX	Kaladp0057s0097	PIF3	597	616	+	13.5333	9.44E-06
Kaladp0045s0236	bHLH	Kaladp0057s0097	PIF3	1548	1561	-	12.0938	5.37E-05
Kaladp0022s0176	CPP	Kaladp0057s0097	PIF3	547	556	+	11.2929	1.00E-04
Kaladp0279s0002	AP2/ERF-ERF	Kaladp0057s0097	PIF3	574	588	+	8.57812	8.20E-05
Kaladp0042s0084	WRKY	Kaladp0057s0097	PIF3	803	816	-	6.8125	9.00E-05
Kaladp1246s0007	HB-HD-ZIP	Kaladp0057s0097	PIF3	273	283	-	12.3382	4.40E-05

Figure 15: Transcription factors as regulatory candidates of *PIF3* Edges with supporting evidence from transcription factor binding site (TFBS) prediction within the promoter regions were classified as candidate regulatory relationships. Multiple TFBS with their associated binding positions were illustrated in the diagram with their statistical scores listed in the table (Edge score $\geq 1e^{-4}$, p-value $\leq 1e^{-4}$, Promoter region = 1,700 bp with 1,500 bp upstream and 200 bp downstream of the target gene starting position).



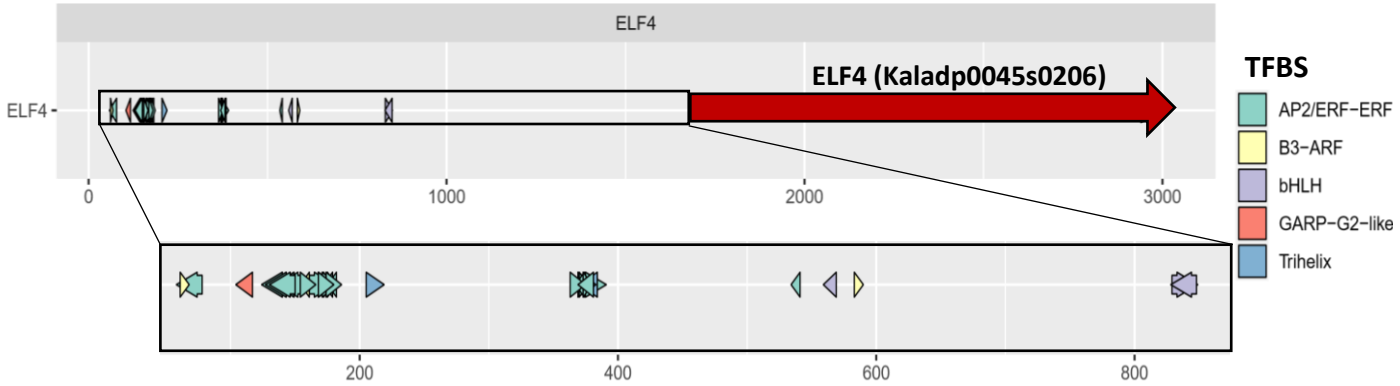
Transcription factor gene ID	Transcription factor family	Target gene ID	Target gene function	start	stop	strand	score	p-value
Kaladp0073s0079	GARP-G2-like	Kaladp0060s0460	HY5	490	502	-	11.4375	5.88E-05
Kaladp1295s0017	MYB-related	Kaladp0060s0460	HY5	778	792	-	12.8594	2.34E-05
Kaladp0043s0101	MYB-related	Kaladp0060s0460	HY5	4	13	-	12.5158	2.49E-05
Kaladp0037s0181	AP2/ERF-ERF	Kaladp0060s0460	HY5	46	66	+	7.15942	8.35E-05

Figure 16: Transcription factors as regulatory candidates of *HY5* Edges with supporting evidence from transcription factor binding site (TFBS) prediction within the promoter regions were classified as candidate regulatory relationships. Multiple TFBS with their associated binding positions were illustrated in the diagram with their statistical scores listed in the table (Edge score ≥ $1e^{-4}$, p-value ≤ $1e^{-4}$, Promoter region = 1,700 bp with 1,500 bp upstream and 200 bp downstream of the target gene starting position).



Transcription factor gene ID	Transcription factor family	Target gene ID	Target gene function	start	stop	strand	score	p-value
Kaladp1295s0017	MYB-related	Kaladp0039s0732	ELF3	1122	1136	-	10.7969	9.41E-05
Kaladp0053s0101	C2C2-Dof	Kaladp0039s0732	ELF3	1402	1422	+	12.2969	3.12E-05
Kaladp0053s0101	C2C2-Dof	Kaladp0039s0732	ELF3	962	982	+	11.4375	5.98E-05
Kaladp0048s0596	C2C2-GATA	Kaladp0039s0732	ELF3	724	738	+	2.67188	8.17E-05
Kaladp0063s0022	MYB	Kaladp0039s0732	ELF3	1123	1137	-	11.2969	7.13E-05
Kaladp0085s0007	C2H2	Kaladp0039s0732	ELF3	197	216	+	15.7969	2.58E-06

Figure 17: Transcription factors as regulatory candidates of *ELF3* Edges with supporting evidence from transcription factor binding site (TFBS) prediction within the promoter regions were classified as candidate regulatory relationships. Multiple TFBS with their associated binding positions were illustrated in the diagram with their statistical scores listed in the table (Edge score ≥ $1e^{-4}$, p-value ≤ $1e^{-4}$, Promoter region = 1,700 bp with 1,500 bp upstream and 200 bp downstream of the target gene starting position).

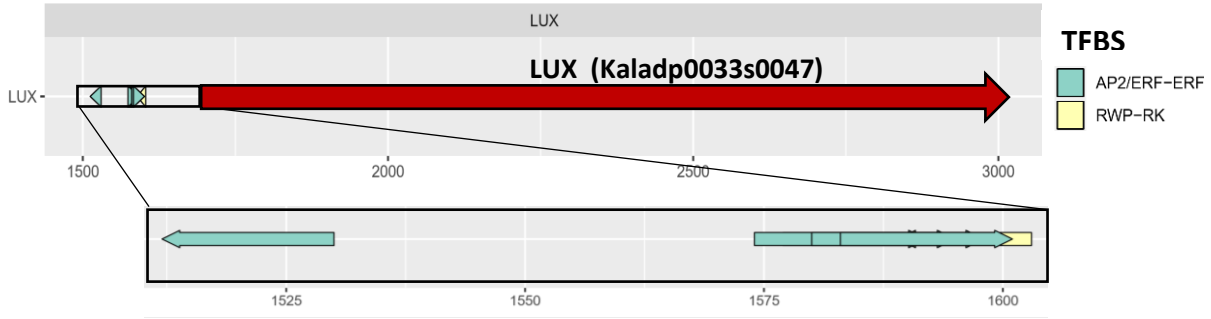


Transcription factor gene ID	Transcription factor family	Target gene ID	Target gene function	start	stop	strand	score	p-value
Kaladp0037s0181	AP2/ERF-ERF	Kaladp0045s0206	ELF4	133	153	-	20.1304	8.77E-08
Kaladp0037s0181	AP2/ERF-ERF	Kaladp0045s0206	ELF4	164	184	+	16.9565	6.60E-07
Kaladp0037s0181	AP2/ERF-ERF	Kaladp0045s0206	ELF4	161	181	+	11.058	1.52E-05
Kaladp0037s0181	AP2/ERF-ERF	Kaladp0045s0206	ELF4	124	144	-	10.8841	1.65E-05
Kaladp0037s0181	AP2/ERF-ERF	Kaladp0045s0206	ELF4	371	391	+	10.7536	1.75E-05
Kaladp0037s0181	AP2/ERF-ERF	Kaladp0045s0206	ELF4	136	156	-	8.98551	3.89E-05
Kaladp0037s0181	AP2/ERF-ERF	Kaladp0045s0206	ELF4	127	147	-	8.97101	3.91E-05
Kaladp0037s0181	AP2/ERF-ERF	Kaladp0045s0206	ELF4	130	150	-	8.92754	3.98E-05
Kaladp0037s0181	AP2/ERF-ERF	Kaladp0045s0206	ELF4	58	78	-	6.92754	9.17E-05
Kaladp0011s0067	B3-ARF	Kaladp0045s0206	ELF4	154	161	-	13.8506	3.81E-05
Kaladp0011s0067	B3-ARF	Kaladp0045s0206	ELF4	61	68	+	7.52874	9.73E-05
Kaladp0011s0067	B3-ARF	Kaladp0045s0206	ELF4	583	590	+	7.52874	9.73E-05
Kaladp0033s0047	GARP-G2-like	Kaladp0045s0206	ELF4	104	117	-	11.0167	5.71E-05
Kaladp0747s0001	bHLH	Kaladp0045s0206	ELF4	559	569	-	12.1351	4.87E-05
Kaladp0011s0356	Trihelix	Kaladp0045s0206	ELF4	205	219	+	14.25	7.50E-06
Kaladp0011s0356	Trihelix	Kaladp0045s0206	ELF4	370	384	-	13.5781	1.12E-05

Figure 18: Transcription factors as regulatory candidates of *ELF4* (continue next page)

Transcription factor gene ID	Transcription factor family	Target gene ID	Target gene function	start	stop	strand	score	p-value
Kaladp0011s0356	Trihelix	Kaladp0045s0206	ELF4	369	383	+	11.3906	3.58E-05
Kaladp0039s0249	AP2/ERF-ERF	Kaladp0045s0206	ELF4	172	179	-	18	1.13E-06
Kaladp0039s0249	AP2/ERF-ERF	Kaladp0045s0206	ELF4	175	182	-	18	1.13E-06
Kaladp0039s0249	AP2/ERF-ERF	Kaladp0045s0206	ELF4	370	377	+	10.0952	2.12E-05
Kaladp0039s0249	AP2/ERF-ERF	Kaladp0045s0206	ELF4	367	374	-	10.0952	2.12E-05
Kaladp0039s0249	AP2/ERF-ERF	Kaladp0045s0206	ELF4	534	541	-	10.0952	2.12E-05
Kaladp0039s0249	AP2/ERF-ERF	Kaladp0045s0206	ELF4	373	380	+	9.29524	5.91E-05
Kaladp0039s0249	AP2/ERF-ERF	Kaladp0045s0206	ELF4	172	179	+	9.12381	7.30E-05
Kaladp0039s0249	AP2/ERF-ERF	Kaladp0045s0206	ELF4	175	182	+	9.12381	7.30E-05
Kaladp0089s0020	AP2/ERF-ERF	Kaladp0045s0206	ELF4	171	183	+	17.8594	4.71E-07
Kaladp0089s0020	AP2/ERF-ERF	Kaladp0045s0206	ELF4	174	186	+	13.7969	1.22E-05
Kaladp0089s0020	AP2/ERF-ERF	Kaladp0045s0206	ELF4	154	166	+	13.2344	1.67E-05
Kaladp0089s0020	AP2/ERF-ERF	Kaladp0045s0206	ELF4	363	375	+	12.5	2.38E-05
Kaladp0089s0020	AP2/ERF-ERF	Kaladp0045s0206	ELF4	168	180	+	9.40625	6.87E-05
Kaladp0089s0020	AP2/ERF-ERF	Kaladp0045s0206	ELF4	369	381	-	9.29688	7.06E-05
Kaladp0050s0101	bHLH	Kaladp0045s0206	ELF4	829	849	+	13.7188	1.38E-05
Kaladp0050s0101	bHLH	Kaladp0045s0206	ELF4	828	848	-	11.0469	5.72E-05

Figure 18: Transcription factors as regulatory candidates of *ELF4* Edges with supporting evidence from transcription factor binding site (TFBS) prediction within the promoter regions were classified as candidate regulatory relationships. Multiple TFBS with their associated binding positions were illustrated in the diagram with their statistical scores listed in the table (Edge score $\geq 1e^{-4}$, p-value $\leq 1e^{-4}$, Promoter region = 1,700 bp with 1,500 bp upstream and 200 bp downstream of the target gene starting position).



Transcription factor gene ID	Transcription factor family	Target gene ID	Target gene function	start	stop	strand	score	p-value
Kaladp0082s0024	RWP-RK	Kaladp0033s0047	LUX	1589	1603	-	13.971	9.15E-06
Kaladp0081s0316	AP2/ERF-ERF	Kaladp0033s0047	LUX	1512	1530	-	13.9531	5.04E-06
Kaladp0081s0316	AP2/ERF-ERF	Kaladp0033s0047	LUX	1577	1595	+	10.0312	3.47E-05
Kaladp0081s0316	AP2/ERF-ERF	Kaladp0033s0047	LUX	1574	1592	+	8.9375	5.57E-05
Kaladp0081s0316	AP2/ERF-ERF	Kaladp0033s0047	LUX	1580	1598	+	8.82812	5.83E-05
Kaladp0081s0316	AP2/ERF-ERF	Kaladp0033s0047	LUX	1583	1601	+	8.75	6.03E-05

Figure 19: Transcription factors as regulatory candidates of *LUX* Edges with supporting evidence from transcription factor binding site (TFBS) prediction within the promoter regions were classified as candidate regulatory relationships. Multiple TFBS with their associated binding positions were illustrated in the diagram with their statistical scores listed in the table (Edge score ≥ $1e^{-4}$, p-value ≤ $1e^{-4}$, Promoter region = 1,700 bp with 1,500 bp upstream and 200 bp downstream of the target gene starting position).

Stomatal subnetwork:

Genes which have been associated with stomata function in *Kalanchoe fedtschenkoi* comprised 20 genes from 14 gene families, namely ABA Insensitive (ABI:2 genes), Plasma membrane protein ATPase 2 (AHA2), Potassium transporter 1 (AKT1), Aluminium-activated malate transporter 9 (ALMT9), BRI1-associated receptor kinase (BAK1), Blue light signalling 1 (BLUS1), Calcium-dependent protein kinase (CPK: 3 genes), High leaf temperature 1 (HT1) Potassium channel in *Arabidopsis thaliana* (KAT:2 genes), Open stomata (OST:2 genes), Phototropin (PHOT:2 genes), Pyrabactin resistance 1-like 9 (PYL9), Quick-activating anion channel 1 (QUAC1/ALMT12), and Slow anion channel-associated 1 (SLAC1). Among these stomatal-function related genes, the Open stomata 1 (OST1) with the gene ID of Kaladp0016s0289 and Slow anion channel-associated 1 (SLAC1) with the gene ID of Kaladp0050s0214 are key genes of interest due to their central role in stomatal function in C3 plants. In addition, Phototropin 2 (PHOT2), with the gene ID of Kaladp0033s0113, is the key gene of interest for the light-signalling pathway in CAM, from evidence of a convergence shift in the diel expression pattern from dawn in C3 *Arabidopsis thaliana* to dusk in two CAM species (*K. fedtschenkoi* and *Ananas comosus*) (Yang *et al.* 2017).

The stomatal subnetwork contained 128 directional edges from 95 transcription factor source nodes associated with the 19 stomatal gene target nodes (Figure 20). These 95 transcription factor nodes belonged to 34 different transcription factor families out of the total number of 69 transcription factor families identified in *K. fedtschenkoi* (Zheng *et al.* 2016). The maximum number of unique target genes sharing the same transcription factor is 4 target genes per 1 transcription factor gene in this stomatal subnetwork. The three transcription factor genes with this characteristic of 1 transcription factor binding to 4 targets were Kaladp0047s0231 (Dof family), Kaladp0073s0028 (RAV family) and Kaladp0085s0007 (C2H2 family) (Figure 21).

With the focus on the transcription factor binding site (TFBS) data, there were 9 transcription factors forming regulatory links onto the promoter of *OST1* gene (Kaladp0016s0289), 5 transcription factors forming regulatory links onto the promoter of *SLAC1* gene (Kaladp0050s0214), and 3 transcription factors forming regulatory links onto the promoter of *PHOT2* gene (Kaladp0033s0113) (Figure 22-24). Among these three target genes of interest, there are no common transcription factor family controlling all 3 genes whereas there are members of bZIP control both *OST1* and *SLAC1* and members of HB-HD-ZIP control *SLAC1* and *PHOT2*.

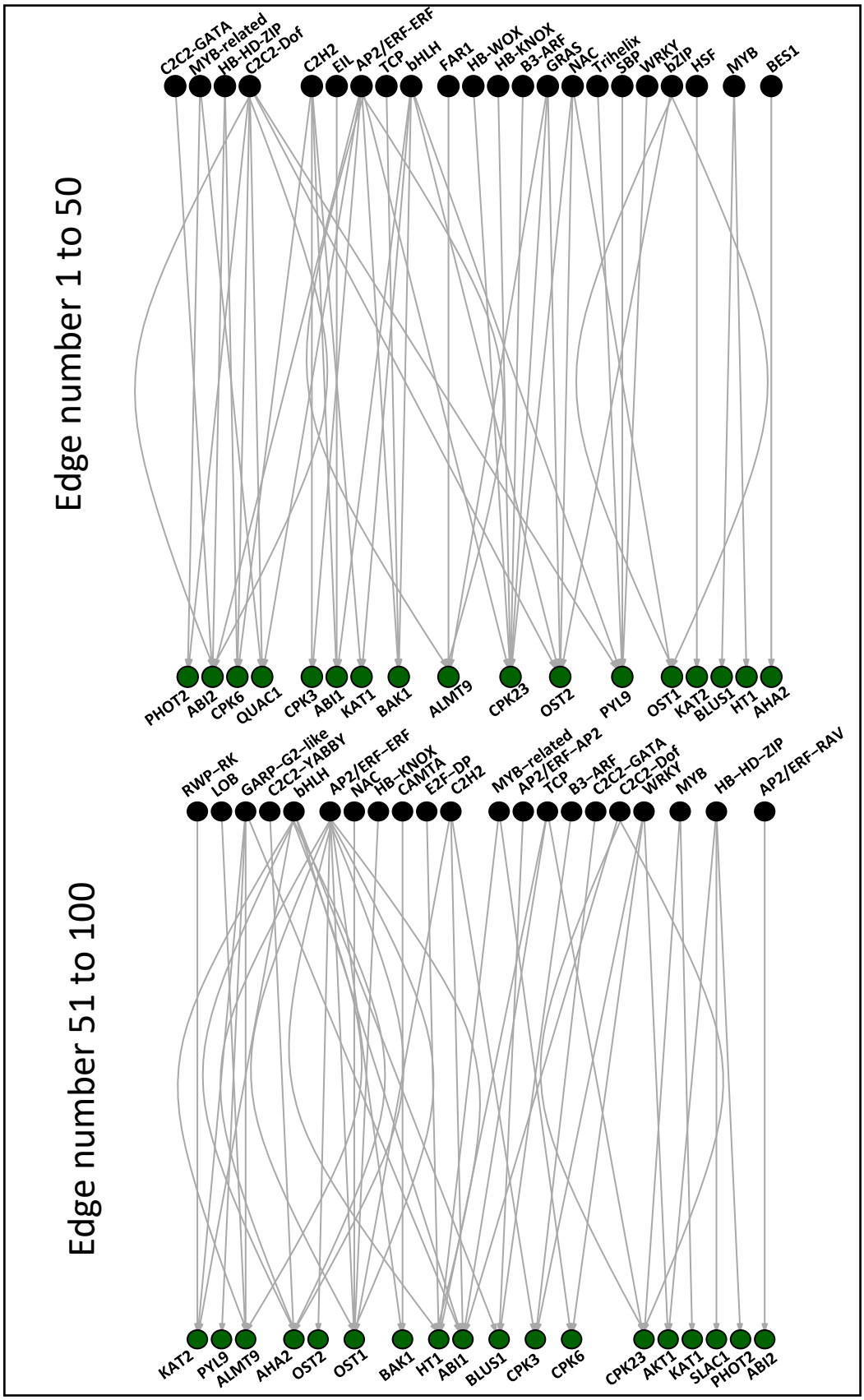


Figure 20: Stomatal subnetwork with transcription factors as source nodes showing the top 100 edges from transcription factor source nodes (black) to stomatal target nodes (green). Transcription factor nodes and stomatal target nodes were grouped by transcription factor families and circadian gene families respectively (Edge score $\geq 1e^{-4}$ and p-value $\leq 1e^{-4}$).

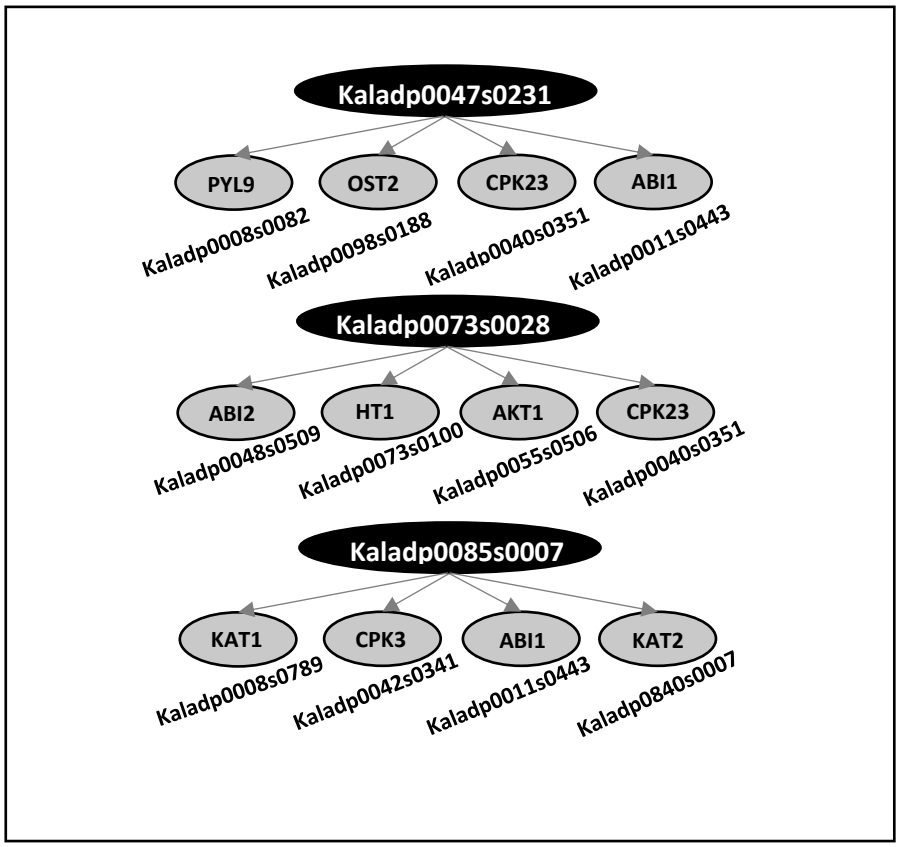
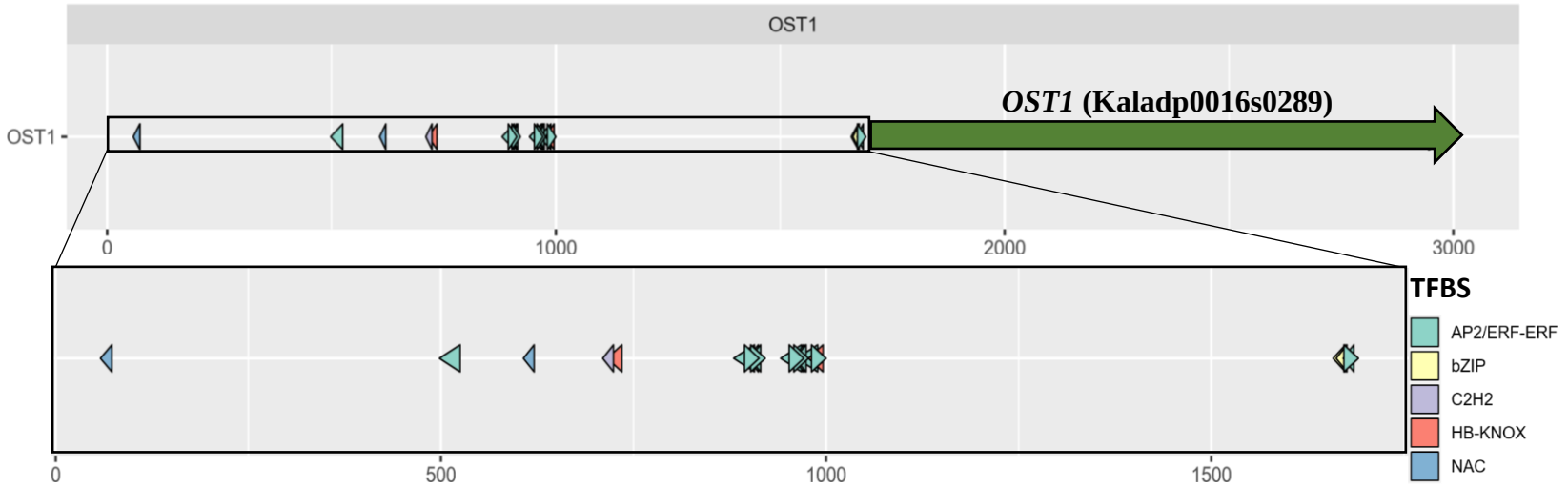


Figure 21: Multiple stomatal genes sharing the same transcription factor source nodes Transcription factors with the maximum number of different target genes within the stomatal subnetwork (4 target genes:1 transcription factor gene) were selected to plot in this figure. Black nodes indicate the transcription factor gene ID and family, whereas grey nodes indicate the target gene ID and gene family (Edge score $\geq 1e^{-4}$ and p-value $\leq 1e^{-4}$).

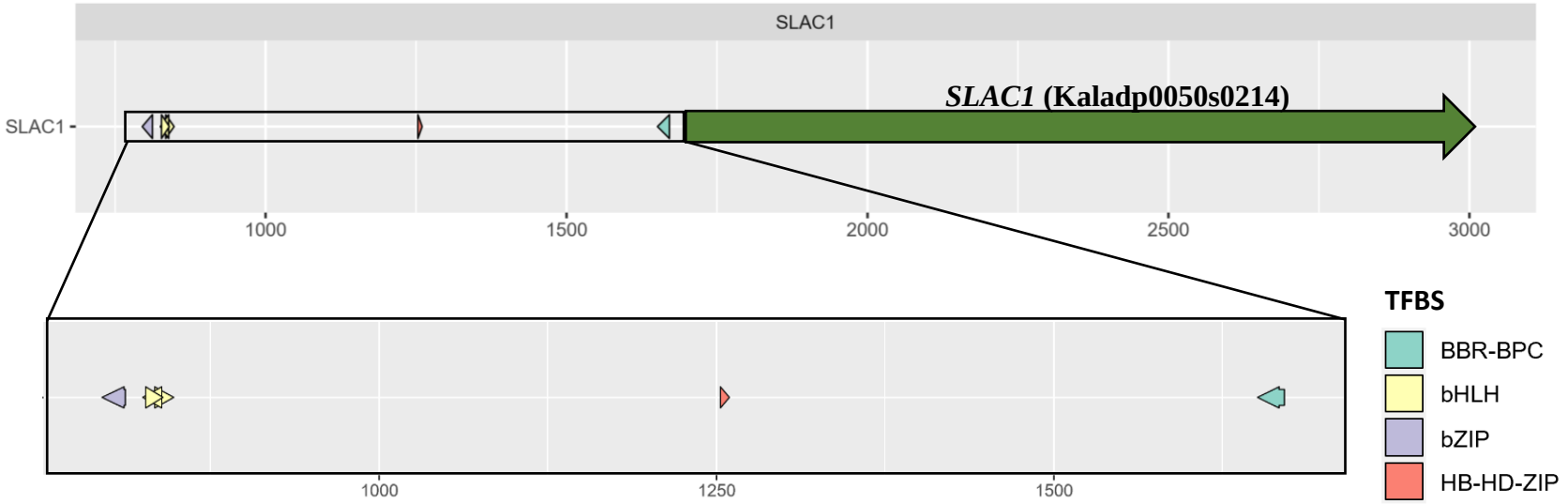


Transcription factor gene ID	Transcription factor family	Target gene ID	Target gene function	start	stop	strand	score	p-value
Kaladp0076s0065	bZIP	Kaladp0016s0289	OST1	1658	1674	-	8.57812	5.28E-05
Kaladp0011s0508	bZIP	Kaladp0016s0289	OST1	59	73	-	8.39062	9.14E-05
Kaladp0011s0508	bZIP	Kaladp0016s0289	OST1	1662	1676	-	8.26562	9.59E-05
Kaladp0091s0012	NAC	Kaladp0016s0289	OST1	607	621	-	13	2.18E-05
Kaladp0085s0049	NAC	Kaladp0016s0289	OST1	58	73	-	9.65625	9.91E-05
Kaladp0040s0059	AP2/ERF-ERF	Kaladp0016s0289	OST1	947	974	-	13.2031	5.26E-06
Kaladp0040s0059	AP2/ERF-ERF	Kaladp0016s0289	OST1	944	971	-	12.6562	6.28E-06
Kaladp0040s0059	AP2/ERF-ERF	Kaladp0016s0289	OST1	941	968	-	10.9219	1.04E-05
Kaladp0040s0059	AP2/ERF-ERF	Kaladp0016s0289	OST1	498	525	-	5.35938	4.93E-05
Kaladp0040s0059	AP2/ERF-ERF	Kaladp0016s0289	OST1	880	907	-	4.375	6.50E-05

Figure 22: Transcription factors as regulatory candidates of *OST1* (continue next page)

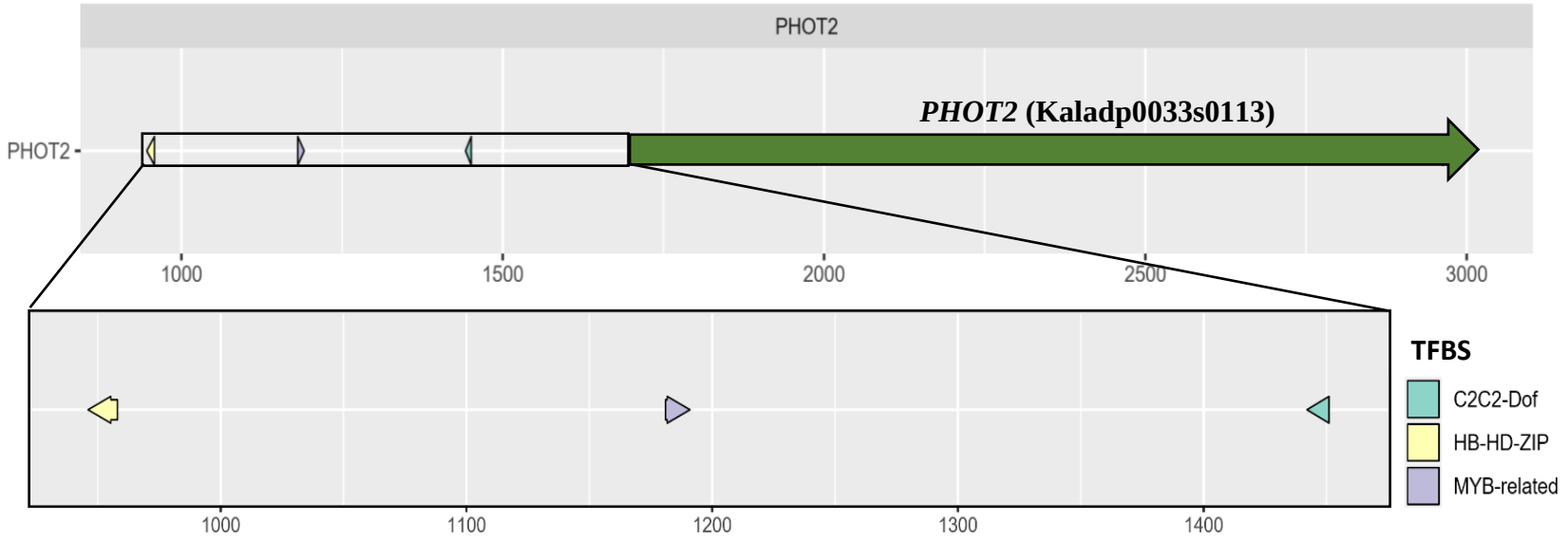
Transcription factor gene ID	Transcription factor family	Target gene ID	Target gene function	start	stop	strand	score	p-value
Kaladp0040s0059	AP2/ERF-ERF	Kaladp0016s0289	OST1	962	989	-	3.51562	8.21E-05
Kaladp0071s0407	HB-KNOX	Kaladp0016s0289	OST1	977	996	-	12.3067	1.86E-05
Kaladp0071s0407	HB-KNOX	Kaladp0016s0289	OST1	716	735	-	10.6933	4.14E-05
Kaladp0039s0495	AP2/ERF-ERF	Kaladp0016s0289	OST1	957	971	-	13.6094	6.22E-06
Kaladp0039s0495	AP2/ERF-ERF	Kaladp0016s0289	OST1	898	912	-	13.1562	7.82E-06
Kaladp0039s0495	AP2/ERF-ERF	Kaladp0016s0289	OST1	1671	1685	-	8.59375	6.05E-05
Kaladp0039s0495	AP2/ERF-ERF	Kaladp0016s0289	OST1	901	915	-	8.4375	6.44E-05
Kaladp0039s0495	AP2/ERF-ERF	Kaladp0016s0289	OST1	960	974	-	8.01562	7.60E-05
Kaladp0008s0200	C2H2	Kaladp0016s0289	OST1	710	724	-	13.2923	1.15E-05
Kaladp0071s0445	AP2/ERF-ERF	Kaladp0016s0289	OST1	902	921	+	14.2656	6.68E-06
Kaladp0071s0445	AP2/ERF-ERF	Kaladp0016s0289	OST1	894	913	+	12.2656	2.06E-05
Kaladp0071s0445	AP2/ERF-ERF	Kaladp0016s0289	OST1	1672	1691	+	12.1406	2.20E-05
Kaladp0071s0445	AP2/ERF-ERF	Kaladp0016s0289	OST1	958	977	+	11.25	3.42E-05
Kaladp0071s0445	AP2/ERF-ERF	Kaladp0016s0289	OST1	952	971	+	10.8906	4.05E-05
Kaladp0071s0445	AP2/ERF-ERF	Kaladp0016s0289	OST1	981	1000	+	10.0469	5.94E-05

Figure 22: Transcription factors as regulatory candidates of *OST1* Edges with supporting evidence from transcription factor binding site (TFBS) prediction within the promoter regions were classified as candidate regulatory relationships. Multiple TFBS with their associated binding positions were illustrated in the diagram with their statistical scores listed in the table (Edge score $\geq 1e^{-4}$, p-value $\leq 1e^{-4}$, Promoter region = 1,700 bp with 1,500 bp upstream and 200 bp downstream of the target gene starting position).



Transcription factor gene ID	Transcription factor family	Target gene ID	Target gene function	start	stop	strand	score	p-value
Kaladp0024s0501	HB-HD-ZIP	Kaladp0050s0214	SLAC1	1253	1260	+	11.6117	6.13E-05
Kaladp0042s0123	bZIP	Kaladp0050s0214	SLAC1	795	812	-	9.89062	2.49E-05
Kaladp0071s0038	bHLH	Kaladp0050s0214	SLAC1	834	848	+	10.625	4.32E-05
Kaladp0071s0038	bHLH	Kaladp0050s0214	SLAC1	825	839	-	7.64062	9.66E-05
Kaladp0053s0637	bHLH	Kaladp0050s0214	SLAC1	827	840	+	12.5811	2.84E-05
Kaladp0001s0060	BBR-BPC	Kaladp0050s0214	SLAC1	1651	1671	-	-2.77027	8.52E-05

Figure 23: Transcription factors as regulatory candidates of *SLAC1* Edges with supporting evidence from transcription factor binding site (TFBS) prediction within the promoter regions were classified as candidate regulatory relationships. Multiple TFBS with their associated binding positions were illustrated in the diagram with their statistical scores listed in the table (Edge score ≥ $1e^{-4}$, p-value ≤ $1e^{-4}$, Promoter region = 1,700 bp with 1,500 bp upstream and 200 bp downstream of the target gene starting position).



Transcription factor gene ID	Transcription factor family	Target gene ID	Target gene function	start	stop	strand	score	p-value
Kaladp1246s0018	Kaladp0033s0113	PHOT2	MYB-related	1181	1191	+	11.5469	7.50E-05
Kaladp0043s0142	Kaladp0033s0113	PHOT2	C2C2-Dof	1442	1451	-	12.1695	1.77E-06
Kaladp0042s0372	Kaladp0033s0113	PHOT2	HB-HD-ZIP	946	958	-	10.4	8.99E-05

Figure 24: Transcription factors as regulatory candidates of *PHOT2* Edges with supporting evidence from transcription factor binding site (TFBS) prediction within the promoter regions were classified as candidate regulatory relationships. Multiple TFBS with their associated binding positions were illustrated in the diagram with their statistical scores listed in the table (Edge score ≥ $1e^{-4}$, p-value ≤ $1e^{-4}$, Promoter region = 1,700 bp with 1,500 bp upstream and 200 bp downstream of the target gene starting position).

Shared regulators between subnetworks:

This study identified 21 transcription factor genes that reach targets across all 4 subnetworks. These 21 transcription factor genes belong to 12 transcription factor families (Table 5). Similar analysis performed at the level of gene family showed that 25 families of transcription factor from the total of 69 families reached targets across all 4 subnetworks (Table 6).

Notably, the same dataset presented in this chapter can be visualised differently by comparing the upstream regions of CAM-specific isoforms of each CAM gene against their homologs/paralogs that are not used in CAM. Subsequently, further comparisons between subnetworks can be made. For example, the carboxylation subnetwork, decarboxylation subnetwork and stomatal subnetwork may have acquired different regulatory outputs from the circadian network components.

Table 5: The 21 transcription factor genes that reach targets across all 4 subnetworks

Transcription factor		Target genes			
Family	**Gene ID**	**Carboxylation**	**Decarboxylation**	**Circadian**	**Stomata**
AP2/ ERF-ERF	Kaladp0051s0049	MDH,ALMT6,PEPC	NAD-ME	STIPL1,ZTL	ALMT9,CPK23
	Kaladp0039s0249	ALMT6,MDH	NAD-ME	CHE,COP1,ELF4,JMJD5	ALMT9,BAK1
	Kaladp0037s0181	PEPC,PPCK,MDH	PEPCK,PPDK	COP1,ELF4,HY5,PRR5/9,RVE8	ABI2
	Kaladp0048s0420	MDH,PEPC	NAD-ME	PHYB,RVE6	OST2/AHA1
	Kaladp0021s0014	MDH, β-CA,PEPC	NAD-ME,PEPCK	RVE1,PRR5/9	CPK3
	Kaladp0039s0495	PPCK	PPDK	PHYA	OST1
BBR-BPC	Kaladp0001s0060	ALMT	NAD-ME,NADP-ME,PEPCK	PHYA	SLAC1
bHLH	Kaladp0057s0097	ALMT, β-CA	NADP-ME,PPDK-RP	PRR5/9, PHYA	AHA2
	Kaladp0747s0001	MDH	PPDK	CRY1,ELF4	AHA2
C2C2-Dof	Kaladp0053s0101	ALMT,MDH	PPDK-RP	ELF3,JMJD5,TEJ	ABI2
C2C2-GATA	Kaladp0015s0060	ALMT,MDH	PPDK-RP	CRY1	CPK3,CPK6
C2H2	Kaladp0085s0007	MDH	NAD-ME	ELF3	ABI1,CPK3,KAT1,KAT2
	Kaladp0060s0097	PEPC2	PPDK-RP	FIO1,LNK1	CPK3
GRAS	Kaladp0059s0271	PEPC,MDH,ALMT6	PEPCK	PIF3,PRR3,GI	ALMT9,CPK23,OST2/AHA1
HB-KNOX	Kaladp0071s0407	MDH	NAD-ME,NADP-ME	PIF3	CPK23,OST1
LOB	Kaladp0089s0088	ALMT6,MDH	NAD-ME	RVE6	ALMT9

Table 5 (continue): The 21 transcription factor genes that reach targets across all 4 subnetworks

Transcription factor		Target genes			
Family	**Gene ID**	**Carboxylation**	**Decarboxylation**	**Circadian**	**Stomata**
MADS-MIKC	Kaladp0016s0148	ALMT,MDH,PEPC1	NADP-ME,PEPCK	JMJD5,PHYB,RVE8,TEJ	ABI2
RWP-RK	Kaladp0082s0024	MDH,β-CA	NAD-ME	LNK2,LUX,SKIP	KAT2
TCP	Kaladp0093s0044	MDH	PPDK-RP	LNK2	CPK23
	Kaladp0630s0020	PEPC	PPDK-RP	LNK2,TEJ	ABI1,HT1
	Kaladp0131s0006	PPCK1,ALMT,β-CA	PEPCK	CHE,CKB4	BAK1

Table 6: All transcription factor families of *K. fedtschenkoi* with 25 families reaching target across all four subnetworks. Total number of transcription factor families is 69 with MYB as the largest family and AP2/ERF-AP2 and CPP as the smallest families. Green box indicates the *presence* of target in the subnetwork. White box indicate the *absence* of target in the subnetwork.

Transcription factor family	Number of members	Targets			
		Carboxylation	Decarboxylation	Circadian	Stomata
MYB	235				
bHLH	218				
bZIP	185				
AP2/ERF-ERF	173				
C2H2	162				
MYB-related	138				
NAC	119				
C3H	110				
B3	104				
WRKY	103				
GARP-G2-like	93				
MADS-MIKC	88				
FAR1	78				
HB-HD-ZIP	61				
C2C2-GATA	60				
GRAS	60				
B3-ARF	59				
MADS-M-type	55				
HSF	53				
LOB	52				
C2C2-Dof	48				
Trihelix	46				
C2C2-YABBY	44				
GARP-ARR-B	44				
RWP-RK	35				
OFP	32				
SBP	31				
NF-YB	30				
HB-BELL	29				

Transcription factor family	Number of members	Targets			
		Carboxylation	Decarboxylation	Circadian	Stomata
TCP	29				
HB-other	28				
NF-YC	28				
TUB	27				
NF-YA	25				
CAMTA	23				
HB-WOX	23				
AP2/ERF-AP2	22				
CPP	22				

Discussion:

As an author, I am responsible for setting these large sets of gene regulatory network data in an evolutionary context, to facilitate the understanding of Crassulacean Acid Metabolism (CAM) physiology for general biologists as well as more general readers. To begin with, CAM lineages evolved independently more than 60 times throughout the history of angiosperms. The major CAM speciation events arose after a reduction in atmospheric CO_2 and climatic shifts associated with the uplift of the Tibetan Plateau in the Miocene-Pliocene: CAM probably originated between 10- 20 Ma, and diversified in many families in the past 5 million years (Givnish *et al.* 2014). During this period, ambient CO_2 concentration was reduced to the lowest levels in palaeohistory, and during the Quaternary has hovered between 180 and 280 ppm for the past 2 million years. Such a reduction in CO_2 was also associated with increased seasonality in many of the savanna regions where C4 systems had developed and diversified for an estimated of 5-10 Ma. The evolution of CAM seems to have been favoured in semi-arid habitats, where a regular recharge of water occurred on a seasonal basis (such as the summer monsoon and winter rains in the Sonoran Desert). Secondly, it is generally accepted that the origins of CAM lie in the molecular and physiological rescheduling of existing C3 systems to generate the inverse cycle of stomatal opening and associated temporally reprogrammed carboxylation/decarboxylation cycles. The aim of this current investigation was to take advantage of the recent proliferation in molecular sequencing data to identify the potential transcriptional regulators which could have signalled this transition. This approach also assumes that cellular succulence and tissue water storage had already evolved developmentally at some earlier point, presumably in response to water limitation in terrestrial and epiphytic habitats. The reduced mesophyll conductance associated with succulence would itself add another selection pressure leading to enhanced CAM activity, owing to the enhanced drawdown in internal CO_2 associated with succulent leaves (Maxwell *et al*. 1997).

Given the critical role in the CAM cycle, the first focus was given to the *activation timing* of the key carboxylation enzyme, Phosphoenolpyruvate carboxylase (PEPC), which drives the nocturnal CO_2 uptake process and perhaps through the initial recycling of respiratory CO_2 which may represent an intermediate step in this progression (Griffiths *et al.* 1989). PEPC is known to be under the control of PEPC kinase (PPCK), which in turn is responsive to (as yet) unidentified transcriptional regulatory controls. In current CAM systems, it is the exact timing of PEPC kinase gene expression that is crucial to the entire PEPC primary fixation process

(Hartwell *et al.* 1996; Hartwell *et al.* 1999; Nimmo 2003). Therefore, PEPC kinase regulators were also key for addressing the development of CAM. It was also timely to focus upon such a well-defined genus *Kalanchoe*, for which so many physiological and molecular advances have been made. Nonetheless, one cannot rule out other possible 'selective-agents' such as the malate accumulation process (thus focusing on malate transporter too, ALMT, tDT), since if the malate cannot be stored at night, a futile cycle would have caused an energetic penalty for CAM (with acknowledgements to J Andrew C Smith, personal discussion).

Alternatively, in the subsequent day-time phase (Phase III), another regulatory step could be the decarboxylation process, which functions in releasing CO_2 from the storage malate to feed the secondary carboxylation reaction catalysed by Rubisco. Interestingly, the decarboxylation process has always been less of a focus than carboxylation processes when one considers CAM functioning. Hence, the decarboxylation subnetwork has been included in our analysis as the decarboxylation process could be as important as carboxylation for CAM to function (Ceusters *et al.* 2021).

Thirdly, the circadian network has been shown on multiple occasions to be central to every organism function ranging from single celled algae to complex plant or mammalian organs. Unfortunately, we need to put the circadian subnetwork of CAM lower down the ranking when it comes to CAM evolution, as the exact identity of CAM central circadian network is not well understood. It would not be surprising to have two-way interactions between CAM carboxylation/decarboxylation processes and the rhythmicity of the core circadian network. This would have ensured that entrainment can be tailored specifically to CAM. Indeed, it has been shown from recent RNAi knockdown experiments using *Kalanchoe laxiflora* (very similar to *K. fedtschenkoi*) that the lack of normal PEPC activity resulted in the clock arrhythmia –and *vice versa* (Boxall *et al.* 2017; Moseley *et al.* 2018; Boxall *et al.* 2020).

Finally, the stomatal subnetwork has also been included in our analysis. This question of stomatal functioning in CAM was the very first question that brought my personal attention to CAM research in 2015. An email asking "How could CAM stomata open at night?" was sent out to two professors at the end of the first year physiology lecture series. One of them replied to this disoriented undergraduate by inviting her to come talk to members of his research group. Although 7 years later, none can provide a satisfactory answer to this question, including myself. The best we know is that the interplay between intercellular CO_2 (Ci) and other canonical stomatal signalling may be sufficient to result in CAM stomatal behaviour.

Intriguingly, my supervisor and his long-term collaborator have kindly pointed out to me that perhaps I was asking a slightly wrong question. I could have asked why CAM stomata close during the day! The limited stomatal opening at night could be a response to the high CO_2 drawdown associated with the PEPC carboxylation capacity and affinity even at relatively low stomatal conductance. The question of whether high Ci triggers stomatal closure during the day remains an active area of study, alongside with the attempts to dissect the complete CO_2 sensing pathway in C3 species by the Julian Schroeder group in the USA (Engineer *et al.* 2016). The best hope for our GRN approach to address these questions is to characterise the candidate regulators of key stomatal signalling genes. If there are any shared regulators between the stomatal subnetwork and the carboxylation subnetwork, then we could propose that there may be a simultaneously selective pressure acting on multiple pre-existing C3 genes in such a way that allows stomatal behaviour and carboxylation timing to be orchestrated/synchronised by the same transcription factors or same transcription factor families.

Finally, it should be noted that the question of regulation through carbohydrate balance (in terms of supplying PEP for the nocturnal carboxylation and demand for resynthesis of reserves and partitioning for growth by day) should also be considered as another key subnetwork. Suffice it to say that the analysis of the transcriptional regulators of carbohydrate metabolism is underway, but was not completed in time for inclusion in this Chapter. Nonetheless, it is very important to emphasise that our overall findings suggested that transcriptional regulations may be the basis of the molecular evolutionary mechanisms for *K. fedtschenkoi,* while leaving the gap open as to whether this would also be conserved in other sister CAM clades. Nevertheless, this study is a crucial step for the understanding of CAM as *Kalanchoë fedtschenkoi* is one of the best model CAM species, being the first eudicot CAM lineage with a genome sequence to date, as well as being the first sequenced species in the distinct eudicot lineage of Saxifragales.

Discussion point 1: Common transcription factor shared between 4 subnetworks can shed light into the orchestration of the CAM system through potential crosstalk between subnetworks.

From the total of 3,048 transcription factors (TFs) from 69 transcription factor families, our analysis has narrowed attention to 21 transcription factor genes that interact with targets across all 4 subnetworks (carboxylation, decarboxylation, circadian and stomata) as summarised in Table 5. In addition to the shared transcription factor genes, the shared transcription factor families have also been identified. These shared transcription factor families include, but are not limited to, the top 5 largest transcription factor families of *Kalanchoe fedtschenkoi* (i.e. MYB, bHLH, bZIP, AP2/ERF-ERF and C2H2) (Table 6). Interestingly, the smallest transcription factor family that spans across all 4 subnetworks is the BBR-BPC family with only 10 TF genes in this family.

In C3 species, transcription factors from MYB, bHLH, bZIP, AP2/ERF-ERF and C2H2 families are involved in various processes including developmental and stress responses (Ambawat *et al.* 2013; Wang *et al.* 2018; Sun *et al.* 2018; Xie *et al.* 2019; Han *et al.* 2020). However, the functions of these transcription factor families in CAM are still inconclusive. The most recent information about transcription factors associated with CAM arose from the following experiments and bioinformatic analysis performed on various CAM species, including both obligate CAM (*Agave tequilana*, *Agave sisalana* and *Kalanchoe fedtschenkoi*) and facultative CAM (*Mesembryanthemum crystallinum* and *Talinum triangulare*).

First, members of AP2/ERF/CRF, NAC, HB, bZIP, MADS/AGAMOUS-LIKE8, and AP2/ERF in *M. crystallinum* were the top 8 transcription factors that showed increased transcript abundance in response to drought stress (Amin *et al.* 2019). The orthologue of these genes in *K. fedtschenkoi* also showed increased transcript abundance during the developmental CAM induction (Amin et al., 2019; Cushman unpublished data). A follow-up functional characterisation of *KfMYB59* by overexpressing this *K. fedtschenkoi* transcription factor in *Arabidopsis thaliana* led to the conclusion that this *KfMYB59* transcription factor functions in plant growth and development (flowering, biomass, rosette size, shoot length) and responses to abiotic stress (increased water-use-efficiency, positive response to selenium stress and higher adaptability to nutrient-limited conditions) in the *Arabidopsis* transgenic line (Amin *et al*. 2019).

Second, there are 47 transcription factors which have been subjected to purifying selection through CAM evolution in monocot *A. tequilana* or *A. sisalana*. These 47 transcription factors belong to bHLH (8 members), zinc finger (23 members), MYB (6 members), AP2 (5 members), NAC (3 members), WRKY (1 member), and bZIP (1 member) families (Huang *et al.* 2018). Third, a comparative transcriptome study found 40 rhythmic transcription factors in *K. fedtschenkoi* with phase-shift from their rhythmic orthologues in *A. thaliana*. These 40 rhythmic transcription factors of *K. fedtschenkoi* belong to 12 families (AP2/ERF-ERF, bHLH, C3H, GRAS, HB-BELL, HB-PHD, HRT, HSF, MYB, SET, Tify and Trihelix) (Moseley *et al.* 2018).

Fourth, Brilhaus *et al,* 2015 identified a number of transcription factors in *Talinum triangulare* that are upregulated to coincide with the onset of CAM and/or sustained throughout the CAM mode (Brilhaus *et al.* 2015). The method by Brilhaus *et al.* provided the most direct evidence of transcription factor expression profiles through the transition from C3 into CAM mode within the timeframe of 4, 9 and 12 days after the onset of drought induction. However, a similar approach has not been undertaken in other CAM species. The study identified the transient upregulation of transcription factors on day 9 after drought induction that include the Drought Responsive element responsive TFs of the C-repeated binding factor (CBF) family. In addition, the same study also identified the sustained upregulation of transcription factors which included ABA responsive transcription factors and growth associated regulators. This information suggested that multiple transcription factors across different families may play important roles in the induction and/or maintenance of CAM. Further, some transcription factors may function in the transient induction phase but not sustained throughout the CAM stage, and *vice versa*. Nevertheless, the functions and identities of transcription factors could differ between CAM clades and CAM types given that the *T. triangulare* is a drought-induced, reversible, facultative CAM system, whereas *K. fedtschenkoi* is an obligate CAM species.

From the 21 transcription factor genes that interact with targets across all 4 subnetworks identified in our study, PIF3 (Kaladp0057s0097) which belongs to bHLH family, and Integrase-type DNA-binding superfamily protein (Kaladp0039s0495), which belongs to AP2/ERF-ERF family, have been previously identified in the comparative rhythmicity analysis to have a phase shift from their orthologues in *A. thaliana* (Moseley *et al.* 2018). The phase of PIF3 in *K. fedtschenkoi* was 6 hours leading its orthologue in *A. thaliana*, whereas the phase of Integrase-type DNA-binding superfamily protein in *K. fedstchenkoi* was 8 hours behind its orthologue in *A. thaliana.* There are 3 out of these 21 transcription factor genes

(Kaladp0093s0044-TCP family, Kaladp0053s0101-C2C2-Dof family and Kaladp0037s0181-AP2/ERF-ERF family) which have rhythmicity in *K. fedschenkoi* but not in its orthologues in *A. thaliana*. There are 6 out of these 21 genes (Kaladp0131s0006-TCP family, Kaladp0087s0163-AP2/ERF-ERF family, Kaladp0071s0407 HB-KNOX family, Kaladp0085s0049-NAC family, Kaladp0089s0088-LOB family, Kaladp0060s0097-C2H2 family) which have rhythmicity in *A. thaliana* but not in its orthologues in *K. fedschenkoi*.

Discussion point 2: Complexity within biological networks may have arisen from gene duplication and expanded gene families.

Reconstructing the regulatory network for any organism is a challenging task due to the complex interplays between various controls, starting from the unfolding of chromatin structure up until the translocation of a functional protein to its right location. It is important to emphasise that the focus of this study is at the transcriptional control, nonetheless, there exists complexity due to a large number of transcription factors forming a combinatorial explosion across the whole network. Understanding the transcriptional regulatory network in plants is further complicated by the fact that plants have larger transcriptional factor families than animal and fungi, as a result of whole genome duplications, local duplications and transposon activities (Hong 2016). A good example of a transcription factor gene family expansion from duplication is the bHLH transcription factor family, the largest family of transcription factor in *A. thaliana*, with 137 and 218 gene members being identified in *A. thaliana* and *K. fedtschenkoi* respectively.

Apart from the complexity arising from an expanded gene family, there is also evidence of transcription factors from different families binding to the same *cis*-element sequence. For example, members of both bHLH and bZIP transcription factor families are known to bind specifically to the same short sequence (G-box element), with a few exceptions of bHLH members binding to the E-box element (Ezer *et al.* 2017). This illustrates the complications in determining the gene regulatory networks based solely on the identification of short *cis*-element sequences in the promoter region because a large number of transcription factors within the same family and across the families may utilise the same short *cis*-element sequence. In other words, due to the complication from genome duplication and shared *cis*-element, a gene regulatory network built from *cis*-element searching alone is not as indicative as a gene

regulatory network built from the direct association between transcription factor identities and their target sites. Previous studies have searched for the enrichment of *cis*-elements within the promoter regions of a sugar-storage obligate CAM species pineapple (*Ananas comosus*). The motif search within promoter regions of key CAM genes in pineapple found the CIRCADIAN CLOCK ASSOCIATED 1 (CCA1)-binding site (CBS; AAAAATCT) and G-box binding site (CACGTG) (Chen *et al.* 2020). This is consistent with our finding that the members of bHLH and bZIP are potential regulators of PPCK1, PPDK-RP and PPDK genes in *K. fedtschenkoi*. It is worth emphasising that the pineapple and *K. fedtschenkoi* belong to different clades of CAM as well as possessing different carbohydrate storage pathways; (pineapple is sugar-storing whereas *K. fedtschenkoi* is starch-storing). On one hand, the similarity in having both bHLH and bZIP controlling key CAM genes perhaps indicate the homologous controls which were key to CAM development. On the other hand, the large family size of both bHLH and bZIP may allow for flexibility in developing CAM by using different family members to control the same key CAM genes, hence relaxing the evolutionary constraint on the exact transcription factor identity but relying more on locating the shared motif in the promoter regions of all key CAM genes.

The network information constructed in this chapter would also allow for the identification of transcription factors that are unique to CAM specific isoforms of each CAM protein compared with the transcription factors of the homologous/paralogous genes which are not used in CAM. This additional analysis could provide a more in-depth information on the regulators upstream of CAM gene expression, which may or may not be involved with the known circadian network.

Discussion point 3: Central dogma of molecular biology does not end at the transcription step but is followed by post-translational and other controls.

It is common to find multiple regulatory steps as a central dogma of molecular biology. The CAM system is no exception. The two key CAM enzymes, phosphoenolpyruvate carboxylase (PEPC) and PYR orthophosphate dikinase (PPDK) are classic examples of the combination of post-translational modification and other forms of regulation. The PEPC kinase (PPCK) enzyme phosphorylates PEPC to make it less sensitive to malate inhibition at night. Similarly, PPDK regulatory protein (PPDK-RP) performs post-translational modifications on PPDK to alter its biochemical properties to either be active or inactive, provided that the PPDK-RP is a bi-functional enzyme as in C4 species (Chastain *et al.* 2018). As a result, PEPC is active at

night for the carboxylation process whereas PPDK is active during the day for the decarboxylation process.

Discussion point 4: Defining 'circadian genes' for CAM is risky. A Circadian network generates the cell-autonomous biological clock, but many genes with rhythmical expression profiles may be the output, not the core components.

Transcription factor with rhythmic diel expression profile do not always reach the targets in CAM systems. Transcription factor from HB-BELL, HB-PHD, HRT, and Tify families had been previously shown to have rhythmical diel expression pattern, however our study suggested that they do not reach targets within CAM 4 subnetworks (Moseley *et al.* 2018 and Table 6 of this chapter). Whereas the transcription factors from AP2/ERF-ERF, bHLH, C3H, GRAS,HSF, MYB, and Trihelix families have rhythmical members and reach targets across multiple subnetworks. Within the circadian subnetwork, HY5 and LUX have the lowest number of transcription factors reaching them as compared to other target genes in our circadian subnetwork.

However, the set of target genes identified in the circadian subnetwork should be viewed with precautions. This is because such '*Circadian genes*' of *K. fedtschenkoi* have been identified from the characteristics of gene expression profiles as opposed to the functional validation experiments in *Arabidopsis thaliana* (Moseley *et al.* 2018). To be more specific, these *K. fedtschenkoi* '*Circadian genes*' are the set of orthologues of *A. thaliana* circadian genes whose gene expression patterns met 3 selective criteria of the theoretical clock-gene behaviour: absolute amplitude > 10;fold-changes > 2; and JTK rhythmicity $p \leq 0.05$. This restriction within the set of *A. thaliana* circadian orthologues could have pre-excluded unique circadian gene of *K. fedtschenkoi.* Alternatively, there is a potential for false positive identifications whereby the orthologues may retain the circadian characteristic in their expression profiles but are not functioning as part of the central circadian oscillators. Evidence in pineapple showed that there are approximately 42% of transcription factors and 45% of transcription co-regulators displaying diel rhythmic expressions (Sharma *et al.* 2017). This emphasises the potential existence of CAM-specific circadian genes at least in the sugar-storage monocot branch of CAM. Therefore, the circadian subnetwork results must be taken forward with caution and updated with additional, true circadian genes which have been identified and validated.

In addition to the post-translational controls, the time-series transcriptome data also showed that these target genes (PEPC and PPDK) and their respective regulators (PPCK and PPDK-RP) all show diel expression patterns with peak expression levels at a specific time of the day. The expression of PPCK at night explains the nocturnal PEPC phosphorylation, whereas the bifunctionality of PPDK may require a more obscure explanation. Our study has advanced a step further by identifying the candidate transcription factors controlling the PEPC, PPCK, PPDK, and PPDK-RP (Figure 8-10). Overall, it is possible that transcriptional control would be operating alongside post-translational controls. Firstly, PPCK and PPDK-RP are known to be under transcriptional control. Subsequently, the PPCK and PPDK-RP exert their post-translational control on their targets: PEPC and PPDK respectively. Finally, PEPC and PPDK are under both transcriptional control and post-translational control to ensure that night-time carboxylation and day-time decarboxylation are tightly regulated to prevent futile cycling. This multitude of controls can be pursued further in the context of the CAM spectrum, reflecting the plasticity in terms of strength and duration of the four phases of CAM (Griffiths *et al.* 2008).

Furthermore, non-coding RNAs may be involved with CAM function. The evidence in pineapple (*Ananas comosus*) has suggested that long non-coding RNAs may compete with PPCK and PEPC genes for being transcribed in the photosynthetic tissue during the daytime, hence preventing these nocturnal CAM enzymes from being transcribed during the day (Bai *et al.* 2019). A separate study also identified 20% of pineapple microRNAs with diel expression pattern and the specific binding sequence to CAM genes including PPCK1 and MDH (Wai *et al.* 2017). The equivalent study of non-coding RNAs has not been undertaken in the Kalanchoe genus.

Discussion point 5: Comparing transcriptional regulation of stomatal-function genes between our network and the Local Edge Machine (LEM) algorithm

A different study has built a gene regulatory network for *K. fedtschenkoi* stomata with a Local Edge Machine (LEM) algorithm (Moseley *et al.* 2021). In that study, 1,605 stomatal genes were obtained from a text-mining approach, 17 of which were controlled by core clock TFs and had rescheduled expression pattern compared to their orthologues in *A. thaliana*. Within these 17 target genes, 3 of which are known to have gene ontology as calcium-dependent protein kinase 26 (CPK26), phototropin-2 (PHOT2), and mitogen-activated protein kinase homolog (MMK2); 5 genes were obtained from the *K. fedtschenkoi* transcriptome project (Yang *et al.*

2017), whereas the other 12 target genes in Moseley's stomatal subnetwork were obtained from a text mining approach. From the algorithmic point of view, Moseley's starting set of target genes are wider than our set of target genes because they not only included 20 genes from homology-based method from the *K. fedtschenkoi* transcriptome project, but also included 1,585 more genes from text-mining approach (Moseley *et al.* 2021; Yang *et al.* 2017). On the contrary, our set of potential regulators are wider than in Moseley's study because we allowed for any transcription factors and any gene across the genome, whereas Moseley restricted the regulators within the set of clock transcription factors.

Phototropin 2 (PHOT2; gene ID Kaladp0033s0113) was selected as gene of interest in both our study and Moseley et al 2020. The Phototropin 2 (PHOT2) has been in the spotlight for stomatal functioning in CAM since 2019 from the evidence of convergence change in the diel expression pattern found in two CAM species (*K. fedtschenkoi* and *Ananas comosus*) relative to the expression pattern in C3 *A. thaliana* (Moseley *et al.* 2018). The timing of peak transcript abundance in *K. fedtschenkoi* and *A. comosus* shifts from the pattern in *A. thaliana* by 11-hr and 9-hr respectively making the PHOT2 gene being most abundant at dusk in CAM, rather than at dawn in *A. thaliana*. The possibility of this PHOT2 participating in pathway that contributes to the inverted timing of the stomatal opening in CAM was later followed by the knockdown study in *K. fedtschenkoi*, the first Crispr/cas protocol paper for CAM (Liu *et al.* 2019). This experiment showed that the lack of PHOT2 resulted in higher stomatal conductance and CO_2 assimilation at night but lower stomatal conductance and CO_2 assimilation in the late afternoon (Phase IV) suggesting that PHOT2 is participating in the timing of stomatal opening in CAM. However, whether PHOT2 is functioning to promote stomatal opening both in the night (Phase I) and day (Phase IV) requires the direct comparisons of the transcript abundance and protein activity between wildtype and knockdown line at both timepoints, which were not reported from such a study (Liu *et al.* 2019).

Conclusions:

Our computational analysis has narrowed down candidate transcription factors governing CAM function in *Kalanchoe fedtschenkoi* from 3,048 transcription factors from 69 families down to exactly 21 transcription factors from 12 families, and the remaining transcription factors have been ranked in order of potential significance for regulating a certain target gene. In the long run, our analysis pipeline can be applied to other CAM species with annotated genome and time-series transcriptome data to allow cross-species comparisons of CAM evolution (Ming *et al.* 2015; Zheng *et al.* 2016; Yang et al., 2017; Wai and VanBuren, 2018; Yin *et al.* 2018; Heyduk *et al.* 2019; Abraham *et al.* 2020). Additionally, this GRN inference method can facilitate attempts at targeted transcriptional reprogramming, because the GRN finding has narrowed down the set of candidate regulators and opened up the opportunity to identify novel regulatory components in an unbiased manner from information derived across the entire genome (Amin *et al.* 2019). With computational prediction, a set of strong candidates has been narrowed down from the genome-wide data. Subsequently, this set of strong candidates could be validated experimentally with the new CRISPR/Cas9 transformation protocol developed specifically for CAM species (Liu *et al.* 2019). The experimental validation would distinguish the functional regulators from the pool of stochastic-binding transcription factors. Subsequently, the up-to-date knowledge on transcriptional factor identities can be fed back into the machine-learning model to develop a better prediction under a supervised method. It is clear that both the computation and the large-scale experimental validation should go hand in hand to advance the CAM physiology field. As a general rule, plant physiology research can advance faster with computational tools with the availability of the large scale -omics data, provided that the fundamental physiological questions are always kept in mind.

Acknowledgements:

I would like to thank Dr Gitanjali Yadav and Dr Manusnan Suriyalaksh for mentioning Gene regulatory network over some tea breaks. I would also like to acknowledge the contribution by Porames Vatanaprasan in reassembling the initial python codes. Without that attempt, the CAM GRN project would have had no starting point.

Chapter 4

Nuclei isolation followed by flow cytometry separation technique for *Kalanchoe fedtschenkoi*

Introduction:

The Gene Regulatory Network Inference in Chapter 3 have provided a list of candidate transcriptional regulators and their predicted target genes with the presence of specific *cis*-element within the promoter regions. Even though these predictions were based on the 24-hour timeseries transcriptome data, they do not represent the direct binding evidence of the transcriptional regulators on their targets nor the timing and condition that such binding may have taken place. Thus, it is essential to pursue the area of Gene Regulatory Network through the attempt to access the physical status of the chromatin and their associated proteins. The knowledge of the Gene Regulatory Network of CAM would be improved by the availability of pure nuclei and the intact binding evidence of the transcription factors and their associated *cis*-elements.

There are multiple options with different limitations for identifying open-chromatin region and the transcription factor binding regions within a genome. First, Chromatin Immunoprecipitation followed by Sequencing (ChIP-seq) relies on the immunoprecipitation of the antigen-antibodies specific for the proteins of interest which can be transcription factors, histones or enhancers, prior to sequencing. Only the chromatin regions associated with the proteins of interest are selectively pulled by the corresponding antibodies and sequenced. This technique was made possible because of the next-generation sequencing and the early studies which adopted this ChIP-seq technique were published in 2007 (Park 2009). However, ChIP-seq in CAM species is challenging because there are not widely available set of antibodies designed to be specific for CAM transcription factors. In general, the antibodies against most plant transcription factors are lacking, thus it requires an additional step of generating the epitope-tagged versions of proteins of interest from the transgenic lines (Lu *et al.* 2017).

Second, micrococcal nuclease digestion followed by high throughput sequencing (MNase-seq) relies on the micrococcal nuclease enzyme to cut the nucleosome-depleted region of the genome and returns the sequencing reads for the nucleosomal regions. This technique relies on laborious MNase enzyme titration and required 1-10 millions cells (Tsompana and Buck 2014). Third, DNase I hypersensitive sites sequencing (DNaseI-seq) uses the DNaseI enzyme to also cut the nucleosome-depleted region of the chromosome but returns the sequencing reads for the open regions with possibilities for the transcription factor binding evidence. DNase-seq also requires several enzyme titration and required 1-10 millions cells similar to MNase-seq (Tsompana and Buck 2014). Fourth, Formaldehyde-Assisted Isolation of Regulatory Elements

(FAIRE-seq), rely on the formaldehyde crosslinking and sonication to extract the nucleosomal-depleted regions. This technique requires 100,000 to 10 million cells and returns low signal-to-noise ratio which hinders the signal interpretation (Tsompana and Buck 2014).

Finally, Assay for Transposase Accessibility Chromatin Sequencing (ATAC-seq) uses Tn5 transposase enzyme which can access and cut accessible regions. The sequencing reads show the patterns of signal drop in the inaccessible regions indicating the evidence of protein binding. ATAC-seq is considered the most convenient method to date due to its concise protocol and its requirement of the smallest number of pure nuclei (500-50,000 nuclei) (Tsompana and Buck 2014). This ATAC-seq technique was adopted for *Arabidopsis thaliana* in 2018 (Bajic *et al.* 2018).

To obtain pure nuclei for any chromatin accessibility protocol, nuclei sorting with flow cytometer offers a promising protocol. The fluorescence-activated nuclei sorting prior to ATAC-sequencing (FAN-ATAC) has been adopted successfully for *A. thaliana* to improve the quality of nuclei purification (Lu *et al.* 2017). Nonetheless, the adopting of any C3 protocol for CAM plants could be challenging due to the distinct anatomical structure of water-storage tissue, mucus tissue, malic acid storage. In addition, the thick epidermal layer may influence the protocol from the physical liquid nitrogen grinding or interact with chemical of the C3-species extraction protocol. Nonetheless, fluorescence-based sorting is still the most promising tool to progress towards pure nuclei extraction for chromatin accessibility assay or other molecular protocols that require nuclei as starting material. Here, we combined the sucrose-gradient nuclei isolation method with the fluorescence-based flow cytometry technique to segregate the particle into different populations which would improve the purity of nuclei.

This chapter reported the absorption/emission patterns of particles in the nuclei extract obtained from sucrose-gradient nuclei isolation method. The absorption/emission patterns specific to DAPI signal were recorded as an indicator for the DNA-containing particles, whilst the red signal was recorded as an indicator of the chlorophyll fluorescence, and the forward scattering signal was an indicator of particle size. In this experiment, the population of particles that had DAPI signal but not chlorophyll fluorescence accounts for at least 44.2% of total particle counts (equivalent to 2,210 particles in total per flow cytometer run).

Material and methods:

Plant growth

Kalanchoe fedtschenkoi from the Cambridge botanical garden were grown and maintained in a glass house for 10 months under the natural light (with supplementary light to keep minimum radiance at 390-410 W/m^2 through 06.00 am-18.00 pm during the short-day period), 40% Relative humidity, daytime temperature of 18 °C, and nighttime temperature of 15 °C. Leaf samples were taken from leaf pair 5^{th} counting from the top starting from the first fully grown leaf pair and snap frozen in liquid nitrogen prior to the nuclei isolation process.

Nuclei isolation (Sucrose-gradient method)

Frozen leaf samples (~0.5 g) were ground in the ice-cold mortar with the liquid nitrogen until the tissue was broken into green homogeneous powder. The ground tissue was suspended in 45 mL extraction buffer (14.6 mM Pipes pH 6.5, 19.5 mM NaCl, 78 mM KCl, 0.975 mM EDTA pH 8.0, 0.293 M Sucrose, 0.05 %V/V β-mercaptoethanol, 2.5 mM Spermidine, 1% V/V Triton-X, 0.125 mM Polyvinylpyrrolidone PVP40 and 1mM Protease inhibitor) and left on ice for 5 minutes before filtering through Miracloth (pore diameter 22-25µm) twice. The filtered solutions were centrifuged at 4,000 rpm at 4 °C for 20 minutes. Subsequently, the pellets were resuspended in 1 mL of extraction buffer (see above) before centrifuged at 12,000 rpm at 4 °C for 10 minutes. The pellets were resuspended in 300 µL of extraction buffer (see above) before layering on top of the clean 300 µL of extraction buffer (see above) in a different Eppendorf tube and centrifuged at 16,000 rpm at 4 °C for 1 hour. Finally, the pellets containing isolated nuclei were resuspended in 1 mL of extraction buffer (see above) to obtain the final product from the nuclei extraction steps.

DAPI (6-Diamidino-2-phenylindol) staining

The 10X DAPI solution (equivalent to 7.2 mM) was prepared by dissolving 10 mg DAPI in 5 mL deionised water. The 10x DAPI solutions were kept in the dark at 4 °C before diluted tenfold into 1x DAPI solution for staining the nuclei extract with the proportion of 100 µL of 1x DAPI in 1 mL of nuclei extract.

Flow cytometer

The nuclei extracts were passed into the Aurora Cytek machine (model: N7-00000-0A, serial number: R0069) with the maximum flow rate of 15 µL/min and total count of 5,000 particle

per run. The instrument were set to detect the DAPI signal (UV absorption with peak emission at 458 nm and emission range 428-720 nm). The laser and detection channels involved being used in this experiment were ultraviolet laser (Excitation 355 nm, Power 20 mW) with 16 UV channel detectors (UV4 channel with centre at 433 nm, bandwidth 15 nm, wavelength range 436-451 nm) and 8 red channel detector (R3 channel with centre at 697 nm, bandwidth 19 nm, wavelength range 688-707 nm). The snap shot were taken after the flow rate was stabilised for 30 second and as triplicates for each sample. The absorbance profiles of all particles being detected by the flow cytometer were recorded in the UV, red, forward scatter (FSC) and side scatter (SSC) channels.

*Notably, the Aurora cytek flow cytometer is able to measure proportion of each subpopulation with their unique absorption/emission spectrum but does not allow the retrieval of subpopulation afterwards. This is different from cell sorter machine which allow the collection of one or more subpopulation from the pool after the proportion of each subpopulation has been determined despite both using the similar flow cytometry approach.

Results:

The absorption/emission patterns of the samples with specific to DAPI channel were recorded through the UV4-A and R3-A channels. The absorption/emission patterns were different between the unstained sample and DAPI-stained sample (Figure 1). This indicated that the nuclei extraction indeed contained the DNA-containing particles (i.e. nuclei, mitochondria and chloroplast). The maximum emission of the unstained sample was in the range of 10-10^3 with the background emission in the range of 10^3-10^5. In contrast, the emission of the DAPI-stained sample appeared as a single peak in the range of 10^3-10^5.

Population-1 was selected based on the absorption/emission patterns that appeared in the DAPI-stained set but not the unstained set , thus are most likely to be nuclei. Notably, the population-1 did not include the particles with red-wavelength emissions of more than 10^3. This is because the Channel R3-A (red channel) was selected to monitor the chlorophyll fluorescence of chloroplast particles that may have been contaminated in the nuclei extracts. The separation of the subpopulations according to the R3-A channel added another level of nuclei purification by excluding the particles that show red-emission (R3-A) despite having been positively stained with the DAPI and also returned blue-emission spectrum (UV4-A).

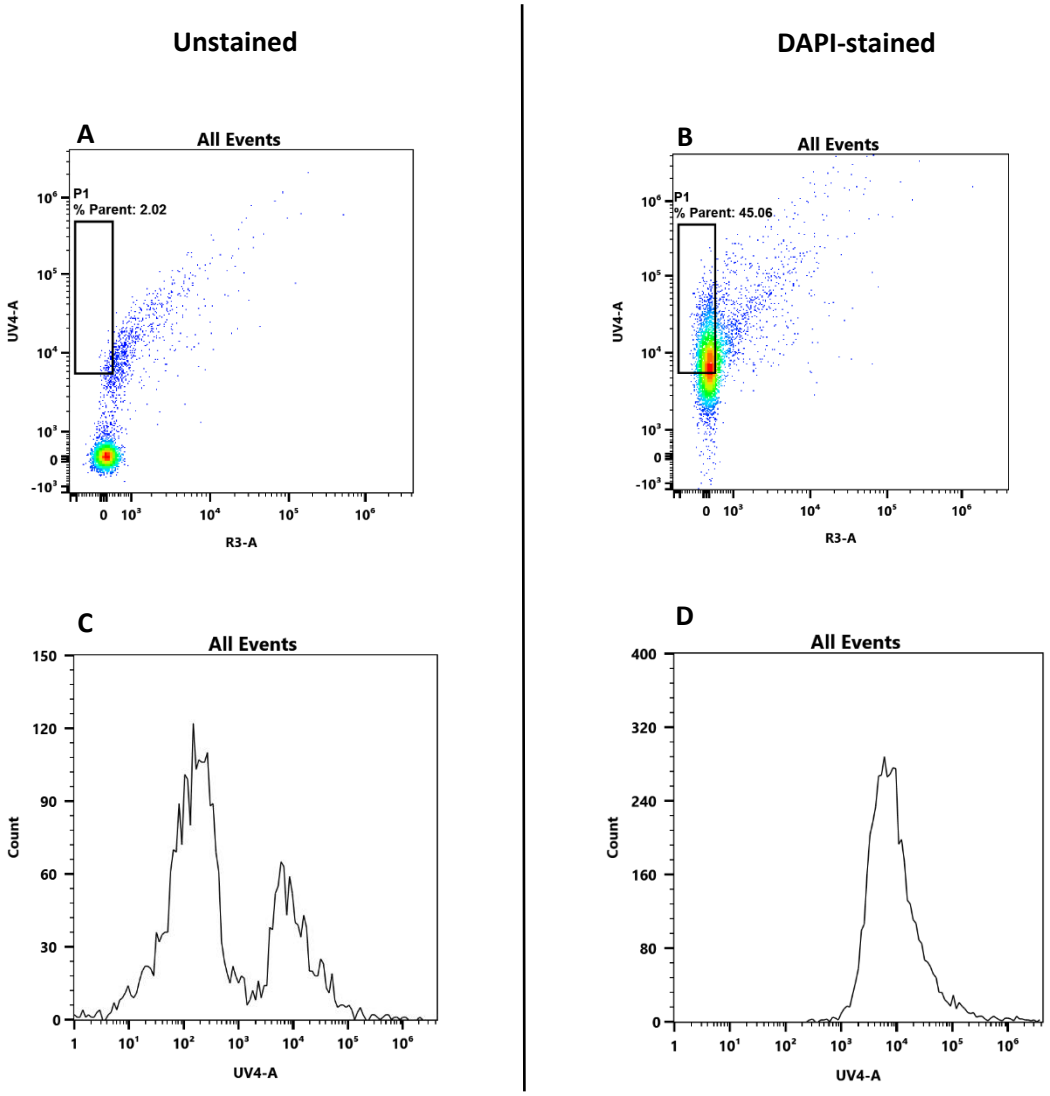


Figure 1: The comparison of flow cytometry signal of unstained and DAPI-stained nuclei extractions from *Kalanchoe fedtschenkoi* (biological replicate 1, technical replicate 1) (A-B)The absorption/emission patterns of the samples with specific to DAPI channel were recorded through the UV4-A and R3-A channels for blue light (wavelength range 436-451 nm) and red light (wavelength range 688-707 nm), respectively. **(C-D)** Particle count at each UV4-A signal were plotted to show double peaks in the unstained sample and single peak in the DAPI-stained sample.

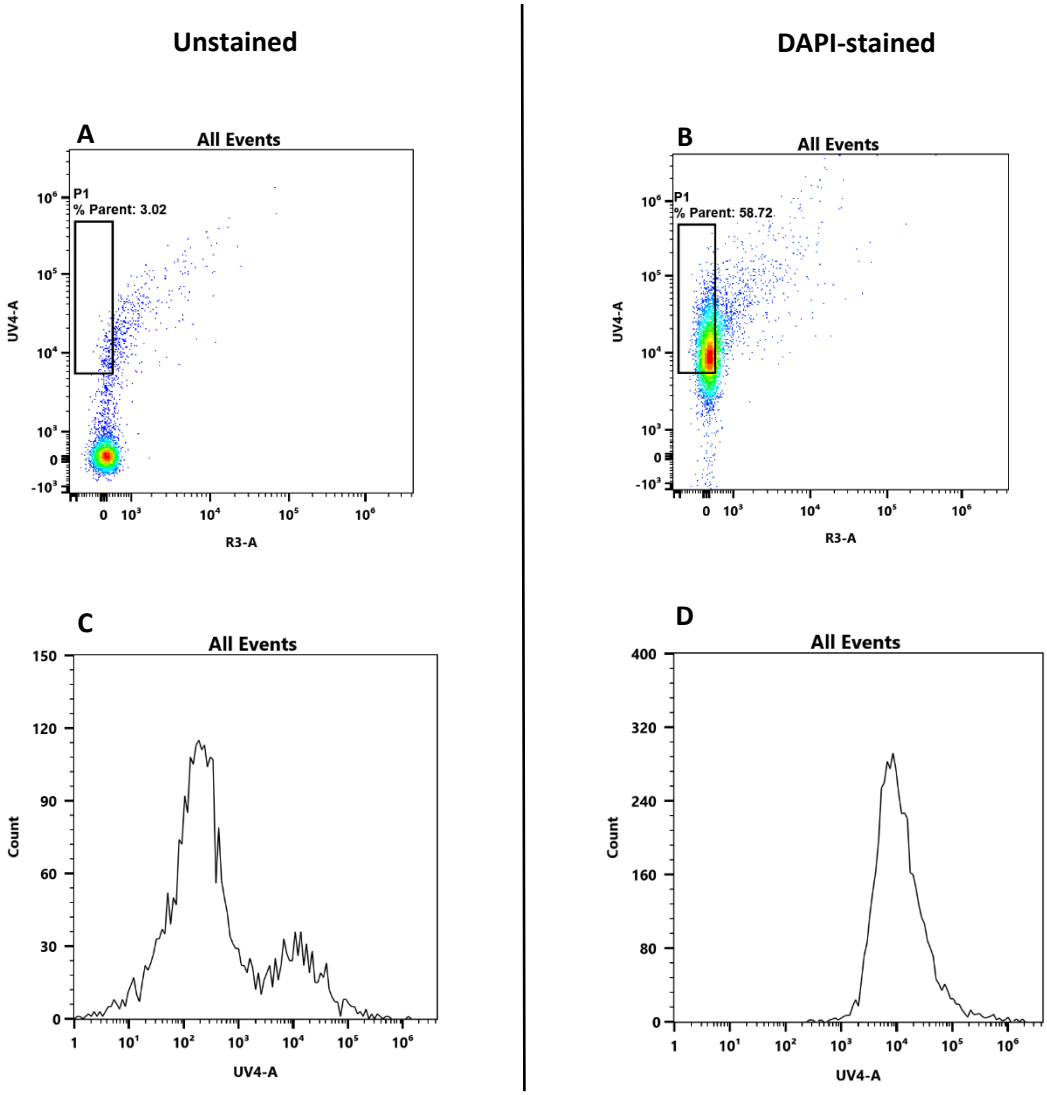


Figure 2: The comparison of flow cytometry signal of unstained and DAPI-stained nuclei extractions from *Kalanchoe fedtschenkoi* (biological replicate 1, technical replicate 2) (A-B)The absorption/emission patterns of the samples with specific to DAPI channel were recorded through the UV4-A and R3-A channels for blue light (wavelength range 436-451 nm) and red light (wavelength range 688-707 nm), respectively. **(C-D)** Particle count at each UV4-A signal were plotted to show double peaks in the unstained sample and single peak in the DAPI-stained sample.

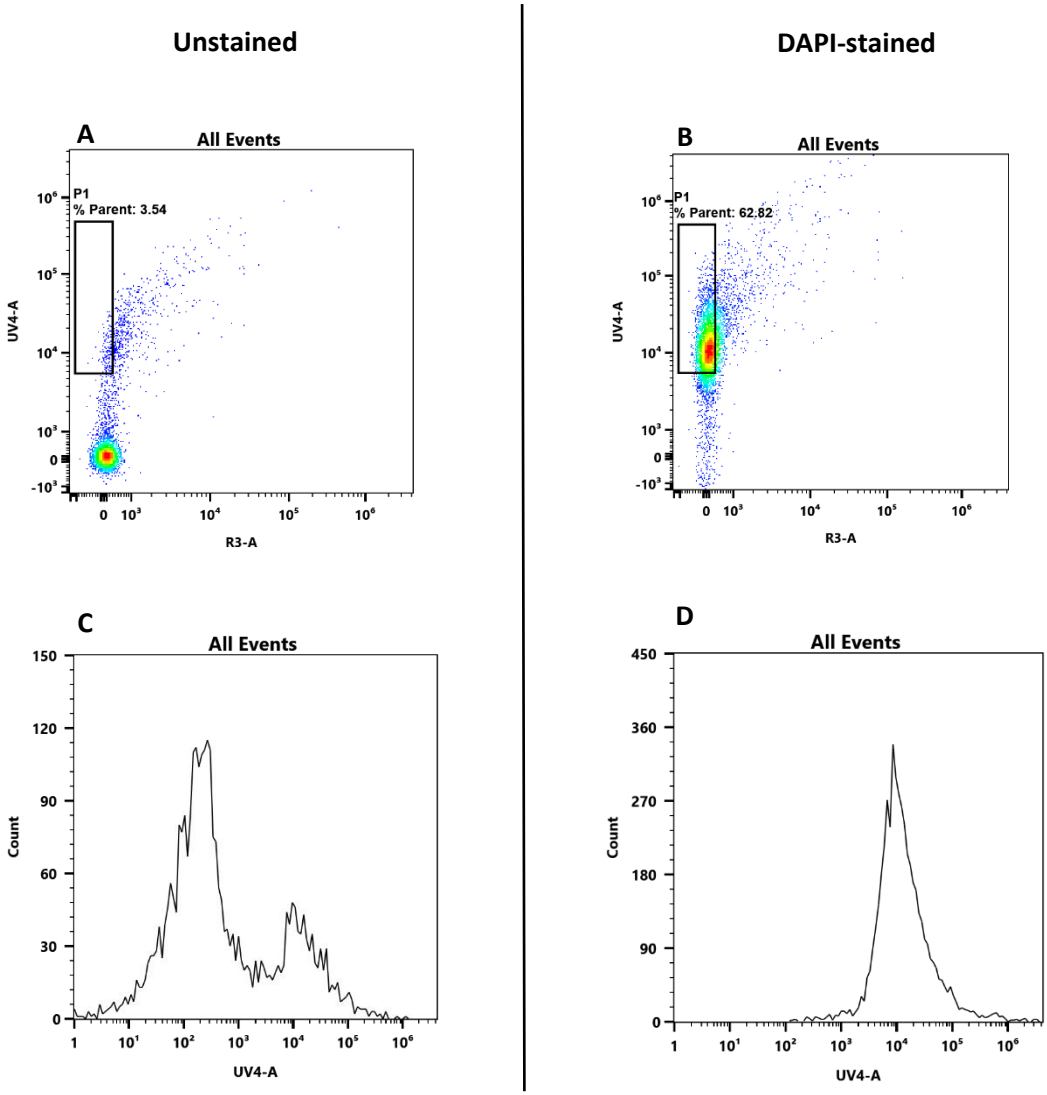


Figure 3: The comparison of flow cytometry signal of unstained and DAPI-stained nuclei extractions from *Kalanchoe fedtschenkoi* (biological replicate 1, technical replicate 3) (A-B)The absorption/emission patterns of the samples with specific to DAPI channel were recorded through the UV4-A and R3-A channels for blue light (wavelength range 436-451 nm) and red light (wavelength range 688-707 nm), respectively. **(C-D)** Particle count at each UV4-A signal were plotted to show double peaks in the unstained sample and single peak in the DAPI-stained sample.

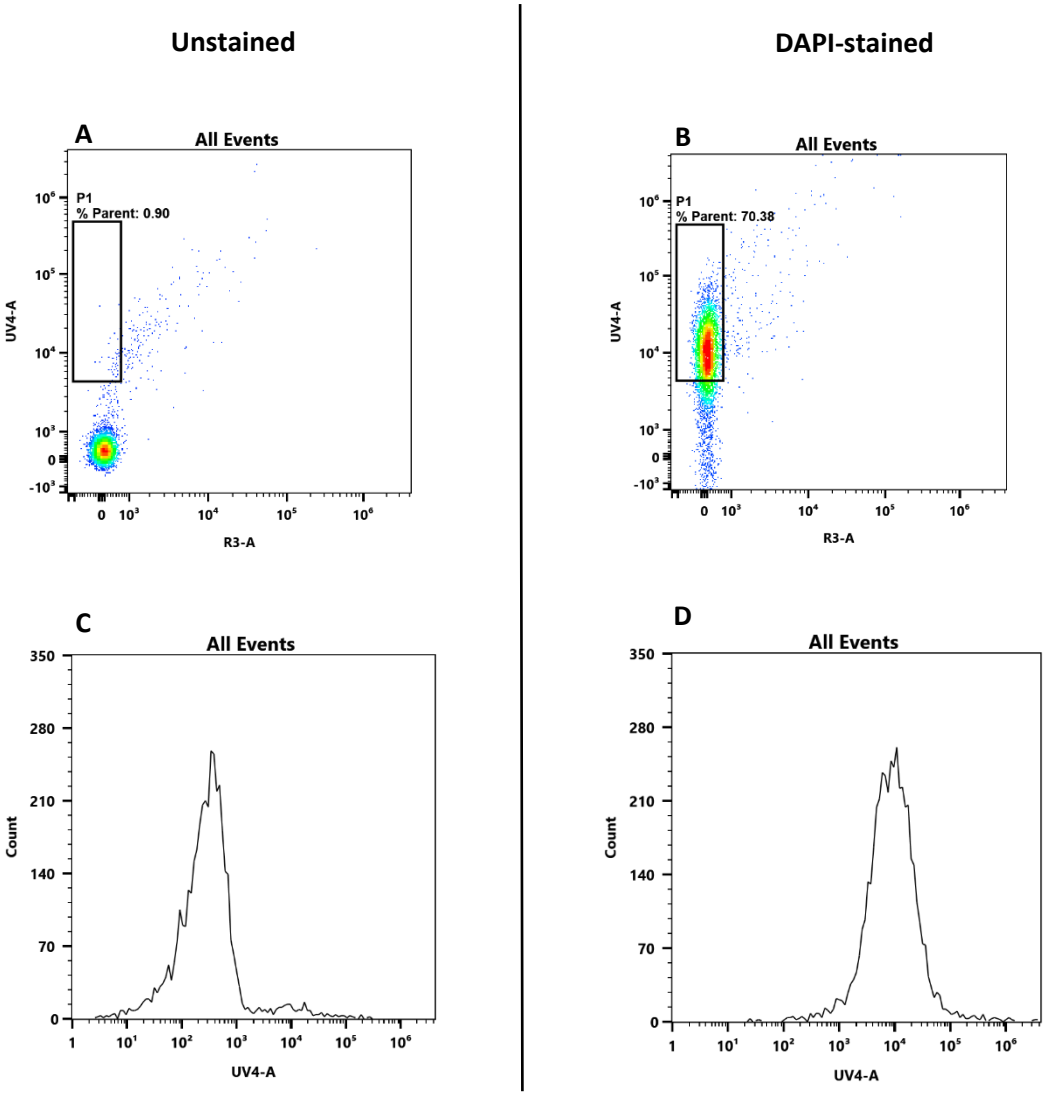


Figure 4: The comparison of flow cytometry signal of unstained and DAPI-stained nuclei extractions from *Kalanchoe fedtschenkoi* (biological replicate 2, technical replicate 1) (A-B)The absorption/emission patterns of the samples with specific to DAPI channel were recorded through the UV4-A and R3-A channels for blue light (wavelength range 436-451 nm) and red light (wavelength range 688-707 nm), respectively. **(C-D)** Particle count at each UV4-A signal were plotted to show double peaks in the unstained sample and single peak in the DAPI-stained sample.

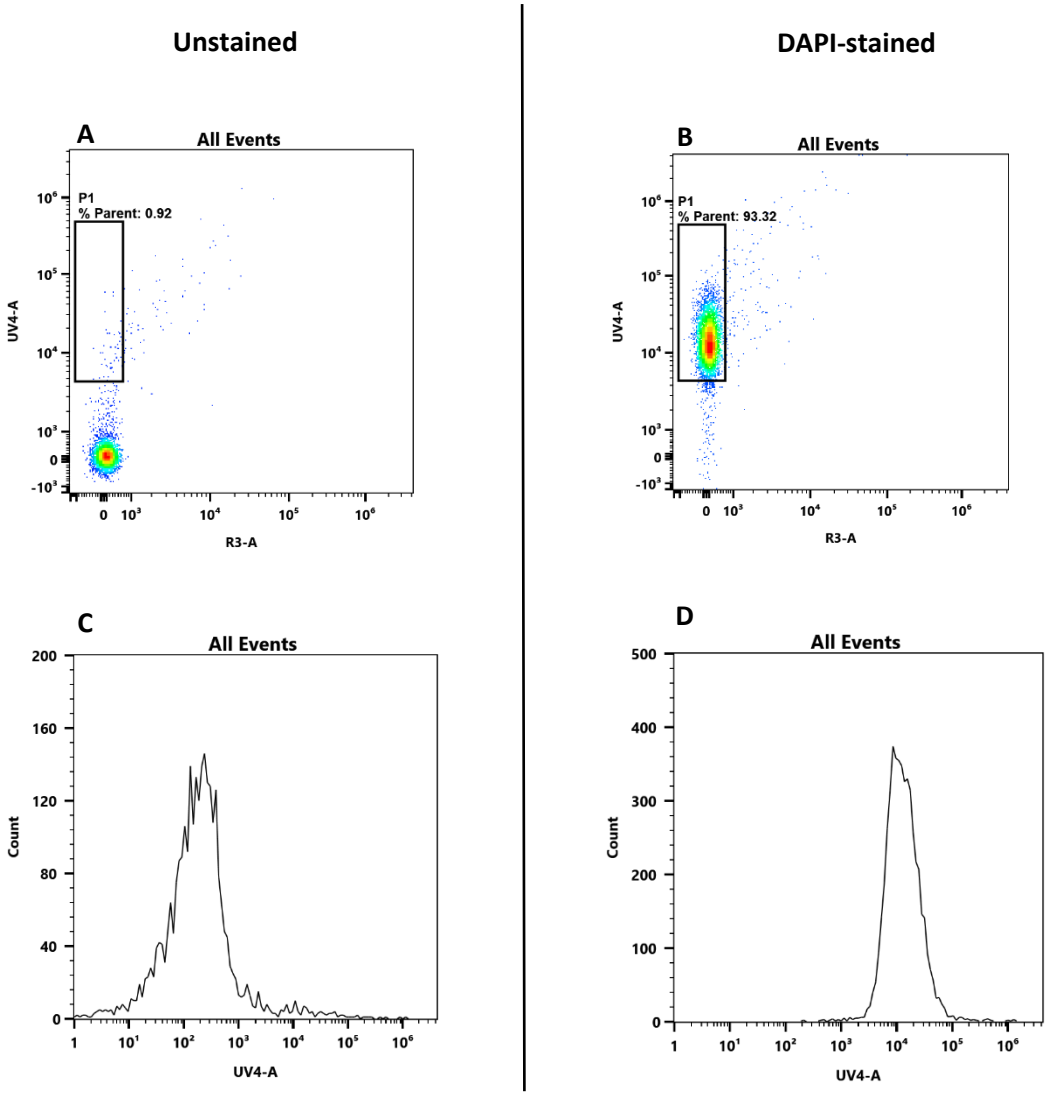


Figure 5: The comparison of flow cytometry signal of unstained and DAPI-stained nuclei extractions from *Kalanchoe fedtschenkoi* (biological replicate 2, technical replicate 2) (A-B)The absorption/emission patterns of the samples with specific to DAPI channel were recorded through the UV4-A and R3-A channels for blue light (wavelength range 436-451 nm) and red light (wavelength range 688-707 nm), respectively. **(C-D)** Particle count at each UV4-A signal were plotted to show double peaks in the unstained sample and single peak in the DAPI-stained sample.

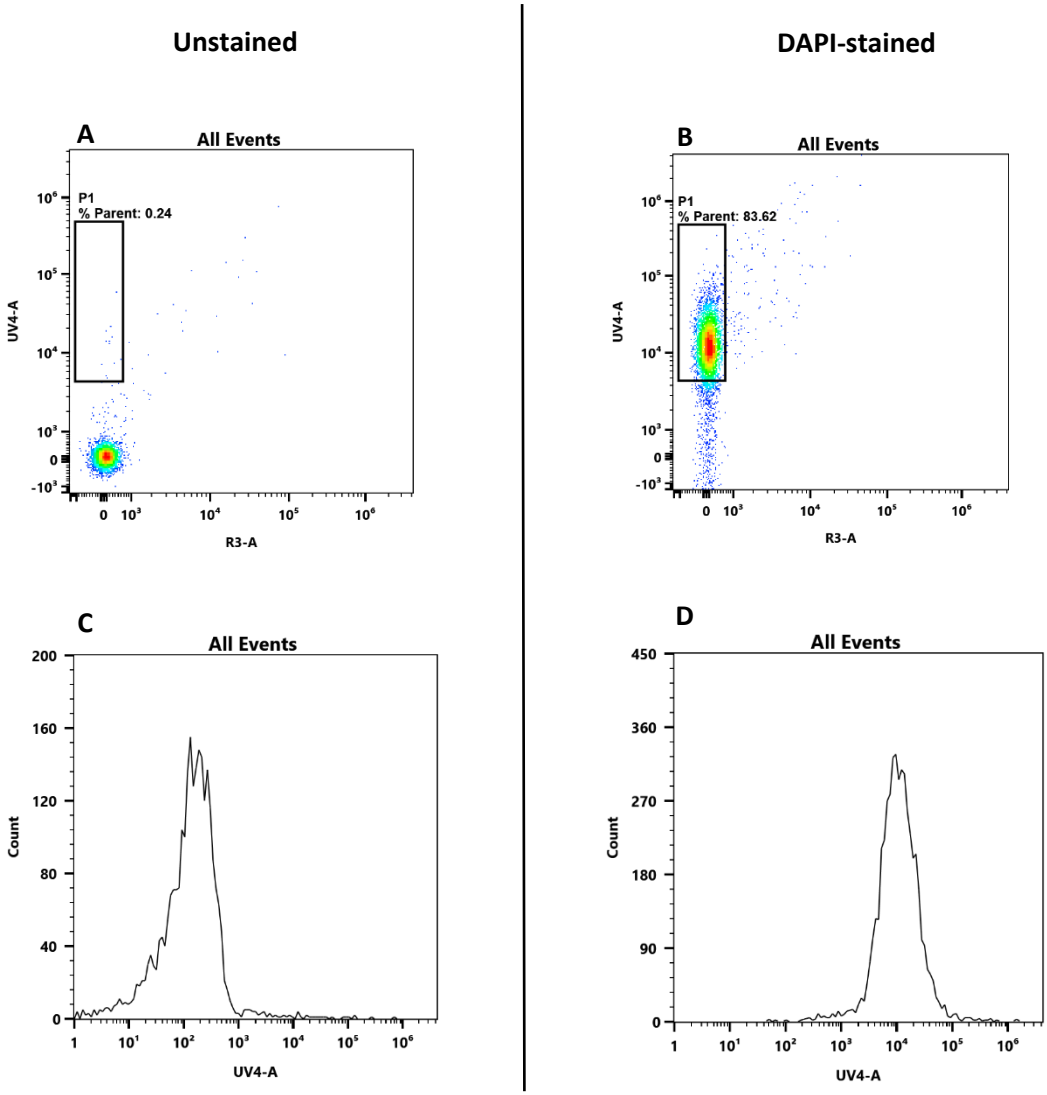


Figure 6: The comparison of flow cytometry signal of unstained and DAPI-stained nuclei extractions from *Kalanchoe fedtschenkoi* (biological replicate 2, technical replicate 3) (A-B)The absorption/emission patterns of the samples with specific to DAPI channel were recorded through the UV4-A and R3-A channels for blue light (wavelength range 436-451 nm) and red light (wavelength range 688-707 nm), respectively. **(C-D)** Particle count at each UV4-A signal were plotted to show double peaks in the unstained sample and single peak in the DAPI-stained sample.

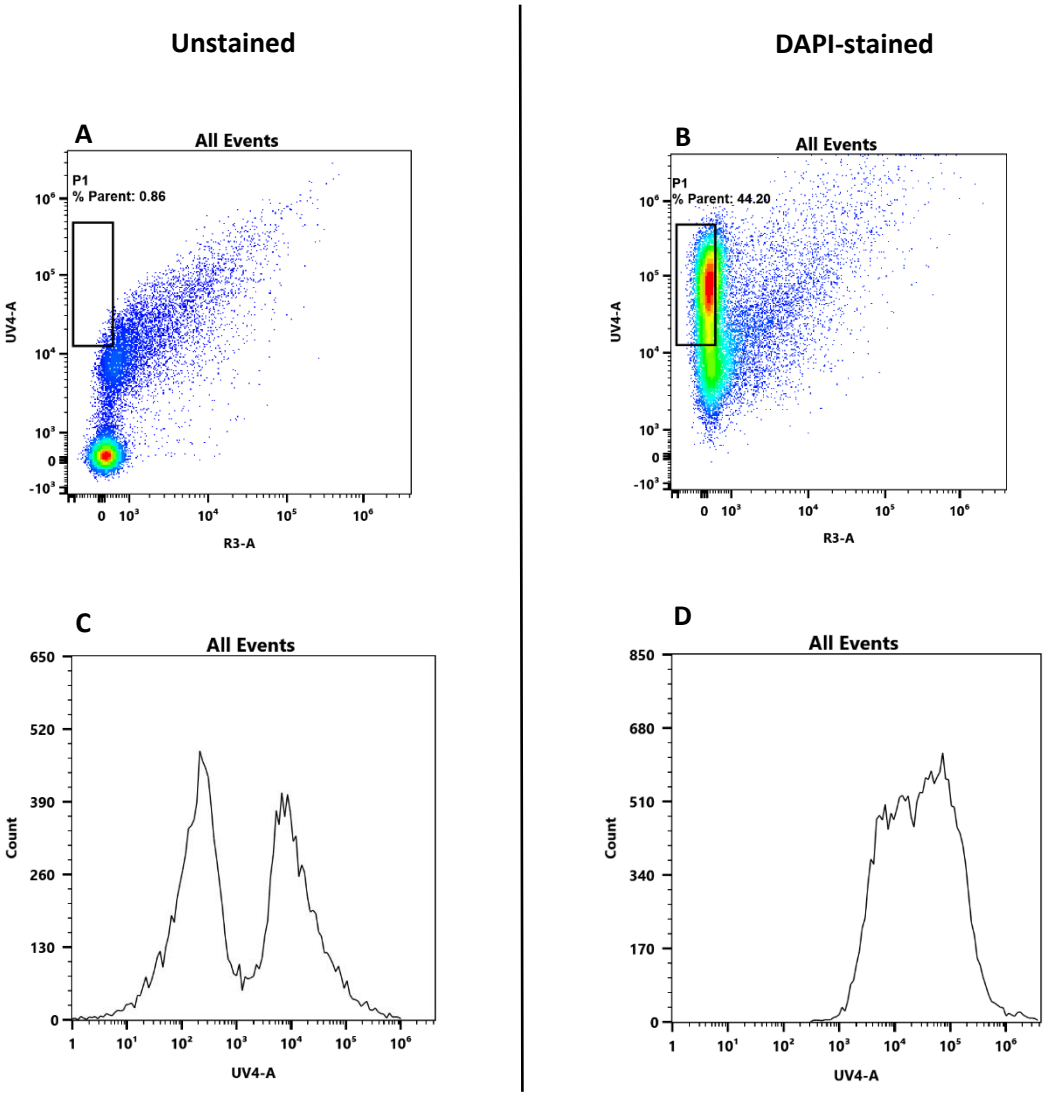
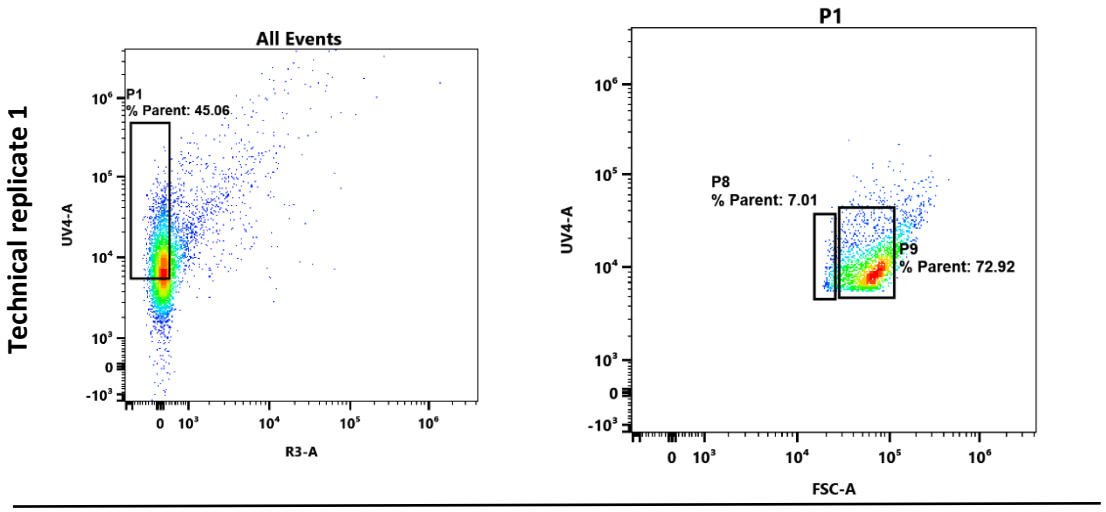
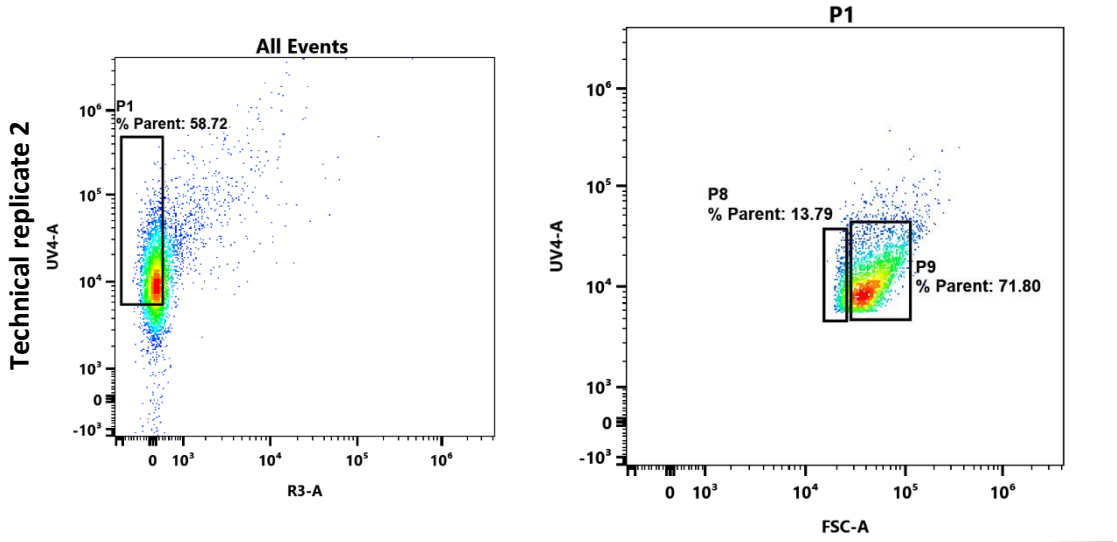


Figure 7: The comparison of flow cytometry signal of unstained and DAPI-stained nuclei extractions from *Kalanchoe fedtschenkoi* (biological replicate 3, technical replicate 1) (A-B)The absorption/emission patterns of the samples with specific to DAPI channel were recorded through the UV4-A and R3-A channels for blue light (wavelength range 436-451 nm) and red light (wavelength range 688-707 nm), respectively. **(C-D)** Particle count at each UV4-A signal were plotted to show double peaks in the unstained sample and single peak in the DAPI-stained sample.

Although the population-1 has been selected to exclude chloroplast, the population-1 may still contain mitochondrial contamination instead of pure nuclei which can be expected from the sucrose-gradient nuclei isolation method. To alleviate this problem with the flow-cytometry approach, the DAPI absorption/emission was plotted against the forward scattering signal (FSC-A) because the DAPI signal indicated the DNA-containing particle whereas the forward scattering signal was positively correlated with the size of the particles being detected. Therefore, the population-1 can be further subdivided into population-8 and population-9 according to the forward scattering signal (FSC-A). Population-8 represented the DNA-containing particles with smaller size than the population-9, thus the population-9 is expected to contained the highest proportion of nuclei and should be selected for further molecular experiments. The similar flow cytometry approach can be repeated in the cell sorter machine prior to collecting the population-9 to be visualised under fluorescence microscope to verify the integrity and identity of nuclei.


Technical replicate 1
All Events
P1
% Parent: 45.06
UV4-A
R3-A
P1
P8
% Parent: 7.01
P9
% Parent: 72.92
UV4-A
FSC-A


Technical replicate 2
All Events
P1
% Parent: 58.72
UV4-A
R3-A
P1
P8
% Parent: 13.79
P9
% Parent: 71.80
UV4-A
FSC-A

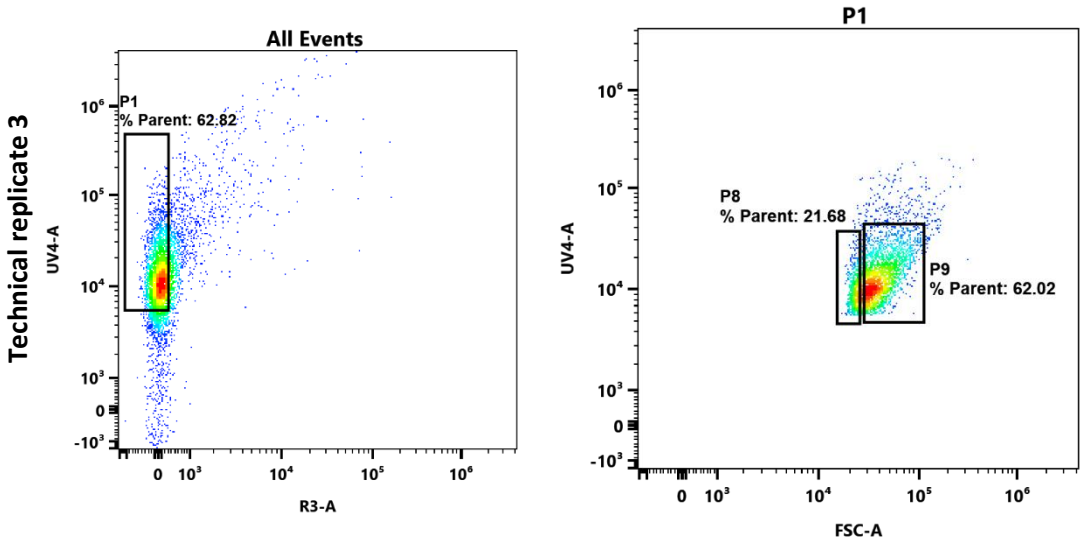

Technical replicate 3
All Events
P1
% Parent: 62.82
UV4-A
R3-A
P1
P8
% Parent: 21.68
P9
% Parent: 62.02
UV4-A
FSC-A

Figure 8: Population-1 were further categorised according to particle size indicated by forward scattering signal (FSC-A) (biological replicate 1)

(A) Population-1 was selected based on the absorption/emission patterns specific to DAPI channel which were recorded through the UV4-A and R3-A channels for blue light (wavelength range 436-451 nm) and red light (wavelength range 688-707 nm), respectively. **(B)** Population-1 was divided into population-8 and population-9 based on FSC-A signal showing the population-8 with smaller particle size and population-9 with larger particle size, while both populations show blue light emission but not red light emission. The data was obtained from *K. fedtschenkoi* biological replicate 1.

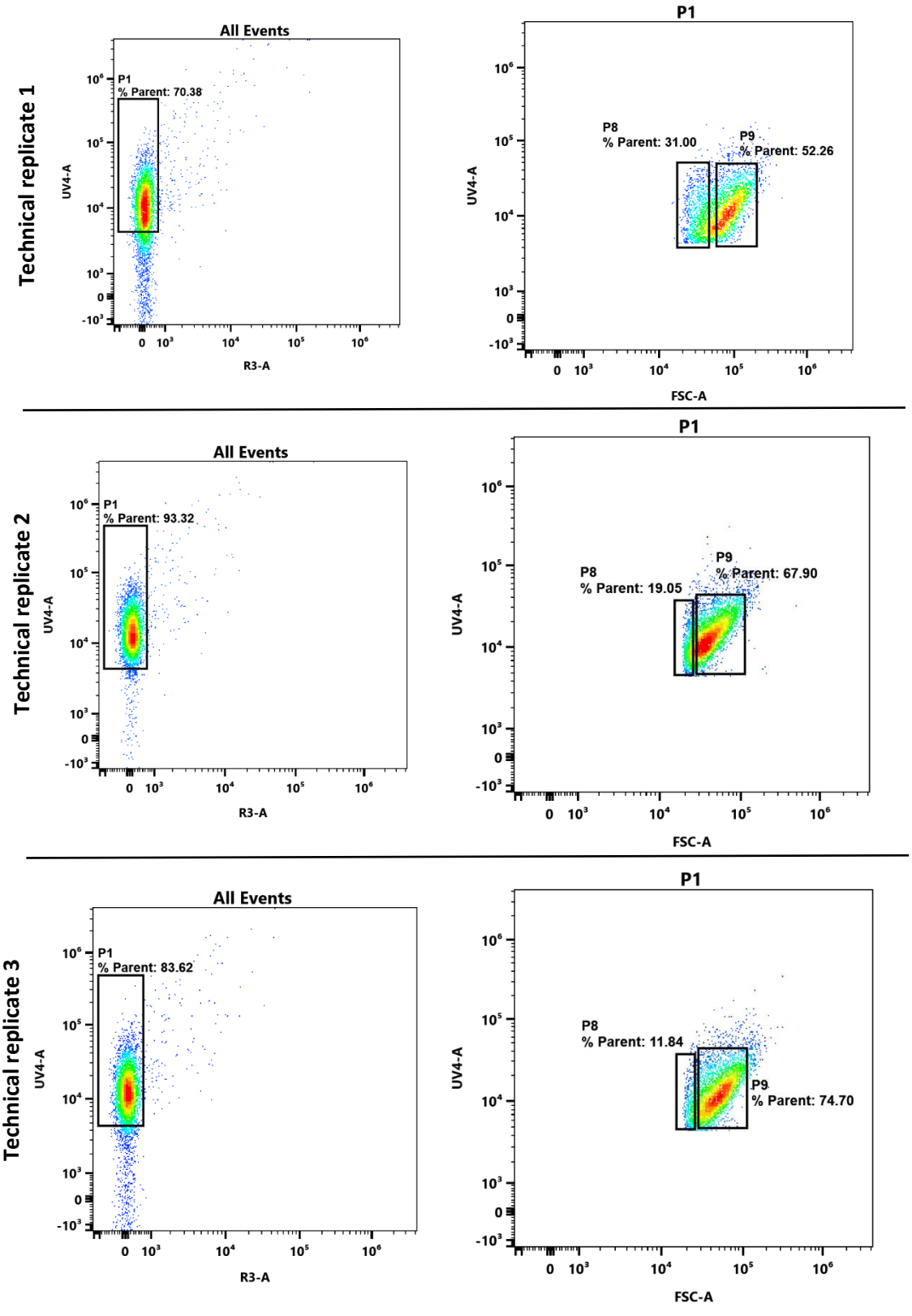

Technical replicate 1
All Events
P1
% Parent: 70.38
UV4-A
R3-A
P1
P8
% Parent: 31.00
P9
% Parent: 52.26
UV4-A
FSC-A
Technical replicate 2
All Events
P1
% Parent: 93.32
UV4-A
R3-A
P1
P8
% Parent: 19.05
P9
% Parent: 67.90
UV4-A
FSC-A
Technical replicate 3
All Events
P1
% Parent: 83.62
UV4-A
R3-A
P1
P8
% Parent: 11.84
P9
% Parent: 74.70
UV4-A
FSC-A

Figure 9: Population-1 were further categorised according to particle size indicated by forward scattering signal (FSC-A) (biological replicate 2)

(A) Population-1 was selected based on the absorption/emission patterns specific to DAPI channel which were recorded through the UV4-A and R3-A channels for blue light (wavelength range 436-451 nm) and red light (wavelength range 688-707 nm), respectively. **(B)** Population-1 was divided into population-8 and population-9 based on FSC-A signal showing the population-8 with smaller particle size and population-9 with larger particle size, while both populations show blue light emission but not red light emission. The data was obtained from *K. fedtschenkoi* biological replicate 2.

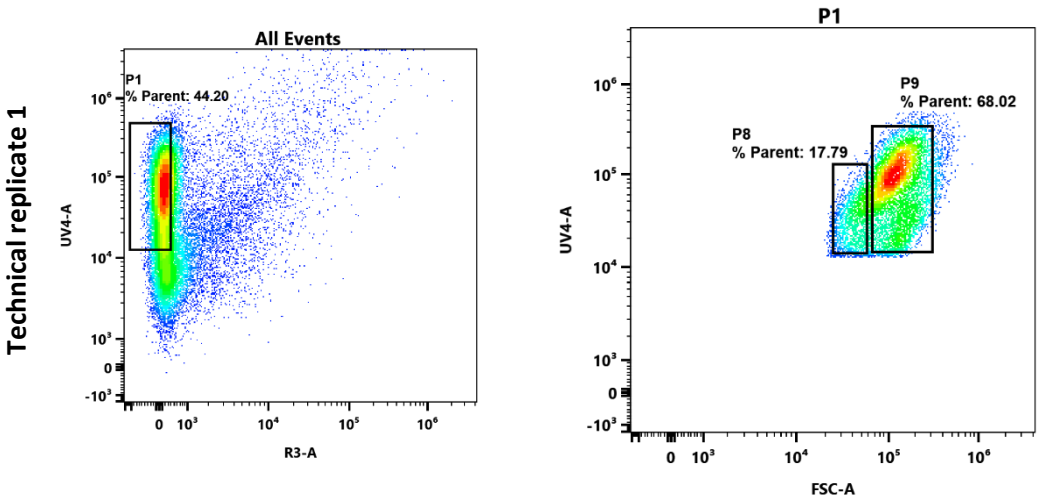


Figure 10: Population-1 were further categorised according to particle size indicated by forward scattering signal (FSC-A) (biological replicate 3)

(A) Population-1 was selected based on the absorption/emission patterns specific to DAPI channel which were recorded through the UV4-A and R3-A channels for blue light (wavelength range 436-451 nm) and red light (wavelength range 688-707 nm), respectively. **(B)** Population-1 was divided into population-8 and population-9 based on FSC-A signal showing the population-8 with smaller particle size and population-9 with larger particle size, while both populations show blue light emission but not red light emission. The data was obtained from *K. fedtschenkoi* biological replicate 3.

Discussion:

The flow cytometer successfully allowed for separation of particles in the nuclei extract of *K. fedtschenkoi* according to their absorption/emission wavelength patterns. This would allow for an improved method for selecting pure nuclei despite the sample being prepared from the sucrose-gradient nuclei isolation method. The combination of sucrose-gradient and flow cytometer can thus circumvent the time and resource limitation in generating transgenic INTACT line in the CAM model species, whereas the purity of nuclei isolation can be improved according to their absorption/emission wavelength patterns.

The fluorescent dye 4'-6-Diamidino-2-phenylindole (DAPI) was used in staining DNA-containing particles in the nuclei extracts because the dye has high specificity for DNA. The DAPI-bound DNA has maximum excitation with the UV light of 358 nm, while the maximum emission is in the blue light range with maximum emission of 461 nm (Karg and Golic 2018). The choice of the dye was a topic of critical decision because it would be preferable to use the dye that do not interfere with the accessibility of DNA for the downstream ATAC-sequencing protocol. Literature review supported the mechanism of DAPI binding to the minor groove area of the DNA with preferential binding to the AT-rich region as opposed to intercalating between the base-pair ladder (Tanious *et al.* 1992; Pineda De Castro and Zacharias 2002). As a result, DAPI was considered permissible for nuclei preparation protocol for ATAC-sequencing because the non-intercalating property whilst the AT-rich signal may be selectively subtracted from the sequencing reads afterwards. In addition, it is common to use DAPI in the nuclei preparation protocol for ATAC-sequencing in other plants (Lu *et al.* 2017; Bajic *et al.* 2018).

The non-zero background emission in the unstained sample indicated the DAPI-like emission without DAPI staining. This background signal is detected in all technical replicates, thus not indicating the cross-contamination in a single preparation batch. Therefore, this signal must represent non-DNA-containing particles within the nuclei extracts which can be accessory pigments. Although the reading overlap with the DAPI-stained particle, the presence of these non-DNA-containing particles would not compete with nuclei for Tn5 reaction, hence the selection of particle from this emission range should not be problematic for the ATAC protocol. In the ATAC study by Liu, the authors also select the corresponding region with maximum population size for the downstream Tn5 reaction with the subsequently successful ATAC-sequencing signals.

Although the chloroplast and mitochondria should have been separated from the nuclei according to their different densities through the centrifugation process, there are possibilities that the chloroplast and mitochondria are contaminated in the final nuclei extracts depending on the density of these organelles in each species. However, the flow cytometry technique allowed for chloroplast population to be visualised due to their chlorophyll fluorescence property (Ni *et al.* 2019). In contrast, mitochondria do not have the inherent fluorescence property but they can be separated from nuclei due to their smaller size which is positively correlated to the forward scattering signal. Moreover, if the information regarding the granularity of these organelles and nuclei in CAM has emerged, the side scattering channel can be used in separating these subpopulations (Cossarizza *et al.* 2017). Finally, similar process as outlined in this experiment can be performed in the cell sorter machine. The selected population can be collected and subsequently visualised under the fluorescence microscope to check the morphology and integrity of nuclei as well as verifying the percentage of mitochondrial and chloroplast in each subpopulation.

Conclusions:

Fluorescence-based flow cytometry with Cytek aurora instrument improved the purity of the nuclei extracts obtained from the sucrose-gradient method. The DAPI staining was the front line for selecting DNA-containing elements. Although the sucrose-gradient method should have minimised the contamination of chloroplasts and mitochondria in the final nuclei extracts due to the different densities, the flow cytometry technique can ensure that these contaminants are separated from the nuclei samples. Red autofluorescence signal and forward scattering signal are the two main axes for separating chloroplasts and mitochondria from the nuclei, respectively.

The same flow cytometry setting can be repeated with the cell sorter instrument. The cell sorter instrument would enable the collection of particles from each fluorescent population according to their absorption/emission profiles. Subsequently, the collected samples can be visualised under a fluorescence microscope or processed with molecular protocols. Overall, the combination of sucrose-gradient nuclei isolation and fluorescence-based flow cytometry technique would improve the purity of final nuclei samples. This is suitable for ATAC-seq preparation as the higher purity of nuclei extracts would reduce the signal loss through the Tn5 enzyme cross-reaction with mitochondrial and chloroplast DNA.

Acknowledgement:

I cannot thank Dr Leonie Luginbuehl and Dr Pallavi Singh enough for sharing the nuclei extraction protocols, their availability for further discussion, and all-time physical and moral support. For the flow cytometry part, I am expressing gratitude to Joana Cerveira and Mercedes Jarana from the flow cytometry facility at the Department of Pathology for introducing me to the flow cytometry concept and assisting throughout.

Chapter 5

General conclusions and recommendations

Summary:

The main theme of this thesis was to explore the regulatory landscape of CAM using systems biology approaches. The scope of the regulatory landscape was drawn around the mesophyll metabolism of the dicotyledonous obligate CAM genus, *Kalanchoe,* even though the computational frameworks that had been developed here can also be applied to other species from different CAM clades.

Firstly, Ordinary Differential Equation modelling of Crassulacean acid metabolism (Chapter 2) was developed to capture the classical gas-exchange patterns as well as the responses to perturbation conditions. The matching between the modelling output and experimental data for the wild-type unperturbed plants suggested that the minimal version of the mechanistic model was sufficient to explain the classical gas-exchange pattern of CAM. The matching between the modelling output and experimental data for the CO_2-perturbation condition suggested that the model can represent the biological plant under such a physical perturbation scenario. In contrast, the discrepancies between the modelling output and the experimental data for the Phosphoenolpyruvate Carboxylase (PEPC) knockdown scenario at the start of the daytime suggested that the knocking down the *pepc* gene may have interfered with other metabolic components, hence the standard model would require more refined parameterisation. These findings from Chapter 2 encouraged me to explore the higher level of CAM regulation that acted upstream of the protein content and activity, and controlled physiological expression. This led to the interest in the Gene Regulatory Network of CAM which was explored in the next chapter.

Secondly, the Gene Regulatory Network Inference of *Kalanchoe fedtschenkoi* (Chapter 3) identifies potential transcriptional regulators on different functional compartments of CAM including the following: Carboxylation subnetwork, Decarboxylation subnetwork, Circadian subnetwork and Stomatal subnetwork. The dynGENIE3 algorithm ranked candidate regulators for each target gene from the most to least likely candidates according to the patterns of transcriptome data collected over a 24-hour period. These candidates were subsequently checked for the presence of the corresponding *cis*-elements within the promoter regions of their target genes using the FIMO tool. This chapter highlighted the potential transcriptional regulators of *ppck, pepc, ppdk-rp, ppdk, phya, phyb, pif3, hy5, elf3, elf4, lux, ost1, slac1, phot2*. From the total of 3,048 transcription factors from 69 transcription factor families, the analyses identified the patterns of multiple target genes from the same subnetwork sharing the same

potential transcriptional regulators. In addition, 21 transcription factor genes interacted with targets across all 4 subnetworks.

Because the Gene Regulatory Network Inference had provided the ranking of the potential transcriptional regulatory candidates, a reasonable step forward would be to probe for the direct binding evidence through molecular approaches. The state-of-the-art genome-wide chromatin accessibility profiling is ATAC-sequencing which had not been done on CAM species to date. Therefore, an improvement in the nuclei extraction protocol was required as it would provide the very first step towards an efficient ATAC-sequencing protocol.

Finally, Chapter 4 explored the Nuclei isolation followed by the flow cytometry separation technique for *Kalanchoe fedtschenkoi*. The DAPI staining signals indicated the presence of DNA materials in the nuclei extracts obtained from the sucrose-gradient method. In addition, the autofluorescence signal and the forward scattering signal could be used for further segregating chloroplast and mitochondria from the nuclei. Overall, the population that has fluorescence characteristics of nuclei can be obtained for further experiments. This concluded the thesis by opening up the way to connect the computational prediction back to the experimental designs and validations. The following section covered each chapter in greater detail with the recommendation for future research.

Significance of the findings under the broader context of CAM:

The system of Ordinary Differential Equations (ODEs) with the set of five equations that represent the key mesophyll metabolic processes were shown to capture the classical gas exchange patterns of *Kalanchoe* species (Chapter 2). Subsequently, the responses of the system were tested with two types of perturbation including the external CO_2 manipulation and the internal downregulation of PEPC activity. The key result showed that the ODE model is capable of responding to both external and internal perturbations. The model simulations captured the broad trends of the experimental gas-exchange data provided that the enzyme Rubisco becomes active earlier in the diurnal phase when modelling the PEPC downregulation scenario.

The early activation of Rubisco was a prerequisite in the PEPC knockdown simulation but not in the CO_2 perturbation simulation. This suggested that CAM mesophyll photosynthetic enzymes may be more sensitive to the concentration of metabolites within the mesophyll layer than the flow rate of carbon dioxide. Griffiths et al 2002 suggested that the environmental conditions limiting nocturnal PEPC activity may have resulted in the early activation of Rubisco in the following diurnal phase (Griffiths *et al.* 2002). The changes in metabolic profiles in the PEPC knockdown line were detected for malate, starch, sucrose, glucose and fructose (Susanna F. Boxall *et al.* 2020). Thus, these metabolites are interesting candidates for functioning as primary stimuli or secondary messengers communicating the nocturnal carbon starvation status to the diurnal carboxylation pathway.

As previously discussed in Chapter 2, the ODE model structure allows for an empirical fine-tuning through parameterisation. Nonetheless, the model outputs remain responsive to key biochemical parameters which include, but are not limited to, PEPC activity and Rubisco activity. Therefore, the ODE system could be a useful tool for modelling more types of perturbations and addressing the gap between the internal metabolic signals and the activities of key CAM proteins. Metabolites that are explicitly written in the ODE model can be parameterised with the data from different perturbation studies. In contrast, metabolites that are implicitly influencing the system through enzyme activities can be introduced into the model either via an addition of inhibition/activation equation or the adjustment of apparent enzyme activity.

Because the parameters of the ODE model represent the level of protein activities, all of the upstream regulations prior to protein function have been implicitly incorporated into the input

parameters. To explore the potential transcriptional regulation of CAM genes, the Gene Regulatory Network (GRN) inference method from timeseries transcriptome data offered a more independent solution (Chapter 3). The gene encoding for enzyme PEPCkinase (PPCK) was expected to be under a strict transcriptional regulation because it had the distinct nocturnal gene expression pattern. With the GRN inference method, the gene encoding for PEPCkinase was predicted to be under the transcriptional regulation of members from BES1, bHLH, TCP, AP2/ERF-ERF, C2H2 and MYB transcription factor families.

GRN inference approach also identified potential regulators of other key CAM genes that showed diel expression patterns but have not previously been probed for transcriptional regulation. These key CAM genes included *ppck, pepc, ppdk-rp, ppdk, phya, phyb, pif3, hy5, elf3, elf4, lux, ost1, slac1, phot2.* Transcriptional responses of CAM genes in *Mesembryanthemum crystallinum* had been curated previously (Cushman and Bohnert 1996). One member of PEPC gene family (Ppc1), NAD(P)-malate dehydrogenase and Pyruvate orthophosphate dikinase were upregulated in responses to salt stress and developmental stage in this inducible CAM species, whereas another member of PEPC gene family (Ppc2) was not responsive to salt stress. This has raised interesting question regarding the similarities or differences of the identity of the transcriptional regulators between *Mesembryanthemum* spp. and *Kalanchoe* spp. From a phylogenetic perspective, due to the distance between *Mesembryanthemum* genus and *Kalanchoe* genus, it would be unexpected for these two clades to rely on the equivalent set of transcriptional regulators.

Nonetheless, the regulatory motif must be accessible for these candidate transcriptional regulators to bind and induce or suppress the transcriptional activity. Methylation status in the promoter region of PEPC gene in C4 species like maize were reported to influence the expression level and enzyme concentration (Langdale *et al.* 1991). The chromatin accessibility would determine whether these potential regulators are functioning in vivo. The availability of the nuclei isolation protocol for *K. fedtschenkoi* (Chapter 4) allows for the further development of the Assay for Transposase Accessibility Chromatin Sequencing (ATAC) to determine the open chromatin region and possibly the transcription factor foot-printing for CAM species for the first time.

Apart from the epigenetics and transcriptional regulation, other levels of regulation play important roles in CAM. The classical post-translational regulation of PEPC has always played important roles in modulating the PEPC sensitivity to malate inhibition. Furthermore, evidence

of the non-canonical regulation by non-coding RNA is also emerging. Overall, the CAM system relies on precise control at every level. Nevertheless, the mRNA product from transcriptional step determines whether the protein can be produced in the first place, prior to any post-translational modifications.

All potential regulators including transcription factors and other types of transcriptional regulators identified in this thesis were ranked for all genes across the genome. This dataset is readily available for future research when one needs to prioritise which candidate regulators are to be genetically manipulated, given limited time and resources. With the available CAM transgenic protocols through RNA interference and CRISPR-Cas methods, the transcription regulators of interest can be manipulated. Subsequently, the system responses of the transgenic lines and the wildtype can be compared in terms of gene expression patterns, enzyme activities and the overall gas-exchange patterns to infer the biological significance of each candidate regulator.

Recommendations for future research:

Chapter 2 Ordinary differential equation modelling of Crassulacean acid metabolism to capture responses to perturbation conditions

Chapter 2 examined the performance of core components of the pre-existing systems dynamic model of CAM and examine the flexibility of the parameterisation and the optimisation steps. The findings supported the following hypotheses.

Hypothesis 1: The ODE system will be sufficient to capture the broad pattern of CAM but may be limited in terms of the parameterisation and optimisation.

Hypothesis 2: The ODE system will be limited by the higher level of regulatory controls including transcriptional controls.

Recommendation for future research for Chapter 2

To explore the roles of metabolites, the metabolic components can be parameterised directly or exert the influence through the addition of inhibition/activation equations into the ODE systems. However, the introducing of more parameters into the systems of equations are projected to introduce a greater degree of uncertainty into the model predictions.

The truncated daytime gas-exchange patterns which mimicked the drought stress response, and the predicted early activation of Rubisco in the transgenic PEPC knockdown line, should be explored in other transgenic plants to verify if such responses were common physiological responses for all PEPC knockdown plants. In addition, any concomitant changes from other photosynthetic components, such as the electron transport rate and other Calvin cycle enzymes that has occurred within the same time frame as Rubisco activation could be explored. The findings would allow for model development as well as physiological insights into the transgenic PEPC lines.

From statistical perspective, it is advisable to gather more experimental data for wildtype and transgenic plants under the same experimental settings, with a sample size of at least 30 biological replicates. Subsequently, the cross-validation method can be applied to verify the discrepancies between the trained and test dataset. The practical step would be to collect data of any new transgenic lines and their respective wildtype across the CAM research community.

Chapter 3 Gene regulatory network inference identifies key transcriptional regulators of CAM in *Kalanchoe fedtschenkoi*

Chapter 3 constructed Gene Regulatory Network from the publicly available timeseries transcriptome data of *Kalanchoe fedtschenkoi* for this model CAM species for the first time. The findings supported the following hypotheses.

Hypothesis 1: The amount of data from the timeseries transcriptome data of *K. fedtschenkoi* was sufficient for the dynGENIE3 algorithm whilst the size of the computer cluster would provide sufficient computational power and storage memory.

Hypothesis 2: The output of the Gene Regulatory Network will return candidates transcription factors of key CAM genes for each functional pathways; namely carboxylation, decarboxylation, circadian and stomata.

Hypothesis 3: A number of the associated candidates transcription factors have transcription factor binding sites (TFBS) in the promoter regions of their corresponding target genes.

Hypothesis 4: There will be shared transcription factors between multiple genes from the same pathway, and between genes from different pathways.

Hypothesis 5: This chapter will identify multiple transcription factors targeting the same target gene.

Recommendation for future research for Chapter 3

When more multi-omics data of the CAM model species has become available, the input functions of the dynGENIE3 algorithm can be replaced with the functional protein data which can capture the concentration, activity and location of proteins. The dynGENIE3 algorithm models the expression level of any gene j to be dependent on the rate of gene expression minus mRNA degradation. Subsequently, the transcription rate was assumed to be a potentially non-linear function f_j of the expression levels of any genes p. In the real biological system, the transcription rate is dependent on the concentration, activity, and location of transcriptional regulators which influence the probability of successful binding to promoter region of the target genes. Whenever the expression levels of gene p does not directly reflect the probability of successful binding, the GRN prediction will not represent the real biological system. Alternatively, by replacing the input function from gene expression level with protein

information (i.e. concentration, activity, and location), the GRN models can be restored back to represent the *in vivo* situations more accurately. Nevertheless, obtaining the timeseries proteomic data can be more challenging than transcriptomic data due to the possibility of signal loss through protein degradation whilst the protein molecules cannot be amplified like the transcript molecules prior to the signal detection.

Chapter 4 Nuclei isolation followed by flow cytometry separation technique for *Kalanchoe fedtschenkoi*

Chapter 4 constructed a protocol for nuclei isolation for *K. fedtschenkoi*. The use of a fluorescence-based spectral flow cytometer enhanced the ability to separate different particles in the nuclei extracts obtained from the sucrose-gradient nuclei isolation protocol. The finding supports the following hypotheses with the evidence of flow cytometry fluorescence absorption/emission signals and scattering signals, but could be additionally verified with the cell sorter instrument and fluorescence microscopy.

Hypothesis 1: The combination of sucrose-gradient isolation and fluorescence-based spectral flow cytometer would yield purify nuclei of greater than 10% of the total number of particles in the nuclei extract.

Hypothesis 2: The combination of sucrose-gradient isolation and fluorescence-based spectral flow cytometer would facilitate the separation of chloroplasts out of the nuclei extract.

Hypothesis 3: The combination of sucrose-gradient isolation and fluorescence-based spectral flow cytometer would facilitate the separation of mitochondria out of the nuclei extracts.

Recommendation for future research for Chapter 4

Product of nuclei isolation from the sucrose-gradient method can be passed through the cell sorting machine with the same wavelength excitation/detection setting as the has been previously used in the Cytek aurora flow cytometer (Chapter 4). Samples from each subpopulation with the distinct fluorescence signals and forward scattering signals can be retrieved and visualised under fluorescence microscope to confirm which absorption/emission characteristics represent the nuclei and other possible contaminants such as chloroplast, mitochondria, ruptured nuclei, and pigments.

Chapter 6

References

Chapter 1 References:

Abraham PE, Hurtado Castano N, Cowan-Turner D, *et al.* 2020. Peeling back the layers of crassulacean acid metabolism: functional differentiation between Kalanchoë fedtschenkoi epidermis and mesophyll proteomes. *Plant Journal* **103**: 869–888.

Abraham PE, Yin H, Borland AM, *et al.* 2016. Transcript, protein and metabolite temporal dynamics in the CAM plant Agave. *Nature Plants* **2**: 16178.

Adams P, Nelson DE, Yamada S, *et al.* 1998. Growth and development of Mesembryanthemum crystallinum (Aizoaceae). *New Phytologist* **138**: 171–190.

Ambawat S, Sharma P, Yadav NR, Yadav RC. **2013**. MYB transcription factor genes as regulators for plant responses: an overview. *Physiology and Molecular Biology of Plants* **19**: 307.

Amin AB. **2019**. Elucidating the Regulatory Role of the Obligate Crassulacean Acid Metabolism (CAM) Plant, Kalanchoe fedtschenkoi MYB Transcription Factor.

Amin AB, Rathnayake KN, Yim WC, *et al.* 2019. Crassulacean acid metabolism abiotic stress-responsive transcription factors: A potential genetic engineering approach for improving crop tolerance to abiotic stress. *Frontiers in Plant Science* **10**: 129.

Artus NN, Edwards GE. **1985**. NAD-malic enzyme from plants. *Federation of European Biochemical Societies* **182**: 225.

Aubry S, Brown NJ, Hibberd JM. **2011**. The role of proteins in C3 plants prior to their recruitment into the C4 pathway. *Journal of Experimental Botany* **62**: 3049–3059.

Bai Y, Dai X, Li Y, *et al.* 2019. Identification and characterization of pineapple leaf lncRNAs in crassulacean acid metabolism (CAM) photosynthesis pathway. *Scientific Reports* **9**: 1–11.

Bajic M, Maher KA, Deal RB. **2018**. Identification of open chromatin regions in plant genomes using ATAC-Seq In: *Methods in Molecular Biology*. Humana Press Inc., 183–201.

Bartlett MS, Vico G, Porporato A. **2014**. Coupled carbon and water fluxes in CAM photosynthesis: modeling quantification of water use efficiency and productivity. *Plant and Soil* **383**: 111–138.

Bastide B, Sipes D, Hann J, Ting IP. **1993**. Effect of severe water stress on aspects of crassulacean acid metabolism in Xerosicyos. *Plant Physiology* **103**: 1089–1096.

Black CC. **1973**. Photosynthetic carbon fixation in relation to net CO_2 uptake. *Annual Review of Plant Physiology* **24:** 253-286.

Black CC, Osmond CB. **2003**. Crassulacean acid metabolism photosynthesis: `working the night shift'. *Photosynthesis Research* **76**: 329–341.

Blasius B, Beck F, Lüttge U. **1998**. Oscillatory model of crassulacean acid metabolism: structural analysis and stability boundaries with a discrete hysteresis switch. *Plant, Cell & Environment* **21**: 775–784.

Borland AM, Griffiths H. **1990**. The Regulation of CAM and Respiratory Recycling by Water Supply and Light Regime in the C 3 -CAM Intermediate Sedum telephium. *Functional Ecology* **4**: 33.

Borland Anne M., Griffiths H. **1997**. A comparative study on the regulation of C3 and C4 carboxylation processes in the constitutive crassulacean acid metabolism (CAM) plant Kalanchoe daigremontiana and the C3-CAM intermediate Clusia minor. *Planta* **201**: 368–378.

Borland AM, Griffiths H, Hartwell J, Smith JAC. **2009**. Exploiting the potential of plants with crassulacean acid metabolism for bioenergy production on marginal lands In: *Journal of Experimental Botany*. Oxford Academic, 2879–2896.

Borland AM, Griffiths H, Maxwell C, Broadmeadow MSJ, Griffiths NM, Barnes JD. **1992**. On the ecophysiology of the Clusiaceae in Trinidad: expression of CAM in Clusia minor L. during the transition from wet to dry season and characterization of three endemic species. *New Phytologist* **122**: 349–357.

Borland Anne M, Guo H-B, Yang X, Cushman JC. **2016**. Orchestration of carbohydrate processing for crassulacean acid metabolism. *Current Opinion in Plant Biology* **31**: 118–124.

Borland AM, Hartwell J, Jenkins GI, Wilkins MB, Nimmo HG, Lüttge U. **1999**. Metabolite Control Overrides Circadian Regulation of Phosphoenolpyruvate Carboxylase Kinase and CO(2) Fixation in Crassulacean Acid Metabolism. *Plant physiology* **121**: 889–896.

Borland AM, Leverett A, Hurtado-Castano N, Hu R, Yang X. **2018**. Functional Anatomical Traits of the Photosynthetic Organs of Plants with Crassulacean Acid Metabolism In: Springer, Cham, 281–305.

Boxall SF, Dever L v., Knerová J, Gould PD, Hartwell J. **2017**. Phosphorylation of Phosphoenolpyruvate Carboxylase Is Essential for Maximal and Sustained Dark CO2 Fixation and Core Circadian Clock Operation in the Obligate Crassulacean Acid Metabolism Species Kalanchoë fedtschenkoi. *The Plant Cell* **29**: 2519–2536.

Boxall SF, Kadu N, Dever L v., *et al.* 2019. Silencing PHOSPHOENOLPYRUVATE CARBOXYLASE1 in the Obligate Crassulacean Acid Metabolism Species Kalanchoë laxiflora causes Reversion to C3-like Metabolism and Amplifies Rhythmicity in a Subset of Core Circadian Clock Genes. *bioRxiv*: 684050.

Boxall Susanna F., Kadu N, Dever L v., *et al.* 2020. Kalanchoë PPC1 Is Essential for Crassulacean Acid Metabolism and the Regulation of Core Circadian Clock and Guard Cell Signaling Genes. *The Plant Cell* **32**: 1136–1160.

Brilhaus D, Bräutigam A, Mettler-Altmann T, Winter K, Weber APM. **2015**. Reversible Burst of Transcriptional Changes during Induction of Crassulacean Acid Metabolism in Talinum triangulare. *Plant Physiology* **170**: 102–122.

von Caemmerer S, Evans JR, Hudson GS, Andrews TJ. **1994**. The kinetics of ribulose-1,5-bisphosphate carboxylase/oxygenase in vivo inferred from measurements of photosynthesis in leaves of transgenic tobacco. *Planta* **195**: 88–97.

von Caemmerer S, Griffiths H. **2009**. Stomatal responses to CO2 during a diel Crassulacean acid metabolism cycle in Kalanchoe daigremontiana and Kalanchoe pinnata. *Plant, Cell and Environment* **32**: 567–576.

Carter PJ, Nimmo HG, Fewson CA, Wilkins MB. **1991**. Circadian rhythms in the activity of a plant protein kinase. *The EMBO Journal* **10**: 2063–2068.

Ceusters N, Borland AM, Ceusters J. **2021**. How to resolve the enigma of diurnal malate remobilisation from the vacuole in plants with crassulacean acid metabolism? *New Phytologist* **229**: 3116–3124.

Ceusters J, Borland AM, Londers E, Verdoodt V, Godts C, de Proft MP. **2008**. Diel shifts in carboxylation pathway and metabolite dynamics in the CAM bromeliad Aechmea “Maya” in response to elevated CO2. *Annals of botany* **102**: 389–97.

Chastain CJ, Baird LM, Walker MT, *et al.* 2018. Maize leaf PPDK regulatory protein isoform-2 is specific to bundle sheath chloroplasts and paradoxically lacks a Pi-dependent PPDK activation activity. *Journal of Experimental Botany* **69**: 1171–1181.

Chen LY, Xin Y, Wai CM, Liu J, Ming R. **2020**. The role of cis-elements in the evolution of crassulacean acid metabolism photosynthesis. *Horticulture Research* **7**: 1–8.

Chiang CP, Yim WC, Sun YH, *et al.* 2016. Identification of ice plant (Mesembryanthemum crystallinum L.) microRNAs using RNA-seq and their putative roles in high salinity responses in seedlings. *Frontiers in Plant Science* **7**.

Chomthong M, Griffiths H. **2020**. Model approaches to advance crassulacean acid metabolism system integration. *The Plant Journal* **101**: 951–963.

Chu C, Dai Z, Ku MSB, Edwards GE. **1990**. Induction of crassulacean acid metabolism in the facultative halophyte Mesembryanthemum crystallinum by abscisic acid. *Plant Physiology* **93**: 1253–1260.

Cockburn W, Ting IP, Sternberg LO. **1979**. Relationships between Stomatal Behavior and Internal Carbon Dioxide Concentration in Crassulacean Acid Metabolism Plants. *Plant Physiology* **63**: 1029.

Cossarizza A, Chang HD, Radbruch A, *et al.* 2017. Guidelines for the use of flow cytometry and cell sorting in immunological studies*. *European Journal of Immunology* **47**: 1584–1797.

Cushman JC. **2001**. Crassulacean acid metabolism. A plastic photosynthetic adaptation to arid environments. *Plant physiology* **127**: 1439–48.

Cushman, J.C., Bohnert, H.J. 1996. Transcriptional Activation of CAM Genes During Development and Environmental Stress. In: Winter, K., Smith, J.A.C., eds. Crassulacean Acid Metabolism. *Ecological Studies*, Springer, Berlin, Heidelberg, 135-158.

Cushman JC, Tillett RL, Wood JA, Branco JM, Schlauch KA. **2008**. Large-scale mRNA expression profiling in the common ice plant, Mesembryanthemum crystallinum, performing C3 photosynthesis and Crassulacean acid metabolism (CAM). *Journal of Experimental Botany* **59**: 1875–1894.

Dai Z, Ku MSB, Zhang D, Edwards GE. **1994**. Effects of growth regulators on the induction of Crassulacean acid metabolism in the facultative halophyte Mesembryanthemum crystallinum L. *Planta* **192**: 287–294.

Davies BN, Griffiths H. **2012**. Competing carboxylases: circadian and metabolic regulation of Rubisco in C3 and CAM Mesembryanthemum crystallinum L. *Plant, Cell & Environment* **35**: 1211–1220.

Dever L v, Boxall SF, Kneřová J, Hartwell J. **2015**. Transgenic perturbation of the decarboxylation phase of Crassulacean acid metabolism alters physiology and metabolism but has only a small effect on growth. *Plant physiology* **167**: 44–59.

Dodd AN, Borland AM, Haslam RP, Griffiths H, Maxwell K. **2002**. Crassulacean acid metabolism: plastic, fantastic. *Journal of Experimental Botany* **53**: 569–580.

Drennan PM, Nobel PS. **2000**. Responses of CAM species to increasing atmospheric CO_2 concentrations. *Plant, Cell & Environment* **23**: 767–781.

Edwards EJ. **2019**. Evolutionary trajectories, accessibility and other metaphors: the case of C4 and CAM photosynthesis. *New Phytologist* **223**: 1742–1755.

Emmerlich V, Linka N, Reinhold T, *et al.* 2003. The plant homolog to the human sodium/dicarboxylic cotransporter is the vacuolar malate carrier. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 11122–11126.

Emmert-Streib F, Dehmer M, Haibe-Kains B. **2014**. Gene regulatory networks and their applications: understanding biological and medical problems in terms of networks. *Frontiers in Cell and Developmental Biology* **2**: 38.

Engineer CB, Hashimoto-Sugimoto M, Negi J, *et al.* 2016. CO2 Sensing and CO2 Regulation of Stomatal Conductance: Advances and Open Questions. *Trends in plant science* **21**: 16–30.

Ezer D, Shepherd SJK, Brestovitsky A, *et al.* 2017. The G-Box Transcriptional Regulatory Code in Arabidopsis. *Plant Physiology* **175**: 628–640.

Flemetakis E, Dimou M, Cotzur D, *et al.* 2003. A Lotus japonicus β-type carbonic anhydrase gene expression pattern suggests distinct physiological roles during nodule development. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* **1628**: 186–194.

Frei B, Eisenach C, Martinoia E, *et al.* 2018. Purification and functional characterization of the vacuolar malate transporter tDT from Arabidopsis. *Journal of Biological Chemistry* **293**: 4180–4190.

Givnish TJ, Barfuss MHJ, Ee B van, *et al.* 2014. Adaptive radiation, correlated and contingent evolution, and net species diversification in Bromeliaceae. *Molecular Phylogenetics and Evolution* **71**: 55–78.

González MC, Sánchez R, Cejudo FJ. **2003**. Abiotic stresses affecting water balance induce phosphoenolpyruvate carboxylase expression in roots of wheat seedlings. *Planta* **216**: 985–992.

Goodstein DM, Shu S, Howson R, *et al.* 2012. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research* **40**: D1178–D1186.

Gotoh E, Oiwamoto K, Inoue S, Shimazaki K, Doi M. **2019**. Stomatal response to blue light in crassulacean acid metabolism plants *Kalanchoe pinnata* and *Kalanchoe daigremontiana*. *Journal of Experimental Botany* **70**: 1367–1374.

Grant CE, Bailey TL, Noble WS. **2011**. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**: 1017–1018.

Greenfield A, Madar A, Ostrer H, Bonneau R. **2010**. DREAM4: Combining Genetic and Dynamic Information to Identify Biological Networks and Dynamical Models. *PLoS ONE* **5**.

Griffiths H, Cousins AB, Badger MR, von Caemmerer S. **2007**. Discrimination in the Dark. Resolving the Interplay between Metabolic and Physical Constraints to Phosphoenolpyruvate Carboxylase Activity during the Crassulacean Acid Metabolism Cycle 1.

Griffiths H, Helliker Bc B, Roberts A, *et al.* 2002. Regulation of Rubisco activity in crassulacean acid metabolism plants: better late than never. *Funct. Plant Biol* **29**: 689–696.

Griffiths H, Ong BL, Avadhani PN, Goh CJ. **1989**. Recycling of respiratory CO2 during Crassulacean acid metabolism: alleviation of photoinhibition in Pyrrosia piloselloides. *Planta 1989 179:1* **179**: 115–122.

Griffiths H, Robe WE, Girnus J, Maxwell K. **2008**. Leaf succulence determines the interplay between carboxylase systems and light use during Crassulacean acid metabolism in Kalanchoë species*. *Journal of Experimental Botany* **59**: 1851–1861.

Haag-kerwer A, Franco AC, Luttge U. **1992**. The Effect of Temperature and Light on Gas Exchange and Acid Accumulation in the C 3 -CAM Plant Clusia minor L. *Journal of Experimental Botany* **43**: 345–352.

Hafke JB, Hafke Y, Smith JAC, Lüttge U, Thiel G. **2003**. Vacuolar malate uptake is mediated by an anion-selective inward rectifier. *The Plant Journal* **35**: 116–128.

Han G, Lu C, Guo J, *et al.* 2020. C2H2 Zinc Finger Proteins: Master Regulators of Abiotic Stress Responses in Plants. *Frontiers in Plant Science* **11**.

Hartwell J, Dever L V, Boxall SF. **2016**. Emerging model systems for functional genomics analysis of Crassulacean acid metabolism. *Current Opinion in Plant Biology* **31**: 100–108.

Hartwell J, Gill A, Nimmo GA, Wilkins MB, Jenkins GI, Nimmo HG. **1999**. Phosphoenolpyruvate carboxylase kinase is a novel protein kinase regulated at the level of expression. *The Plant journal : for cell and molecular biology* **20**: 333–42.

Hartwell J, Smith LH, Wilkins MB, Jenkins GI, Nimmo HG. **1996**. Higher plant phosphoenolpyruvate carboxylase kinase is regulated at the level of translatable mRNA in response to light or a circadian rhythm. *The Plant Journal* **10**: 1071–1078.

Hartzell S, Bartlett MS, Inglese P, Consoli S, Yin J, Porporato A. **2020**. Modeling nonlinear dynamics of CAM productivity and water use for global predictions . *Plant, Cell & Environment*: 1–15.

Hartzell S, Bartlett MS, Porporato A. **2018**. Unified representation of the C3, C4, and CAM photosynthetic pathways with the Photo3 model. *Ecological Modelling* **384**: 173–187.

Heyduk K, Hwang M, Albert V, *et al.* 2019. Altered gene regulatory networks are associated with the transition from C 3 to crassulacean acid metabolism in erycina (Oncidiinae: Orchidaceae). *Frontiers in Plant Science* **9**: 2000.

Heyduk K, Ray JN, Ayyampalayam S, *et al.* 2019. Shared expression of crassulacean acid metabolism (CAM) genes pre-dates the origin of CAM in the genus Yucca. *Journal of Experimental Botany* **70:** 6597-6609.

Heyne B. **1816**. XVII. On the Deoxidation of the Leaves of Cotyledon calycina; in a Letter to A. B. Lambert, Esq., Vice-President of the Linnean Society. *Transactions of the Linnean Society of London* **os-11**: 213–215.

Hoang C v, Chapman KD. **2002**. Regulation of carbonic anhydrase gene expression in cotyledons of cotton (Gossypium hirsutum L.) seedlings during post-germinative growth. *Plant Molecular Biology* **49**: 449–458.

Hoang C v, Wessler HG, Local A, Turley RB, Benjamin RC, Chapman KD. **1999**. Identification and Expression of Cotton (Gossypium hirsutum L.) Plastidial Carbonic Anhydrase. *Plant Cell Physiol* **40**: 1262–1270.

Holtum JAM, Andrew J, Smith C, Neuhaus HE. **2005**. Intracellular transport and pathways of carbon flow in plants with crassulacean acid metabolism. *Functional Plant Biology* **32**: 429–449.

Holtum JAM, Winter K. **1982**. Activity of enzymes of carbon metabolism during the induction of Crassulacean acid metabolism in Mesembryanthemum crystallinum L. *Planta* **155**: 8–16.

Hong JC. **2016**. General Aspects of Plant Transcription Factor Families In: Gonzalez DH, eds. Plant Transcription Factors: *Evolutionary, Structural and Functional Aspects*. Academic Press, 35–56.

Hsu PY, Harmer SL. **2014**. Wheels within wheels: The plant circadian system. *Trends in Plant Science* **19**: 240–249.

Hu H, Boisson-Dernier A, Israelsson-Nordström M, *et al.* 2010. Carbonic Anhydrases are Upstream Regulators in Guard Cells of CO2-Controlled Stomatal Movements. *Nature cell biology* **12**: 87.

Huang X, Wang B, Xi J, *et al.* 2018. Transcriptome comparison reveals distinct selection patterns in domesticated and wild agave species, the important cam plants. *International Journal of Genomics* **2018**.

Hughes TE, Langdale JA, Kelly S. **2014**. The impact of widespread regulatory neofunctionalization on homeolog gene evolution following whole-genome duplication in maize. *Genome Research* **24**: 1348.

Huynh-Thu VA, Geurts P. **2018**. dynGENIE3: dynamical GENIE3 for the inference of gene networks from time series expression data. *Scientific Reports 2018 8:1* **8**: 1–12.

Huynh-Thu VA, Irrthum A, Wehenkel L, Geurts P. **2010**. Inferring Regulatory Networks from Expression Data Using Tree-Based Methods. *PLOS ONE* **5**: e12776.

Imaizumi N, Ku MSB, Ishihara K, Samejima M, Kaneko S, Matsuoka M. **1997**. Characterization of the gene for pyruvate,orthophosphate dikinase from rice, a C3 plant, and a comparison of structure and expression between C3 and C4 genes for this protein. *Plant Molecular Biology* **34**: 701–716.

Karg TJ, Golic KG. **2018**. Photoconversion of DAPI and Hoechst dyes to green and red-emitting forms after exposure to UV excitation. *Chromosoma* **127**: 235–245.

Kavroulakis N, Flemetakis E, Aivalakis G, Katinakis P. **2007**. Carbon Metabolism in Developing Soybean Root Nodules: The Role of Carbonic Anhydrase. *Molecular Plant-Microbe Interactions* **13**: 14–22.

Kluge M, Ting IP. **1978a**. The Metabolic Pathway of CAM In: Billings WD, Golley F, Lange OL, Olson JS, eds. Crassulacean Acid Metabolism. *Analysis of an Ecological Adaptation*. Springer-Verlag Berlin, Heidelberg, New York, 45–72.

Kluge M, Ting IP. **1978b**. Morphology, Anatomy, and Ultrastructure of CAM Plants In: Billings WD, Golley F, Lange OL, Olson JS, eds. Crassulacean Acid Metabolism. *Analysis of an Ecological Adaptation*. Springer-Verlag Berlin, Heidelberg, New York, 29–44.

Kluge M, Ting IP. **1978c**. Gas Exchange of CAM Plants In: Billings WD, Golley F, Lange OL, Olson JS, eds. Crassulacean Acid Metabolism. *Analysis of an Ecological Adaptation*. Springer-Verlag Berlin, Heidelberg, New York, 108–152.

Kondo A, Nose A, Yuasa H, Ueno O. **2000**. Species variation in the intracellular localization of pyruvate, Pi dikinase in leaves of crassulacean-acid-metabolism plants: an immunogold electron-microscope study. *Planta 2000 210:4* **210**: 611–621.

Langdale JA, Taylor WC, Nelson T. **1991**. Cell-specific accumulation of maize phosphoenolpyruvate carboxylase is correlated with demethylation at a specific site > 3 kb upstream of the gene. *MGG Molecular & General Genetics* **225**: 49–55.

Lee DM, Assmann SM. **1992**. Stomatal responses to light in the facultative Crassulacean acid metabolism species, Portulacaria afra. *Physiologia Plantarum* **85**: 35–42.

Lim SD, Lee S, Choi WG, Yim WC, Cushman JC. **2019**. Laying the foundation for crassulacean acid metabolism (CAM) biodesign: Expression of the c 4 metabolism cycle genes of CAM in arabidopsis. *Frontiers in Plant Science* **10**.

Liu D, Chen M, Mendoza B, Cheng H, Hu R, Li L, Trinh Cong T., *et al.* 2019. CRISPR/Cas9-mediated targeted mutagenesis for functional genomics research of crassulacean acid metabolism plants. *Journal of Experimental Botany* **70**: 6621–6629.

Lu Z, Hofmeister BT, Vollmers C, DuBois RM, Schmitz RJ. **2017**. Combining ATAC-seq with nuclei sorting for discovery of cis-regulatory regions in plant genomes. *Nucleic Acids Research* **45**: e41–e41.

Lüttge U. **2004**. Ecophysiology of Crassulacean Acid Metabolism (CAM). *Annals of Botany* **93**: 629–652.

Maleckova E, Brilhaus D, Wrobel TJ, Weber APM. **2019**. Transcript and metabolite changes during the early phase of abscisic acid-mediated induction of crassulacean acid metabolism in Talinum triangulare. *Journal of Experimental Botany* **70**: 6581–6596.

Males J, Griffiths H. **2017**. Stomatal Biology of CAM Plants. *Plant physiology* **174**: 550–560.

Mantel N, Myers M. **1971**. Problems of Convergence of Maximum Likelihood Iterative Procedures in Multiparameter Situations. *Journal of the American Statistical Association* **66:** 484-491.

Marbach D, Costello JC, Küffner R, *et al.* 2012. Wisdom of crowds for robust gene network inference. *Nature methods* **9**: 796.

Masumoto C, Miyazawa SI, Ohkawa H, *et al.* 2010. Phosphoenolpyruvate carboxylase intrinsically located in the chloroplast of rice plays a crucial role in ammonium assimilation. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 5226–5231.

Maurice CY, Poolman MG, Fell DA, George Ratcliffe R, Sweetlove LJ. **2014**. A diel flux balance model captures interactions between light and dark metabolism during day-night cycles in C3 and crassulacean acid metabolism leaves. *Plant Physiology* **165**: 917–929.

Maxwell K, Borland AM, Haslam RP, Helliker BR, Roberts A, Griffiths H. **1999**. Modulation of rubisco activity during the diurnal phases of the Crassulacean acid metabolism plant Kalanchoe daigremontiana. *Plant Physiology* **121**: 849–856.

Maxwell K, Von Caemmerer S, Evans JR. **1997**. Is a low internal conductance to CO2 diffusion a consequence of succulence in plants with crassulacean acid metabolism? *Australian Journal of Plant Physiology* **24**: 777–786.

Maxwell K, Griffiths H, Helliker B, *et al.* 2002. Regulation of Rubisco activity in crassulacean acid metabolism plants: better late than never. *Functional Plant Biology* **29**: 689.

Mercatelli D, Scalambra L, Triboli L, Ray F, Giorgi FM. **2020**. Gene regulatory network inference resources: A practical overview. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1863**: 194430.

Meyer S, Scholz-Starke J, de Angeli A, *et al.* 2011. Malate transport by the vacuolar AtALMT6 channel in guard cells is subject to multiple regulation. *The Plant Journal* **67**: 247–257.

Ming R, VanBuren R, Wai CM, *et al.* 2015. The pineapple genome and the evolution of CAM photosynthesis. *Nature Genetics* **47**: 1435–1442.

Miyao M, Fukayama H. **2003**. Metabolic consequences of overproduction of phosphoenolpyruvate carboxylase in C3 plants. *Archives of Biochemistry and Biophysics* **414**: 197–203.

Mochida K, Koda S, Inoue K, Nishii R. **2018**. Statistical and Machine Learning Approaches to Predict Gene Regulatory Networks From Transcriptome Datasets. *Frontiers in Plant Science* **9**: 1770.

Moseley RC, Mewalal R, Motta F, Tuskan GA, Haase S, Yang X. **2018**. Conservation and diversification of circadian rhythmicity between a model crassulacean acid metabolism plant kalanchoë fedtschenkoi and a model C3 photosynthesis plant arabidopsis thaliana. *Frontiers in Plant Science* **871**: 1757.

Moseley RC, Motta F, Tuskan GA, Haase SB, Yang X. **2021**. Inference of Gene Regulatory Network Uncovers the Linkage between Circadian Clock and Crassulacean Acid Metabolism in Kalanchoë fedtschenkoi. *Cells 2021, Vol. 10, Page 2217* **10**: 2217.

Neales TF. **1975**. The Gas Exchange Patterns of CAM Plants In: *Environmental and Biological Control of Photosynthesis*. Springer Netherlands, 299–310.

Ni Z, Lu Q, Huo H, Zhang H. **2019**. Estimation of Chlorophyll Fluorescence at Different Scales: A Review. *Sensors 2019, Vol. 19, Page 3000* **19**: 3000.

Nimmo HG. **2000**. The regulation of phosphoenolpyruvate carboxylase in CAM plants. *Trends in plant science* **5**: 75–80.

Nimmo HG. **2003**. How to tell the time: The regulation of phosphoenolpyruvate carboxylase in Crassulacean acid metabolism (CAM) plants. *Biochemical Society Transactions* **31**: 728–730.

Nott D, Osmond C. **1982**. Purification and Properties of Phosphoenolpyruvate Carboxylase From Plants With Crassulacean Acid Metabolism. *Functional Plant Biology* **9**: 409.

Olson DL. **2003**. Software Process Simulation. *Encyclopedia of Information Systems* **4:** 143–153.

Orth JD, Thiele I, Palsson BO. **2010**. What is flux balance analysis? *Nature Biotechnology* **28**: 245–248.

Osmond C. **1978**. Crassulacean Acid Metabolism: A Curiosity in Context. *Annual Review of Plant Physiology* **29:** 379-414.

Owen NA, Fahy KF, Griffiths H. **2016**. Crassulacean acid metabolism (CAM) offers sustainable bioenergy production and resilience to climate change. *GCB Bioenergy* **8**: 737–749.

Owen NA, Griffiths H. **2013**. A system dynamics model integrating physiology and biochemical regulation predicts extent of crassulacean acid metabolism (CAM) phases. *New Phytologist* **200**: 1116–1131.

Park PJ. **2009**. ChIP–seq: advantages and challenges of a maturing technology. *Nature Reviews Genetics 2009 10:10* **10**: 669–680.

Parry MAJ, Keys AJ, Madgwick PJ, Carmo-Silva AE, Andralojc PJ. **2008**. Rubisco regulation: A role for inhibitors In: *Journal of Experimental Botany*. Oxford Academic, 1569–1580.

Pineda De Castro LF, Zacharias M. **2002**. DAPI binding to the DNA minor groove: a continuum solvent analysis. *Journal of Molecular Recognition* **15**: 209–220.

Price GD, von Caemmerer S, Evans JR, *et al.* 1994. Specific reduction of chloroplast carbonic anhydrase activity by antisense RNA in transgenic tobacco plants has a minor effect on photosynthetic CO2 assimilation. *Planta 1994 193:3* **193**: 331–340.

Sánchez R, Flores A, Cejudo FJ. **2006**. Arabidopsis phosphoenolpyruvate carboxylase genes encode immunologically unrelated polypeptides and are differentially expressed in response to drought and salt stress. *Planta* **223**: 901–909.

Santelia D, Lawson T. **2016**. Rethinking guard cell metabolism. *Plant Physiology* **172**: 1371–1392.

de Saussure T. **1804**. *Recherches chimiques sur la végétation*. A Paris : Chez la ve. Nyon.

Shameer S, Baghalian K, Cheung C. Y.Maurice, Ratcliffe RG, Sweetlove LJ. **2018**. Computational analysis of the productivity potential of CAM. *Nature Plants* **4**: 165–171.

Sharma A, Wai CM, Ming R, Yu Q. **2017**. Diurnal Cycling Transcription Factors of Pineapple Revealed by Genome-Wide Annotation and Global Transcriptomic Analysis. *Genome Biology and Evolution* **9**: 2170.

Silvera K, Neubig KM, Whitten WM, *et al.* 2010. Evolution along the crassulacean acid metabolism continuum. *Functional Plant Biology* **37**: 995–1010.

Smith JAC, Winter K. **1996**. Taxonomic Distribution of Crassulacean Acid Metabolism In: Winter K, Smith JAC, eds. Crassulacean Acid Metabolism. *Biochemistry, Ecophysiology and Evolution.* Springer, 427–436.

Sun X, Wang Y, Sui N. **2018**. Transcriptional regulation of bHLH during plant response to stress. *Biochemical and Biophysical Research Communications* **503**: 397–401.

Tallman G, Zhu J, Mawson BT, *et al.* 1997. Induction of CAM in Mesembryanthemum crystallinum Abolishes the Stomatal Response to Blue Light and Light-Dependent Zeaxanthin Formation in Guard Cell Chloroplasts. *Plant and Cell Physiology* **38**: 236–242.

Tanious FA, Veal JM, Buczak H, Ratmeyer LS, Wilson WD. **1992**. DAPI (4',6-diamidino-2-phenylindole) binds differently to DNA and RNA: minor-groove binding at AT sites and intercalation at AU sites. *Biochemistry* **31**: 3103–3112.

Taybi T, Cushman JC. **1999**. Signaling events leading to Crassulacean acid metabolism induction in the common ice plant. *Plant Physiology* **121**: 545–555.

Taybi T, Sotta B, Gehrig H, Güçlü S, Kluge M, Brulfert J. **1995**. Differential Effects of Abscisic Acid on Phosphoenolpyruvate Carboxylase and CAM Operation in *Kalanchoë blossfeldiana*. *Botanica Acta* **108**: 240–246.

Thomas M. **1949**. Physiological Studies on Acid Metabolism in Green Plants. I. CO2 Fixation and CO2 Liberation in Crassulacean Acid Metabolism. *The New Phytologist* **48:** 390-420.

Töpfer N, Braam T, Shameer S, Ratcliffe RG, Sweetlove LJ. **2020**. Alternative Crassulacean Acid Metabolism Modes Provide Environment-Specific Water-Saving Benefits in a Leaf Metabolic Model. *The Plant Cell* **32**: 3689–3705.

Tsompana M, Buck MJ. **2014**. Chromatin accessibility: a window into the genome. *Epigenetics & Chromatin* **7:** 33.

Varala K, Marshall-Colón A, Cirrone J, *et al.* 2018. Temporal transcriptional logic of dynamic regulatory networks underlying nitrogen signaling and use in plants. *Proceedings of the National Academy of Sciences of the United States of America* **115**: 6494–6499.

Wai CM, VanBuren R. **2018**. Circadian Regulation of Pineapple CAM Photosynthesis In: Ming R, ed. Genetics and Genomics of Pineapple, Plant Genetics and Genomics. *Crops and Models*. Springer Nature Switzerland, 247–258.

Wai CM, VanBuren R, Zhang J, *et al.* 2017. Temporal and spatial transcriptomic and microRNA dynamics of CAM photosynthesis in pineapple. *The Plant Journal* **92**: 19–30.

Wai CM, Weise SE, Ozersky P, Mockler TC, Michael TP, VanBuren R. **2019**. Time of day and network reprogramming during drought induced CAM photosynthesis in Sedum album (NM Springer, Ed.). *PLOS Genetics* **15**: e1008209.

Wang Y, Zhang Y, Zhou R, *et al.* 2018. Identification and characterization of the bZIP transcription factor family and its expression in response to abiotic stresses in sesame. *PLOS ONE* **13**: e0200850.

Wedding RT, Black MK, Pap D. **1976**. Malate Dehydrogenase and NAD Malic Enzyme in the Oxidation of Malate by Sweet Potato Mitochondria. *Plant Physiology* **58**: 740–743.

Weiss I, Mizrahi Y, Raveh E. **2009**. Chamber response time: A neglected issue in gas exchange measurements. *Photosynthetica* **47**: 121–124.

Winter K. **2019**. Ecophysiology of constitutive and facultative CAM photosynthesis. *Journal of Experimental Botany* **70:** 6495–6508.

Winter K, Holtum JAM. **2007**. Environment or development? Lifetime net CO2 exchange and control of the expression of Crassulacean acid metabolism in Mesembryanthemum crystallinum. *Plant Physiology* **143**: 98–107.

Winter K, Holtum JAM, Smith JAC. **2015**. Crassulacean acid metabolism: a continuous or discrete trait? *New Phytologist* **208**: 73–78.

Wolf J. **1949**. Beiträge zur Kenntnis des Säurestoffwechsels Sukkulenter Crassulaceen. *Planta* **37**: 510–534.

Xie Z, Nolan TM, Jiang H, Yin Y. **2019**. AP2/ERF transcription factor regulatory networks in hormone and abiotic stress responses in Arabidopsis. *Frontiers in Plant Science* **10**: 228.

Yang X, Cushman JC, Borland AM, *et al.* 2015. A roadmap for research on crassulacean acid metabolism (CAM) to enhance sustainable food and bioenergy production in a hotter, drier world. *New Phytologist* **207**: 491–504.

Yang X, Hu R, Yin H, Jenkins J, Shu S, Tang H, Liu D, Weighill DA, Cheol Yim W, Ha J, Heyduk K, Goodstein DM, Guo HB, *et al.* 2017. The Kalanchoë genome provides insights into convergent evolution and building blocks of crassulacean acid metabolism. *Nature Communications 2017 8:1* **8**: 1–15.

Yin H, Guo HB, Weston DJ, *et al.* 2018. Diel rewiring and positive selection of ancient plant proteins enabled evolution of CAM photosynthesis in Agave. *BMC Genomics* **19**: 1–16.

Zhang L, Chen F, Zhang G-Q, *et al.* 2016. Origin and mechanism of crassulacean acid metabolism in orchids as implied by comparative transcriptomics and genomics of the carbon fixation pathway. *The Plant Journal* **86**: 175–185.

Zhang N, Portis AR. **1999**. Mechanism of light regulation of Rubisco: A specific role for the larger rubisco activase isoform involving reductive activation by thioredoxin-f. *Proceedings of the National Academy of Sciences of the United States of America* **96**: 9438–9443.

Zheng Y, Jiao C, Sun H, *et al.* 2016. iTAK: A Program for Genome-wide Prediction and Classification of Plant Transcription Factors, Transcriptional Regulators, and Protein Kinases. *Molecular Plant* **9**: 1667–1670.

Chapter 2 References:

Artus NN, Edwards GE. **1985**. NAD-malic enzyme from plants. *Federation of European Biochemical Societies* **182**: 225.

Bartlett MS, Vico G, Porporato A. **2014**. Coupled carbon and water fluxes in CAM photosynthesis: modeling quantification of water use efficiency and productivity. *Plant and Soil* **383**: 111–138.

Bastide B, Sipes D, Hann J, Ting IP. **1993**. Effect of severe water stress on aspects of crassulacean acid metabolism in Xerosicyos. *Plant Physiology* **103**: 1089–1096.

Black CC. **1973**. Photosynthetic carbon fixation in relation to net CO_2 uptake. *Annual Review of Plant Physiology* **24:** 253-286.

Blasius B, Beck F, Lüttge U. **1998**. Oscillatory model of crassulacean acid metabolism: structural analysis and stability boundaries with a discrete hysteresis switch. *Plant, Cell & Environment* **21**: 775–784.

Borland AM., Griffiths H. **1997**. A comparative study on the regulation of C3 and C4 carboxylation processes in the constitutive crassulacean acid metabolism (CAM) plant Kalanchoe daigremontiana and the C3-CAM intermediate Clusia minor. *Planta* **201**: 368–378.

Borland AM, Griffiths H, Hartwell J, Smith JAC. **2009**. Exploiting the potential of plants with crassulacean acid metabolism for bioenergy production on marginal lands In: *Journal of Experimental Botany*. Oxford Academic, 2879–2896.

Borland Anne M, Guo H-B, Yang X, Cushman JC. **2016**. Orchestration of carbohydrate processing for crassulacean acid metabolism. *Current Opinion in Plant Biology* **31**: 118–124.

Borland AM, Hartwell J, Jenkins GI, Wilkins MB, Nimmo HG, Lüttge U. **1999**. Metabolite Control Overrides Circadian Regulation of Phosphoenolpyruvate Carboxylase Kinase and CO(2) Fixation in Crassulacean Acid Metabolism. *Plant physiology* **121**: 889–896.

Boxall Susanna F., Kadu N, Dever L v., *et al.* 2020. Kalanchoë PPC1 Is Essential for Crassulacean Acid Metabolism and the Regulation of Core Circadian Clock and Guard Cell Signaling Genes. *The Plant Cell* **32**: 1136–1160.

von Caemmerer S, Evans JR, Hudson GS, Andrews TJ. **1994**. The kinetics of ribulose-1,5-bisphosphate carboxylase/oxygenase in vivo inferred from measurements of photosynthesis in leaves of transgenic tobacco. *Planta* **195**: 88–97.

von Caemmerer S, Griffiths H. **2009**. Stomatal responses to CO2 during a diel Crassulacean acid metabolism cycle in Kalanchoe daigremontiana and Kalanchoe pinnata. *Plant, Cell and Environment* **32**: 567–576.

Chomthong M, Griffiths H. **2020**. Model approaches to advance crassulacean acid metabolism system integration. *The Plant Journal* **101**: 951–963.

Dodd AN, Borland AM, Haslam RP, Griffiths H, Maxwell K. **2002**. Crassulacean acid metabolism: plastic, fantastic. *Journal of Experimental Botany* **53**: 569–580.

Emmerlich V, Linka N, Reinhold T, *et al.* 2003. The plant homolog to the human sodium/dicarboxylic cotransporter is the vacuolar malate carrier. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 11122–11126.

Frei B, Eisenach C, Martinoia E, *et al.* 2018. Purification and functional characterization of the vacuolar malate transporter tDT from Arabidopsis. *Journal of Biological Chemistry* **293**: 4180–4190.

Griffiths H, Helliker Bc B, Roberts A, *et al.* 2002. Regulation of Rubisco activity in crassulacean acid metabolism plants: better late than never. *Funct. Plant Biol* **29**: 689–696.

Griffiths H, Robe WE, Girnus J, Maxwell K. **2008**. Leaf succulence determines the interplay between carboxylase systems and light use during Crassulacean acid metabolism in Kalanchoë species*. *Journal of Experimental Botany* **59**: 1851–1861.

Hafke JB, Hafke Y, Smith JAC, Lüttge U, Thiel G. **2003**. Vacuolar malate uptake is mediated by an anion-selective inward rectifier. *The Plant Journal* **35**: 116–128.

Hartwell J, Dever L V, Boxall SF. **2016**. Emerging model systems for functional genomics analysis of Crassulacean acid metabolism. *Current Opinion in Plant Biology* **31**: 100–108.

Hartzell S, Bartlett MS, Inglese P, Consoli S, Yin J, Porporato A. **2020**. Modeling nonlinear dynamics of CAM productivity and water use for global predictions . *Plant, Cell & Environment*: 1–15.

Hartzell S, Bartlett MS, Porporato A. **2018**. Unified representation of the C3, C4, and CAM photosynthetic pathways with the Photo3 model. *Ecological Modelling* **384**: 173–187.

Holtum JAM, Andrew J, Smith C, Neuhaus HE. **2005**. Intracellular transport and pathways of carbon flow in plants with crassulacean acid metabolism. *Functional Plant Biology* **32**: 429–449.

Holtum JAM, Winter K. **1982**. Activity of enzymes of carbon metabolism during the induction of Crassulacean acid metabolism in Mesembryanthemum crystallinum L. *Planta* **155**: 8–16.

Lüttge U. **2004**. Ecophysiology of Crassulacean Acid Metabolism (CAM). *Annals of Botany* **93**: 629–652.

Mantel N, Myers M. **1971**. Problems of Convergence of Maximum Likelihood Iterative Procedures in Multiparameter Situations. *Journal of the American Statistical Association* **66:** 484-491.

Maxwell K, Borland AM, Haslam RP, Helliker BR, Roberts A, Griffiths H. **1999**. Modulation of rubisco activity during the diurnal phases of the Crassulacean acid metabolism plant Kalanchoe daigremontiana. *Plant Physiology* **121**: 849–856.

Maxwell K, Von Caemmerer S, Evans JR. **1997**. Is a low internal conductance to CO2 diffusion a consequence of succulence in plants with crassulacean acid metabolism? *Australian Journal of Plant Physiology* **24**: 777–786.

Meyer S, Scholz-Starke J, de Angeli A, *et al.* 2011. Malate transport by the vacuolar AtALMT6 channel in guard cells is subject to multiple regulation. *The Plant Journal* **67**: 247–257.

Neales TF. **1975**. The Gas Exchange Patterns of CAM Plants In: *Environmental and Biological Control of Photosynthesis*. Springer Netherlands, 299–310.

Nott D, Osmond C. **1982**. Purification and Properties of Phosphoenolpyruvate Carboxylase From Plants With Crassulacean Acid Metabolism. *Functional Plant Biology* **9**: 409.

Orth JD, Thiele I, Palsson BO. **2010**. What is flux balance analysis? *Nature Biotechnology* **28**: 245–248.

Osmond C. **1978**. Crassulacean Acid Metabolism: A Curiosity in Context. *Annual Review of Plant Physiology* **29:** 379-414.

Owen NA, Fahy KF, Griffiths H. **2016**. Crassulacean acid metabolism (CAM) offers sustainable bioenergy production and resilience to climate change. *GCB Bioenergy* **8**: 737–749.

Owen NA, Griffiths H. **2013**. A system dynamics model integrating physiology and biochemical regulation predicts extent of crassulacean acid metabolism (CAM) phases. *New Phytologist* **200**: 1116–1131.

Parry MAJ, Keys AJ, Madgwick PJ, Carmo-Silva AE, Andralojc PJ. **2008**. Rubisco regulation: A role for inhibitors In: *Journal of Experimental Botany*. Oxford Academic, 1569–1580.

Shameer S, Baghalian K, Cheung C. Y.Maurice, Ratcliffe RG, Sweetlove LJ. **2018**. Computational analysis of the productivity potential of CAM. *Nature Plants* **4**: 165–171.

Smith JAC, Winter K. **1996**. Taxonomic Distribution of Crassulacean Acid Metabolism In: Winter K, Smith JAC, eds. Crassulacean Acid Metabolism. *Biochemistry, Ecophysiology and Evolution.* Springer, 427–436.

Töpfer N, Braam T, Shameer S, Ratcliffe RG, Sweetlove LJ. **2020**. Alternative Crassulacean Acid Metabolism Modes Provide Environment-Specific Water-Saving Benefits in a Leaf Metabolic Model. *The Plant Cell* **32**: 3689–3705.

Wedding RT, Black MK, Pap D. **1976**. Malate Dehydrogenase and NAD Malic Enzyme in the Oxidation of Malate by Sweet Potato Mitochondria. *Plant Physiology* **58**: 740–743.

Weiss I, Mizrahi Y, Raveh E. **2009**. Chamber response time: A neglected issue in gas exchange measurements. *Photosynthetica* **47**: 121–124.

Yang X, Cushman JC, Borland AM, *et al.* 2015. A roadmap for research on crassulacean acid metabolism (CAM) to enhance sustainable food and bioenergy production in a hotter, drier world. *New Phytologist* **207**: 491–504.

Yang X, Hu R, Yin H, Jenkins J, Shu S, Tang H, Liu D, Weighill DA, Cheol Yim W, Ha J, Heyduk K, Goodstein DM, Guo H-B, *et al.* 2017. The Kalanchoë genome provides insights into convergent evolution and building blocks of crassulacean acid metabolism. *Nature Communications* **8**: 1899.

Zhang N, Portis AR. **1999**. Mechanism of light regulation of Rubisco: A specific role for the larger rubisco activase isoform involving reductive activation by thioredoxin-f. *Proceedings of the National Academy of Sciences of the United States of America* **96**: 9438–9443.

Chapter 3 References:

Abraham PE, Hurtado Castano N, Cowan-Turner D, *et al.* 2020. Peeling back the layers of crassulacean acid metabolism: functional differentiation between Kalanchoë fedtschenkoi epidermis and mesophyll proteomes. *Plant Journal* **103**: 869–888.

Ambawat S, Sharma P, Yadav NR, Yadav RC. **2013**. MYB transcription factor genes as regulators for plant responses: an overview. *Physiology and Molecular Biology of Plants* **19**: 307.

Amin AB. **2019**. Elucidating the Regulatory Role of the Obligate Crassulacean Acid Metabolism (CAM) Plant, Kalanchoe fedtschenkoi MYB Transcription Factor.

Amin AB, Rathnayake KN, Yim WC, *et al.* 2019. Crassulacean acid metabolism abiotic stress-responsive transcription factors: A potential genetic engineering approach for improving crop tolerance to abiotic stress. *Frontiers in Plant Science* **10**: 129.

Bai Y, Dai X, Li Y, *et al.* 2019. Identification and characterization of pineapple leaf lncRNAs in crassulacean acid metabolism (CAM) photosynthesis pathway. *Scientific Reports 2019 9:1* **9**: 1–11.

Borland Anne M, Guo H-B, Yang X, Cushman JC. **2016**. Orchestration of carbohydrate processing for crassulacean acid metabolism. *Current Opinion in Plant Biology* **31**: 118–124.

Boxall SF, Dever L v., Knerová J, Gould PD, Hartwell J. **2017**. Phosphorylation of Phosphoenolpyruvate Carboxylase Is Essential for Maximal and Sustained Dark CO2 Fixation and Core Circadian Clock Operation in the Obligate Crassulacean Acid Metabolism Species Kalanchoë fedtschenkoi. *The Plant Cell* **29**: 2519–2536.

Boxall Susanna F., Kadu N, Dever L v., *et al.* 2020. Kalanchoë PPC1 Is Essential for Crassulacean Acid Metabolism and the Regulation of Core Circadian Clock and Guard Cell Signaling Genes. *The Plant Cell* **32**: 1136–1160.

Brilhaus D, Bräutigam A, Mettler-Altmann T, Winter K, Weber APM. **2015**. Reversible Burst of Transcriptional Changes during Induction of Crassulacean Acid Metabolism in Talinum triangulare. *Plant Physiology* **170**: 102–122.

Ceusters N, Borland AM, Ceusters J. **2021**. How to resolve the enigma of diurnal malate remobilisation from the vacuole in plants with crassulacean acid metabolism? *New Phytologist* **229**: 3116–3124.

Chastain CJ, Baird LM, Walker MT, *et al.* 2018. Maize leaf PPDK regulatory protein isoform-2 is specific to bundle sheath chloroplasts and paradoxically lacks a Pi-dependent PPDK activation activity. *Journal of Experimental Botany* **69**: 1171–1181.

Chen LY, Xin Y, Wai CM, Liu J, Ming R. **2020**. The role of cis-elements in the evolution of crassulacean acid metabolism photosynthesis. *Horticulture Research 2020 7:1* **7**: 1–8.

Cockburn W, Ting IP, Sternberg LO. **1979**. Relationships between Stomatal Behavior and Internal Carbon Dioxide Concentration in Crassulacean Acid Metabolism Plants. *Plant Physiology* **63**: 1029.

Edwards EJ. **2019**. Evolutionary trajectories, accessibility and other metaphors: the case of C4 and CAM photosynthesis. *New Phytologist* **223**: 1742–1755.

Emmert-Streib F, Dehmer M, Haibe-Kains B. **2014**. Gene regulatory networks and their applications: understanding biological and medical problems in terms of networks. *Frontiers in Cell and Developmental Biology* **2**: 38.

Engineer CB, Hashimoto-Sugimoto M, Negi J, *et al.* 2016. CO2 Sensing and CO2 Regulation of Stomatal Conductance: Advances and Open Questions. *Trends in plant science* **21**: 16–30.

Ezer D, Shepherd SJK, Brestovitsky A, *et al.* 2017. The G-Box Transcriptional Regulatory Code in Arabidopsis. *Plant Physiology* **175**: 628–640.

Givnish TJ, Barfuss MHJ, Ee B van, *et al.* 2014. Adaptive radiation, correlated and contingent evolution, and net species diversification in Bromeliaceae. *Molecular Phylogenetics and Evolution* **71**: 55–78.

Goodstein DM, Shu S, Howson R, *et al.* 2012. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research* **40**: D1178–D1186.

Grant CE, Bailey TL, Noble WS. **2011**. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**: 1017–1018.

Greenfield A, Madar A, Ostrer H, Bonneau R. **2010**. DREAM4: Combining Genetic and Dynamic Information to Identify Biological Networks and Dynamical Models. *PLoS ONE* **5**.

Griffiths H, Ong BL, Avadhani PN, Goh CJ. **1989**. Recycling of respiratory CO2 during Crassulacean acid metabolism: alleviation of photoinhibition in Pyrrosia piloselloides. *Planta 1989 179:1* **179**: 115–122.

Han G, Lu C, Guo J, *et al.* 2020. C2H2 Zinc Finger Proteins: Master Regulators of Abiotic Stress Responses in Plants. *Frontiers in Plant Science* **11**.

Hartwell J, Dever L v., Boxall SF. **2016**. Emerging model systems for functional genomics analysis of Crassulacean acid metabolism. *Current Opinion in Plant Biology* **31**: 100–108.

Hartwell J, Gill A, Nimmo GA, Wilkins MB, Jenkins GI, Nimmo HG. **1999**. Phosphoenolpyruvate carboxylase kinase is a novel protein kinase regulated at the level of expression. *The Plant Journal* **20**: 333–342.

Hartwell J, Smith LH, Wilkins MB, Jenkins GI, Nimmo HG. **1996**. Higher plant phosphoenolpyruvate carboxylase kinase is regulated at the level of translatable mRNA in response to light or a circadian rhythm. *The Plant Journal* **10**: 1071–1078.

Heyduk K, Hwang M, Albert V, *et al.* 2019. Altered gene regulatory networks are associated with the transition from C 3 to crassulacean acid metabolism in erycina (Oncidiinae: Orchidaceae). *Frontiers in Plant Science* **9**: 2000.

Hong JC. **2016**. General Aspects of Plant Transcription Factor Families In: Gonzalez DH, eds. Plant Transcription Factors: *Evolutionary, Structural and Functional Aspects*. Academic Press, 35–56.

Huang X, Wang B, Xi J, *et al.* 2018. Transcriptome comparison reveals distinct selection patterns in domesticated and wild agave species, the important cam plants. *International Journal of Genomics* **2018**.

Huynh-Thu VA, Geurts P. **2018**. dynGENIE3: dynamical GENIE3 for the inference of gene networks from time series expression data. *Scientific Reports 2018 8:1* **8**: 1–12.

Liu D, Chen M, Mendoza B, Cheng H, Hu R, Li L, Trinh Cong T., *et al.* 2019. CRISPR/Cas9-mediated targeted mutagenesis for functional genomics research of crassulacean acid metabolism plants. *Journal of Experimental Botany* **70**: 6621–6629.

Marbach D, Costello JC, Küffner R, *et al.* 2012. Wisdom of crowds for robust gene network inference. *Nature methods* **9**: 796.

Ming R, VanBuren R, Wai CM, *et al.* 2015. The pineapple genome and the evolution of CAM photosynthesis. *Nature Genetics* **47**: 1435–1442.

Mochida K, Koda S, Inoue K, Nishii R. **2018**. Statistical and Machine Learning Approaches to Predict Gene Regulatory Networks From Transcriptome Datasets. *Frontiers in Plant Science* **9**: 1770.

Moseley RC, Mewalal R, Motta F, Tuskan GA, Haase S, Yang X. **2018**. Conservation and diversification of circadian rhythmicity between a model crassulacean acid metabolism plant kalanchoë fedtschenkoi and a model C3 photosynthesis plant arabidopsis thaliana. *Frontiers in Plant Science* **871**: 1757.

Moseley RC, Motta F, Tuskan GA, Haase SB, Yang X. **2021**. Inference of Gene Regulatory Network Uncovers the Linkage between Circadian Clock and Crassulacean Acid Metabolism in Kalanchoë fedtschenkoi. *Cells 2021, Vol. 10, Page 2217* **10**: 2217.

Nimmo HG. **2003**. How to tell the time: The regulation of phosphoenolpyruvate carboxylase in Crassulacean acid metabolism (CAM) plants. *Biochemical Society Transactions* **31**: 728–730.

Sharma A, Wai CM, Ming R, Yu Q. **2017**. Diurnal Cycling Transcription Factors of Pineapple Revealed by Genome-Wide Annotation and Global Transcriptomic Analysis. *Genome Biology and Evolution* **9**: 2170.

Silvera K, Neubig KM, Whitten WM, *et al.* 2010. Evolution along the crassulacean acid metabolism continuum. *Functional Plant Biology* **37**: 995–1010.

Sun X, Wang Y, Sui N. **2018**. Transcriptional regulation of bHLH during plant response to stress. *Biochemical and Biophysical Research Communications* **503**: 397–401.

Wai CM, VanBuren R. **2018**. Circadian Regulation of Pineapple CAM Photosynthesis In: Ming R, ed. Genetics and Genomics of Pineapple, Plant Genetics and Genomics. *Crops and Models*. Springer Nature Switzerland, 247–258.

Wai CM, VanBuren R, Zhang J, *et al.* 2017. Temporal and spatial transcriptomic and microRNA dynamics of CAM photosynthesis in pineapple. *The Plant Journal* **92**: 19–30.

Wai CM, Weise SE, Ozersky P, Mockler TC, Michael TP, VanBuren R. **2019**. Time of day and network reprogramming during drought induced CAM photosynthesis in Sedum album (NM Springer, Ed.). *PLOS Genetics* **15**: e1008209.

Wang Y, Zhang Y, Zhou R, *et al.* 2018. Identification and characterization of the bZIP transcription factor family and its expression in response to abiotic stresses in sesame. *PLOS ONE* **13**: e0200850.

Xie Z, Nolan TM, Jiang H, Yin Y. **2019**. AP2/ERF transcription factor regulatory networks in hormone and abiotic stress responses in Arabidopsis. *Frontiers in Plant Science* **10**: 228.

Yang X, Cushman JC, Borland AM, *et al.* 2015. A roadmap for research on crassulacean acid metabolism (CAM) to enhance sustainable food and bioenergy production in a hotter, drier world. *New Phytologist* **207**: 491–504.

Yang X, Hu R, Yin H, Jenkins J, Shu S, Tang H, Liu D, Weighill DA, Cheol Yim W, Ha J, Heyduk K, Goodstein DM, Guo H-B, *et al.* 2017. The Kalanchoë genome provides insights into convergent evolution and building blocks of crassulacean acid metabolism. *Nature Communications* **8**: 1899.

Yin H, Guo HB, Weston DJ, *et al.* 2018. Diel rewiring and positive selection of ancient plant proteins enabled evolution of CAM photosynthesis in Agave. *BMC Genomics* **19**: 1–16.

Zheng Y, Jiao C, Sun H, *et al.* 2016. iTAK: A Program for Genome-wide Prediction and Classification of Plant Transcription Factors, Transcriptional Regulators, and Protein Kinases. *Molecular Plant* **9**: 1667–1670.

Chapter 4 References:

Bajic M, Maher KA, Deal RB. **2018**. Identification of open chromatin regions in plant genomes using ATAC-Seq. In: Methods in Molecular Biology. Humana Press Inc., 183–201.

Cossarizza A, Chang HD, Radbruch A, Akdis M, Andrä I, Annunziato F, Bacher P, Barnaba V, Battistini L, Bauer WM, *et al.* 2017. Guidelines for the use of flow cytometry and cell sorting in immunological studies*. *European Journal of Immunology* **47**: 1584–1797.

Dimauro I, Pearson T, Caporossi D, Jackson MJ. **2012**. A simple protocol for the subcellular fractionation of skeletal muscle cells and tissue. *BMC Research Notes* **5**: 1–5.

Karg TJ, Golic KG. **2018**. Photoconversion of DAPI and Hoechst dyes to green and red-emitting forms after exposure to UV excitation. *Chromosoma* **127**: 235–245.

Lu Z, Hofmeister BT, Vollmers C, DuBois RM, Schmitz RJ. **2017**. Combining ATAC-seq with nuclei sorting for discovery of cis-regulatory regions in plant genomes. *Nucleic Acids Research* **45**: e41–e41.

Ni Z, Lu Q, Huo H, Zhang H. **2019**. Estimation of Chlorophyll Fluorescence at Different Scales: A Review. *Sensors 2019, Vol. 19, Page 3000* **19**: 3000.

Park PJ. **2009**. ChIP–seq: advantages and challenges of a maturing technology. *Nature Reviews Genetics 2009 10:10* **10**: 669–680.

Pineda De Castro LF, Zacharias M. **2002**. DAPI binding to the DNA minor groove: a continuum solvent analysis. *Journal of Molecular Recognition* **15**: 209–220.

Tanious FA, Veal JM, Buczak H, Ratmeyer LS, Wilson WD. **1992**. DAPI (4',6-diamidino-2-phenylindole) binds differently to DNA and RNA: minor-groove binding at AT sites and intercalation at AU sites. *Biochemistry* **31**: 3103–3112.

Tsompana M, Buck MJ. **2014**. Chromatin accessibility: a window into the genome. *Epigenetics & Chromatin* **7:** 33.

Chapter 5 References:

Boxall SF, Kadu N, Dever L v., Knerová J, Waller JL, Gould PJD, Hartwell J. **2020**. Kalanchoë PPC1 Is Essential for Crassulacean Acid Metabolism and the Regulation of Core Circadian Clock and Guard Cell Signaling Genes. *The Plant Cell* **32**: 1136–1160.

Cushman JC, Bohnert HJ. **1996**. Transcriptional Activation of CAM Genes During Development and Environmental Stress In: Winter K, Smith JAC, eds. *Crassulacean Acid Metabolism. Biochemistry, Ecophysiology and Evolution*. Springer, 135–158.

Griffiths H, Helliker Bc B, Roberts A, Haslam RP, Girnus J, Robe WE, Borland AM, Maxwell K. **2002**. Regulation of Rubisco activity in crassulacean acid metabolism plants: better late than never. *Funct. Plant Biol* **29**: 689–696.

Langdale JA, Taylor WC, Nelson T. **1991**. Cell-specific accumulation of maize phosphoenolpyruvate carboxylase is correlated with demethylation at a specific site > 3 kb upstream of the gene. *MGG Molecular & General Genetics* **225**: 49–55.

Appendices

Appendix A:

1. ODE_PEPCexperiment_WT.m

```
function dydt = ODE_PEPCexperiment_WT(t,y)
dydt = zeros(5,1);
%y1 is Aci, y2 is Acc, y3 is Amc, y4 is Amv, y5 is Ac

% Physical parameters. These are fixed.
Cca= 400;
Henrys_law_constant = 0.03445;
Conversion_1 = 0.025; % Owen paper 2013

% Anatomical parameters. Measurable.
Succulence_leaf_area_to_volume = 2.8;
Percentage_Chlorenchyma_air_space = 0.143;
Chloren_V = Succulence_leaf_area_to_volume*Percentage_Chlorenchyma_air_space;
gm = 0.053;
Xvmax = 150000;

% Biochemical parameters
Vm_mitochondria = 0.6;% Have no experimental data but might be able to measure.

% Set condition for multiple days
t_hour = t/3600;
r = mod(t_hour,24);

if r>12
%if t>43200 %night time
    Malic_acid_efflux_rate = 0;
else
    Malic_acid_efflux_rate = 33.19;
end

if r>12
%if t>43200 %night time
    Malic_acid_influx_rate = 7.34;%Nick doesnt have this but set it to mirror PEPC
activity
else
    Malic_acid_influx_rate = 0;
end

% Enzyme constants
Vdmax = 5.5;
Kd = 11;

Vcmax = 10.82;
Kc = 10.8;

Vpmax = 5;
Kp = 12;

if r<12
%if t<43200
    gs = gs_WT_day (t/3600);
else
    gs = gs_WT_night(t/3600);
end
```

```
Rubisco_activation_term = sin(2*pi*t/86400);%active at first half of 24 cycle
if Rubisco_activation_term<=0
    Rubisco_activation_term = 0;
end

PEPC_activation_term = -sin(2*pi*t/86400);%active at second half of 24 cycle
if PEPC_activation_term<=0
    PEPC_activation_term = 0;
end

Decarboxylation_activation_term = sin(2*pi*t/86400);%active at first half of 24
cycle
if Decarboxylation_activation_term<=0
    Decarboxylation_activation_term = 0;
end

% Differential equations
dydt(1) = gs*(Cca-Conversion_1*y(1))-gm*(y(1)*Conversion_1-
y(2)/Chloren_V/Henrys_law_constant);%Aci

Ki = 8000; %reference from Nick's table and Borland&Griffiths 2008
PEPC_reaction = PEPC_activation_term*(Vpmax  * y(2)/Chloren_V) / (Kp*(1+y(3)/Ki)
+ y(2)/Chloren_V);

dydt(2) = gm*(y(1)*Conversion_1-
y(2)/Chloren_V/Henrys_law_constant)+Decarboxylation_activation_term*(Vdmax  *
y(3)/(Kd+y(3)) )+Vm_mitochondria-Rubisco_activation_term*(Vcmax  * y(2)/Chloren_V)
/ (Kc  + y(2)/Chloren_V) - PEPC_reaction;
dydt(3) = PEPC_reaction - Decarboxylation_activation_term*(Vdmax  * y(3)/(Kd+y(3))
)-0.05*Malic_acid_influx_rate*(1-
y(4)/Xvmax)*y(3)+Malic_acid_efflux_rate*(y(4)/Xvmax);
dydt(4) = 0.05*Malic_acid_influx_rate*(1-y(4)/Xvmax)*y(3)-
Malic_acid_efflux_rate*(y(4)/Xvmax);
dydt(5) = Rubisco_activation_term*Vcmax*y(2)/Chloren_V/(Kc+y(2)/Chloren_V)-
Vm_mitochondria;
```

2. plotJs.m

```
%% PlotJs

%Solve
[t,y] =
ode15s(@ODE_PEPCexperiment_WT,[0:1:86400*120],[6.7104,5.8921,1.5617e+05,6.5678e+04
,4.7408e+05]);


figure
%read experimental data from gas_exchange.xlsx file.
%plot(t_WT_exp,Js_WT_exp,'+b')

%plot5cycles
t_exp_5cycles = [t_WT_exp;t_WT_exp+24;t_WT_exp+24*2;t_WT_exp+24*3;t_WT_exp+24*4];
Js_exp_5cycles = repmat(Js_WT_exp,5,1);
plot(t_exp_5cycles,Js_exp_5cycles,'+k', 'LineWidth', 1)

hold on

%% Get Js from model

%generate repeat gs for 1 cycle
n = 1;
t_hour = [0:1:86400*n]/3600;
r = mod(t_hour,24);

gs_array = [];
for i = 1:length(t_hour)-1
    if r(i)<12
        gs = gs_WT_day (t_hour(i));
    else
        gs = gs_WT_night(t_hour(i));
    end
    gs_array = [gs_array,gs];
end

%concatenate into 5 cycles
gs_array_5cycles = repmat(gs_array,1,5)/1.6;
gs_array_120cycles = repmat(gs_array,1,120)/1.6;

%select to plot 100th to 105th cycle
y_5cycles = y(86400*100+1:86400*105,:);

%calculate Js
Cca = 400;
Conversion_1 = 0.025;
Js = gs_array_5cycles.*(Cca-Conversion_1*y_5cycles(:,1)');%WT
%Js = gs_array.*(Cca-Conversion_1*y_rPPC1B(:,1));%rPPC1-B

plot([0:1:86400*5-1]/3600, Js, 'k', 'LineWidth', 3);

%% Label plot

xlabel ('Time (Hours)','FontSize', 18)
ylabel ('CO_{2} assimilation rate (\mumol m^{-2} s^{-1})','FontSize', 18)
hold on
```

```
%vertical_line = [-2:12];
for i = 1:9
    plot(86400*i/2/3600*ones(1,length(vertical_line)),vertical_line, 'Color', [0.5
0.5 0.5])
end
xticks([0 12 12*2 12*3 12*4 12*5 12*6 12*7 12*8 12*9 12*10])

%add grey filled areas
fill([12 24 24 12],[-1 -1 7 7],'k','FaceAlpha', 0.3)
fill([36 48 48 36],[-1 -1 7 7],'k','FaceAlpha', 0.3)
fill([60 72 72 60],[-1 -1 7 7],'k','FaceAlpha', 0.3)
fill([84 96 96 84],[-1 -1 7 7],'k','FaceAlpha', 0.3)
fill([108 120 120 108],[-1 -1 7 7],'k','FaceAlpha', 0.3)
```

3. gs_WT_day.m

```
function gs_fitted = gs_WT_day(x)

%Coefficients (with 95% confidence bounds):
    a =    0.006533 ;% (0.005408, 0.007658)
    b =     -0.2216 ;% (-0.2892, -0.1541)

    gs_fitted = a*exp(b*x);
end
```

4. gs_WT_night.m

```
function gs_fitted = gs_WT_night(x)

%Coefficients (with 95% confidence bounds):
    a1_phase1 =       0.02799;%  (0.02311, 0.03287)
    b1_phase1 =      18.39   ;%  (17.88, 18.9)
    c1_phase1 =       4.408  ;%  (3.6, 5.217)

    gs_fit_phase1 = a1_phase1*exp(-((x-b1_phase1)/c1_phase1).^2);
    gs_fitted = gs_fit_phase1;
end
```

Appendix B:

Chomthong M, Griffiths H. **2020**. Model approaches to advance crassulacean acid metabolism system integration. *The Plant Journal* **101**: 951–963.

the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

The Plant Journal (2020) **101,** 951–963 doi: 10.1111/tpj.14691

SI ADVANCES IN PHOTOSYNTHESIS

Model approaches to advance crassulacean acid metabolism system integration

Methawi Chomthong* and **Howard Griffiths**

Department of Plant Sciences, University of Cambridge, Downing street, Cambridge CB2 3EA, UK

Received 1 October 2019; accepted 2 January 2020; published online 14 January 2020.
*For correspondence (e-mail mc901@cam.ac.uk).

SUMMARY

This review summarises recent progress in understanding crassulacean acid metabolism (CAM) systems and the integration of internal and external stimuli to maximise water-use efficiency. Complex CAM traits have been reduced to their minimum and captured as computational models, which can now be refined using recently available data from transgenic manipulations and large-scale omics studies. We identify three key areas in which an appropriate choice of modelling tool could help capture relevant comparative molecular data to address the evolutionary drivers and plasticity of CAM. One focus is to identify the environmental and internal signals that drive inverse stomatal opening at night. Secondly, it is important to identify the regulatory processes required to orchestrate the diel pattern of carbon fluxes within mesophyll layers. Finally, the limitations imposed by contrasting succulent systems and associated hydraulic conductance components should be compared in the context of water-use and evolutionary strategies. While network analysis of transcriptomic data can provide insights via co-expression modules and hubs, alternative forms of computational modelling should be used iteratively to define the physiological significance of key components and informing targeted functional gene manipulation studies. We conclude that the resultant improvements of bottom-up, mechanistic modelling systems can enhance progress towards capturing the physiological controls for phylogenetically diverse CAM systems in the face of the recent surge of information in this omics era.

Keywords: stomatal physiology, mesophyll metabolism, hydraulic conductance, systems dynamic.

INTRODUCTION

Crassulacean acid metabolism (CAM) species are known for their remarkable photosynthetic traits, including nighttime primary CO_2 fixation by the enzyme phosphoenol pyruvate carboxylase (PEPC), which conserves water when stomata are closed for part of the day. At a morphological level, succulence – of cells, leaves or stems – is typically associated with both semi-arid desert environments and the epiphytic niches in tropical rainforests, leading to remarkable global diversity arising from multiple independent origins of CAM over the last 5–10 million years (Dodd *et al.*, 2002; Borland *et al.*, 2018; Wai and VanBuren, 2018; Edwards, 2019). Many physiological, biochemical and molecular features have evolved convergently to regulate the expression of CAM in such contrasting systems. This review will provide a brief introductory framework summarising these features, and then identify modelling approaches that could help to integrate the wealth of new data arising from recent molecular analyses. The specific aim will be to develop a better understanding of three major regulatory processes constraining CAM productivity across the day–night cycle, namely stomatal functioning, mesophyll photosynthetic biochemistry and metering of water use.

The distinct regulation of guard cells allows stomatal opening at night, directing CO_2 flux into the process of primary carboxylation by PEPC during Phase I of CAM (Osmond, 1978). This temporal shift in initial carboxylation by PEPC is controlled through reduced inhibition of PEPC by a phosphorylation reaction mediated by a circadian clock-controlled protein kinase called phosphoenol pyruvate carboxylase kinase (PEPC kinase; Hartwell *et al.*, 1999; Nimmo, 2000; Dodd *et al.*, 2002). The final product of this primary carboxylation process, malic acid, is then sequestered into large storage vacuoles that can occupy up to

95% of the mesophyll cell volume. During the subsequent light period (Phase III), regulation can occur during the decarboxylation of malate, which releases CO_2 for secondary carboxylation by ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO) as well as the activation of RubisCO (Griffiths *et al.*, 2002). The CO_2 is concentrated around RubisCO behind closed stomata during the daytime, allowing CAM RubisCO to adopt a more relaxed specificity factor (i.e. the selectivity of CO_2 over O_2 is lower) compared with C3 plants, albeit with the retention of a high affinity for CO_2 (low $K_{m,CO2}$; Griffiths *et al.*, 2008). This temporal separation between primary and secondary CO_2 fixation within a single mesophyll cell enhances overall operating carboxylation and water-use efficiencies compared with C3 species (Borland *et al.*, 2009).

A number of complex CAM phenotypes demonstrate remarkable physiological and ecological plasticity which still engender intriguing research questions (Yang *et al.*, 2015; Winter *et al.*, 2015; Borland *et al.*, 2018; Edwards, 2019; Winter, 2019). In strong CAM species such as *Kalanchoe daigremontiana*, *Agave tequilana* and *Tillandsia usneoides* the strength and duration of the intermediate Phases II and IV (respectively early morning and late afternoon) are adversely affected by a degree of drought, but can be restored quickly upon rehydration (Owen *et al.*, 2016). Plasticity, as represented by facultative CAM, has also been documented for woody dicotyledonous species such as *Clusia minor*, temperate perennial herbs such as *Sedum telephium* and many members of the Aizoaceae which can rapidly and reversibly undergo CAM induction (Borland and Griffiths, 1990; Borland *et al.*, 1992; Haag-kerwer *et al.*, 1992; Dodd *et al.*, 2002; Winter, 2019). Finally, there is the well-defined induction of CAM by stress in *Mesembryanthemum crystallinum* (Adams *et al.*, 1998; Cushman, 2001; Winter and Holtum, 2007) or during leaf development and aging in *Kalanchoe* species (Hartwell *et al.*, 2016).

These examples of the inducibility of CAM and Phase II and IV responsiveness suggest that the mechanisms regulating CAM drive a core temporal switch which can then be modified depending on environmental conditions or internal metabolic feedback. This is supported by the analysis of the pineapple (*Ananas comosus*) genome showing that CAM has evolved from re-wiring of existing C3 components through regulatory neofunctionalisation, rather than by acquisition of new genes with specialised CAM functions through duplication processes (Ming *et al.*, 2015). Additional evidence comes from the extent that the circadian expression of the CAM cycle can be disrupted when key metabolic enzymes have been manipulated (Dever *et al.*, 2015; Hartwell *et al.*, 2016; Boxall *et al.*, 2017). These are explored in more detail in a subsequent section of this review.

The availability of CAM genomes has provided a springboard for analysis of orthologues and shared transcriptional control elements (Ming *et al.*, 2015; Yang *et al.*, 2017; Wai *et al.*, 2019), whereas emerging comparative 'omics' analysis of various CAM species has also contributed to a more in-depth understanding of regulatory networks (Abraham *et al.*, 2016; Chiang *et al.*, 2016; Zhang *et al.*, 2016; Yin *et al.*, 2018; Heyduk *et al.*, 2019). Comparative transcriptomic studies have provided insights into the evolutionary trajectory of CAM and suggest that transcriptional regulation is primarily associated with specific expression profiles of key CAM genes. Additional analysis has listed candidate CAM-specific transcription factors which mediate transition from C3 to CAM mode in facultative (*M. crystallinum* and *Talinum triangulare*; Brilhaus *et al.*, 2016; Amin *et al.*, 2019; Maleckova *et al.*, 2019) and obligate (*Kalanchoe fedtschenkoi* and multiple *Agave* species; Moseley *et al.*, 2018; Yin *et al.*, 2018) CAM species. Finally, additional layers of control are also being explored at the level of post-transcriptional regulation via microRNA (miRNA) as well as less conventional mechanisms such as long non-coding RNA (lncRNA) which could function competitively with endogenous RNAs to alter the expression profile of key CAM pathway genes such as *PEPC* and *PPDK* (Yang *et al.*, 2015; Wai *et al.*, 2017; Bai *et al.*, 2019). Hence, computational modelling is becoming increasingly important to make sense of the explosion of data in the genomics era and not be deafened by the associated noise (Schatz, 2012; Fernie, 2016; Smita *et al.*, 2019; Yang *et al.*, 2019).

In this review, we set the context for certain fundamental questions concerning CAM that need to be resolved. We then suggest how newly available large-scale data can be aligned with various computational techniques, including systems models developed for CAM (Owen and Griffiths, 2013a; Maurice Cheung *et al.*, 2014; Hartzell *et al.*, 2018), to make more rapid progress in addressing fundamental physiological questions (Figure 1). By identifying key molecular targets associated with regulatory hubs controlling key elements of the CAM system, such an approach could be used to inform experimental genetic manipulation approaches to define these responses empirically. In order to develop these approaches, the focus of this review is first on the regulatory processes leading to control of stomata across the day–night cycle and secondly on mesophyll control of biochemistry, which is highly coordinated across the CAM phases. Finally, we shift the scale from the photosynthetic site to the whole-plant system by considering how hydraulic conductance, vasculature and succulence may have co-evolved to integrate stomatal and mesophyll processes, allowing the gradual metering of water loss, yet permitting rapid recharge, for CAM plant tissues experiencing intermittent precipitation events (Griffiths, 2013).

STOMATAL PHYSIOLOGY: SENSITIVITY AND SIGNALLING PROCESSES ACROSS THE CAM CYCLE

'What mechanisms underlie CAM stomatal behaviour of night-time opening, day-time closing and the flexible day-

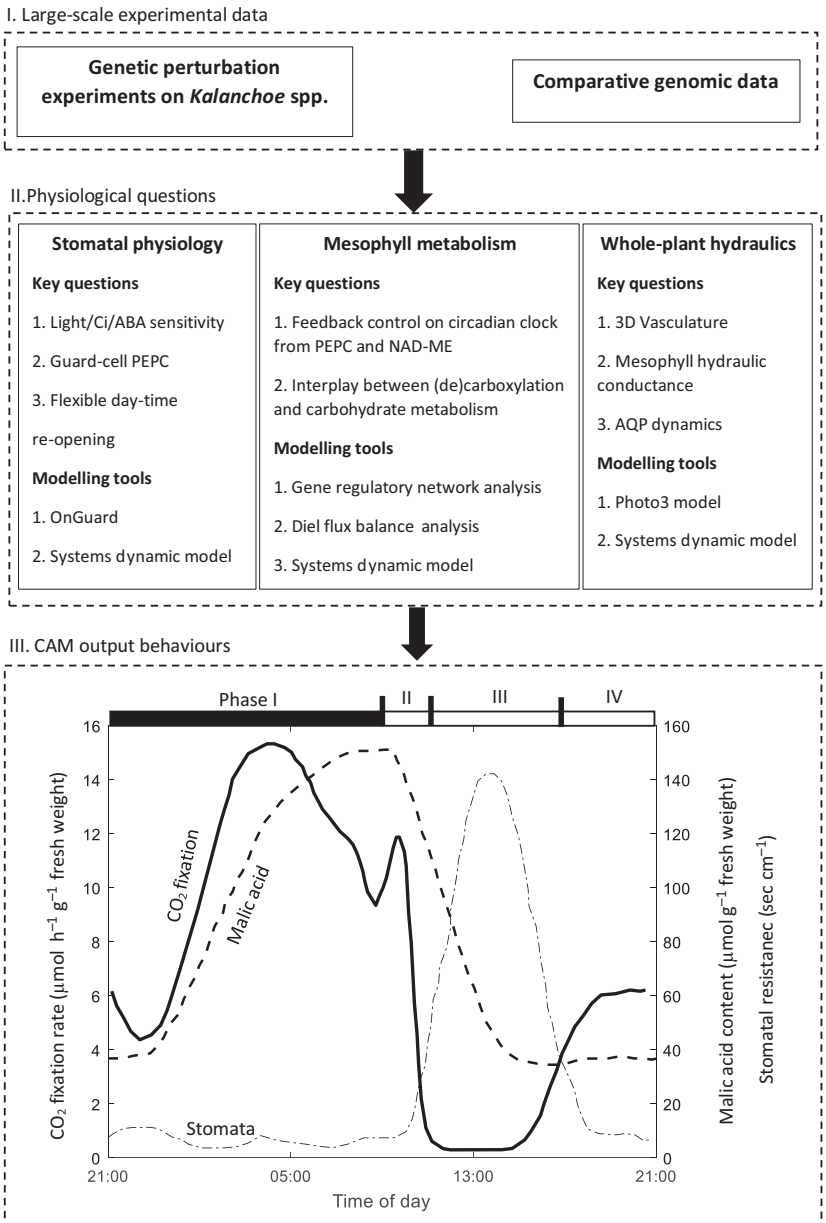


Figure 1. Summary of the three key areas for crassulacean acid metabolism (CAM) model development to address regulatory insights arising from recent genetic perturbation and integrated genomic approaches. Abbreviations: ABA, abscisic acid; AQP, aquaporin; Ci, intercellular CO_2; NAD-ME, NAD-dependent malic enzyme; PEPC, phosphoenolpyruvate carboxylase.

time re-opening in Phase IV?': this question was posed in 1978 in a classic article by C Barry Osmond (Osmond, 1978) and it remains an important, and as yet unanswered, problem (Osmond, 1978; Males and Griffiths, 2017; Borland *et al.*, 2018).

General models of C3 guard cell physiology suggest that stomatal opening responds to light (both blue and red wavelengths) and reduced atmospheric CO_2 concentration, whereas closure during the day is in response to elevated CO_2 and abscisic acid (ABA; Kim *et al.*, 2010; Assmann and Jegla, 2016). The magnitude of the responses can be tempered by environmental conditions (atmospheric vapour pressure and wind speed) as well as leaf and guard cell ultrastructure (boundary layers, guard cell dimensions and mechanical coupling with epidermal cells; Henry *et al.*, 2019). These generalised responses of C3 species are superimposed upon a potentially distinct guard cell circadian control system (Engineer *et al.*, 2016; Males and Griffiths, 2017; Hassidim *et al.*, 2017). To transition from diurnal stomatal opening in C3 mode to nocturnal stomatal opening in CAM mode, light-induced stomatal opening must either be blocked/downregulated or overridden by other cue(s) such as high intercellular CO_2 (Ci) or ABA to eliminate opening behaviour during the daytime. However, it is unclear how these canonical C3 pathways are replaced or rewired with additional

control mechanisms to allow for the stomatal behaviour seen with CAM (Kong *et al.*, 2019).

First, the evidence of the CAM stomatal response to light is still inconclusive. Initial reports on facultative CAM species (*M. crystallinum* and *Portulacaria afra*) suggested that stomata no longer respond to light signals once induced into CAM mode (Lee and Assmann, 1992; Tallman *et al.*, 1997). However, the stomata of obligate CAM species (*Kalanchoe pinnata* and *K. daigremontiana*) have been shown to open in response to blue light, similar to C3 stomata (Gotoh *et al.*, 2019). These contradictory observations provide a focus for the question originally posed by Osmond (1978) and the interplay between light cues and the CAM pathway. In addition, it also highlights the potential difference in regulatory mechanisms and evolutionary trajectory between facultative and obligate CAM species, which have gone through environmental induction and ontogenetic induction, respectively (Winter, 2019).

Secondly, nocturnal opening behaviour of stomata in CAM is hypothesised to be mediated by low Ci as a result of PEPC activity, but no precise mechanism has been elucidated to date (Borland and Griffiths, 1997; Drennan and Nobel, 2000; Ceusters *et al.*, 2008; Von Caemmerer and Griffiths, 2009). The fluctuation of Ci associated with the two distinct CAM carboxylation processes, coupled to decarboxylation, can result in significant fluctuations of Ci over the range of 0.011% to 0.5% with a defined diel pattern (Borland and Taybi, 2004). For *Kalanchoe* species with contrasting degrees of succulence, this 24-h pattern of Ci synchronised well with the night-time opening/daytime closing behaviour and the underlying photosynthetic activity (Wyka *et al.*, 2005; Griffiths *et al.*, 2007, 2008; Males and Griffiths, 2017). However, the presence of a Ci signalling pathway operating across the 24-h cycle has yet to be verified. The transcriptomic work on *Agave americana* has shown that genes involved with CO_2 sensing, such as *HT1*, have an inverted expression pattern to that of a C3 plant. This suggested that the stomatal responsiveness to Ci of *A. americana* might not be intact throughout the 24-h cycle (Abraham *et al.*, 2016). Nonetheless, the intermediate components of the Ci signalling pathway are yet to be identified (Engineer *et al.*, 2016; Hsu *et al.*, 2018). Moreover, it may be that responses to internal CO_2 supply have different sensing/signalling pathways depending on whether Ci is decreasing or increasing. Hence a more systematic identification and analysis of a Ci signalling pathway in a comparative C3–CAM system is required at this stage.

Thirdly, it is not clear whether guard cell PEPC activity at night contributes to the stomatal behaviour in CAM species by providing malate as one of the major solutes driving the opening process (Santelia and Lawson, 2016; Males and Griffiths, 2017). In C3 species, malate accumulation as a result of PEPC activity during the day is critical for the light-induced stomatal opening process (Asai *et al.*, 2000). Evidence from a transgenic RNA interference (RNAi) experiment on the CAM species *Kalanchoe laxiflora* has recently shown that stomata revert back to C3 daytime opening in a *PEPC* knock-down line (Boxall *et al.*, 2019). However, because *PEPC* was knocked down in all tissues in this RNAi experiment, the reduced stomatal conductance cannot be attributed solely to either the internal metabolism of guard cells or Ci signals generated from mesophyll metabolism. Further investigations could use a more targeted genetic approach, such as a *PEPC* knock-down line, but with a guard cell-specific promoter.

Finally, ABA has been implicated in the upregulation of CAM gene expression and enzymatic activities, although the direct effect of ABA on the inverted timing of CAM stomatal behaviour remains elusive (Holtum and Winter, 1982; Chu *et al.*, 1990; Bastide *et al.*, 1993; Dai *et al.*, 1994; Taybi *et al.*, 1995; Taybi and Cushman, 1999). The analysis of transcriptomic data from *Agave* showed that ABA biosynthesis and signalling components are enriched in the expression cluster that shows peak expression at 0 and 3 h before dawn (Abraham *et al.*, 2016). This suggested a potential role for synthesis and perception of ABA at the precise moment before CAM plants enter Phase III stomatal closure. Emerging evidence has also proposed candidate components that may link ABA to CAM stomatal responses through transcriptional regulation (Lian *et al.*, 2018; Maleckova *et al.*, 2019), reactive oxygen species signalling and/or protein kinase activities (Moseley *et al.*, 2019b). These details require integration into a framework for stomatal signalling and further functional verification tests.

Recent contributions to address these questions have come from comparative transcriptomic approaches and functional genomics (Fernie, 2016; Abraham *et al.*, 2016; Wai and VanBuren, 2018; Liu *et al.*, 2019; Moseley *et al.*, 2019b). Transcriptomic studies elucidated genes that have transcript abundance patterns with inverted timing in CAM species when compared with their orthologues in C3 *Arabidopsis thaliana*. These genes encode for stomatal signalling components and transporters in *Agave* spp. (e.g. orthologues of *OST1, HT1, PP2C, RCAR3*, K^+ transporter *AKT2/3*, chloride channel *CLC-c*, Ca^{2+} transporter *ACA2* and *ECA4*), stomatal signalling components in pineapple (e.g. orthologues of *BLUS1, HT1, CPK6*), an H_2O_2 detoxification enzyme (orthologue of *CAT2*) implicated in the stomatal signalling process and a blue-light receptor (*PHOT2*) in *K. fedtschenkoi* (Abraham *et al.*, 2016; Yang *et al.*, 2017; Wai and VanBuren, 2018; Moseley *et al.*, 2019b). Among these genes, only *PHOT2* has been shown through a functional study to have a direct implication for CAM stomatal behaviour in *K. fedtschenkoi* (Liu *et al.*, 2019). However, the interplay between signalling components and transporters is not always intuitive and would require computational modelling tool(s) to interpret (see the next section).

In addition, an important caveat concerning the comparative transcriptomic approach is that transcript abundance is not guard cell specific; rather, it represents the integrated pattern across the whole leaf.

A recent genetic perturbation experiment also addressed this question by investigating the relative gene expression profiles in guard cell-enriched epidermal tissues. In this experiment, differential regulation and phasing of key genes involved in stomatal responses were associated with the shift to daytime CO_2 fixation in a PEPC knock-down line (Boxall *et al.*, 2019). The authors concluded that the differential regulation of *HT1*, *SKOR* and *MYB61* genes (upregulated) compared with *SLAC1*, *PPC2C, SOS2*, *ACA2*, *ECA4* and *MYB60* (downregulated) in the wild-type line were consistent with nocturnal opening responses and light period closure during CAM (Boxall *et al.*, 2019). In parallel, guard cell-specific single-cell transcriptomic profiling of *M. crystallinum* has suggested that specific transcriptional reprogramming of 495 transcripts (including 18 transcription factors, 285 guard cell genes and 7 CAM metabolism genes) could be the underlying mechanism for the transition from C3 to CAM stomatal behaviour (Kong *et al.*, 2019). Such guard cell-specific insights could also feed into computational analyses, although it is worth noting that transcript abundance does not necessarily reflect protein activity.

How can existing modelling platforms help?

Possible roles of the inverted gene expression profiles could be further investigated with modelling approaches and subsequent functional tests based around stomatal signalling pathway(s) and solute transporter networks. Starting from the canonical C3 stomatal signalling models, the specific CAM transcriptional timing can be superimposed onto the components of the signalling pathway (s) to investigate if any specific signalling pathways are up/downregulated coinciding with the four phases of CAM. Concomitantly, timing of the expression of CAM guard cell transporters can be mapped onto the biophysical modelling platform 'OnGuard' to develop a specific CAM version (Hills *et al.*, 2012). However, gene expression data alone do not provide conclusive evidence about the time during which a particular signalling pathway is operating. Genes that are not expressed simply yield missing model components, whereas the protein abundance and activity for genes that are expressed cannot simply be inferred until other layers of control (e.g. post-translational modification or protein turnover) have also been measured (Yang *et al.*, 2015; Fernie, 2016). In addition, the effect of redundancy or alternative pathways that might appear exclusively in CAM must also be taken into account. Therefore, more data can be extracted from comparative transcriptomic/proteomic approaches to feed into signalling/biophysical modelling platforms to address the specific timings of CAM stomatal sensitivity to various stimuli.

Alternatively, systems dynamic (SD) computational modelling (Owen and Griffiths, 2013a) has provided a framework for studying the role of Ci in governing CAM stomatal behaviour. The current version of the model can be improved by using the empirical data on CAM stomatal responses and Ci to construct the stomatal response module in the SD model (Bohn *et al.*, 2001). As a result, the entire Ci signalling pathway that is not fully characterised in CAM can be by-passed using the experimental Ci pattern across a 24-h period as the input into a black box which generates stomatal conductance responses as outputs. These stomatal conductance outputs, together with the CO_2 gradient, would then dictate the CO_2 flux from the atmosphere into the mesophyll and identify the extent to which the observed Ci pattern was a direct cause of stomatal opening or an effect of guard cell or mesophyll processes (Sakamoto *et al.*, 2015). If the minimal model with only a Ci cue fails to capture CAM stomatal behaviour throughout the four phases, alternative control loops such as light (in)sensitivity, internal guard cell metabolism and/or ABA sensitivity must be added to the model to better reflect physiological scenarios.

MODELLING APPROACHES CAPTURE THE COMPLEX CONTROL OF CAM MESOPHYLL PROCESSES

Mesophyll metabolism in CAM integrates complex traits, coupled via feedback effects, and control of gene expression (Figure 2). To ensure that such a complex interlocking system is amenable to modelling, CAM mesophyll processes can be subdivided into three metabolic modules: carbon assimilation (carboxylation), malic acid storage and processing (decarboxylation) and carbohydrate supply and demand. It is important to understand the control mechanisms which ensure the accurate activation and precise timing needed to synchronise carbon flow between these three pools. One of the most important control points is PEPC which integrates circadian and metabolic control in the CAM system. However, more recent evidence from genetic perturbation experiments and comparative genomics studies has suggested other compelling candidate control points that can be analysed further using existing CAM mesophyll models.

To sustain the diel pattern of CAM mesophyll metabolism, circadian controls set the diel phases of carboxylation processes while metabolites provide another layer of refinement (Borland *et al.*, 2016). Circadian regulation of the activation of key enzymes, including the carboxylases PEPC and RubisCO, has been shown to play an important role in governing the temporal separation of CAM photosynthetic activity (RubisCO: Griffiths *et al.*, 2002; Maxwell *et al.*, 2002; Davies and Griffiths, 2012; PEPC: Carter *et al.*, 1991; Hartwell *et al.*, 1999; Dodd *et al.*, 2002). The most

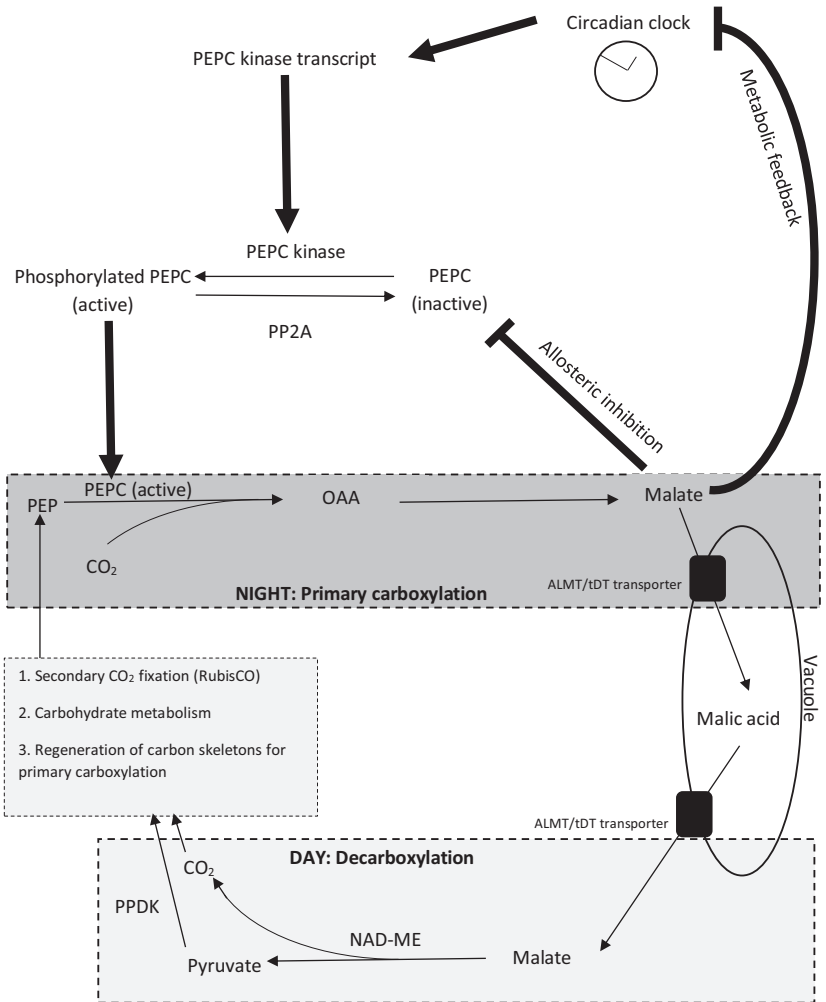


Figure 2. A simplified diagram of the mesophyll metabolic processes including key enzymes and the integrated control between circadian clock and metabolic feedback of crassulacean acid metabolism (CAM) species that operate via the NAD-ME/PPDK decarboxylation pathway. Key to metabolites and enzymes/transporters: ALMT, aluminium-activated malate transporter; NAD-ME, NAD-dependent malic enzyme; OAA, oxaloacetate; PEP, phosphoenol-pyruvate; PEPC, phosphoenol-pyruvate carboxylase; PP2A, protein phosphatase 2A; PPDK, pyruvate orthophosphate dikinase; tDT, tonoplast dicarboxylate transporter. Metabolic feedback is represented with malate effect; however, the exact identity of the metabolite which exerts control on the circadian clock remains inconclusive and could be other molecules, including but not limited to sucrose (discussed in text).

prominent integration point between circadian control and metabolite control is at PEPC. As the primary CO_2-fixation enzyme, PEPC is phosphorylated by PPCK at night, leading to reduced sensitivity to malate allosteric inhibition, whilst the abundance of PPCK itself is under transcriptional control from circadian clock outputs (Carter *et al.*, 1991; Hartwell *et al.*, 1999). In addition, PPCK gene expression can be modified, with increased malate accumulation exerting negative feedback on PPCK gene expression, possibly as a secondary effect via the circadian clock (Borland *et al.*, 1999; Nimmo, 2000; Figure 2).

Experimental evidence: manipulation of gene expression

Controls of CAM mesophyll metabolism extend beyond the carboxylation module (Borland and Taybi, 2004). Firstly, it is possible that other mesophyll enzymes/transporters are under circadian control or can be modified by CAM metabolites through similar secondary circadian responses to those proposed for PPCK (Borland *et al.*, 1999; Cushman, 2001; Davies and Griffiths, 2012; Dever *et al.*, 2015; Hartwell *et al.*, 2016). Recent transgenic manipulation experiments suggest a close link between circadian and metabolite control in CAM. Downregulation of the carboxylation activity (PEPC) or decarboxylation activity [mitochondrial NAD–malic enzyme (NAD-ME)] has been shown to disrupt the rhythmicity of CAM in other mesophyll components distant from the knockdown point, in addition to altering the oscillation pattern of circadian clock genes through unidentified metabolic feedbacks (Dever *et al.*, 2015; Boxall *et al.*, 2019).

In terms of carboxylation, the regulation of PEPC has recently been investigated through two RNAi experiments. In the first approach, the *PEPC kinase* gene of *K. fedtschenkoi* was silenced, resulting in reduced dark phosphorylation of PEPC and a 66% reduction in nighttime CO_2 assimilation (Boxall *et al.*, 2017). This study emphasised the importance of PEPC phosphorylation for maintaining PEPC activity through the dark period, which in turn affected the stability of the circadian clock and overall CAM productivity. Subsequently, this group developed

transgenic lines of *K. laxiflora* in which the *PEPC* gene was downregulated directly (Boxall *et al.*, 2019). The plants suffered a complete loss of dark CO_2 assimilation, fixing CO_2 directly only in the light, and the associated impacts on the regulation of stomatal aperture have been discussed above. Additionally, transcripts for PPCK were downregulated at night and peaked at the start of the light period, and gas exchange suffered arrhythmia under constant light, despite the transcript abundances of some core circadian clock genes being enhanced (Boxall *et al.*, 2019).

In terms of the decarboxylation component of the CAM cycle, the RNAi approach has been used to reduce the activities of two CAM enzymes, NAD-ME and pyruvate orthophosphate dikinase (PPDK), both of which are involved in the decarboxylation of malate and release of CO_2 internally during the light period (Figure 2) (Dever *et al.*, 2015). The expression of CAM was suppressed in both sets of plants, which tended to take up the majority of CO_2 directly in the light. In addition, there was a reduction in transcripts of many genes associated with the CAM cycle, among which were *PEPC* and genes associated with storage carbohydrate trafficking, including *PPDK* (in a *NAD-ME* transgenic line), glucan water dikinase and glucose-6-phosphate translocator. Whilst the expression patterns of *PPCK* and core clock gene transcripts were not affected, the circadian cycle became arrhythmic under continuous light, again suggesting a role for metabolite feedback in resetting the CAM cycle on a daily basis (Dever *et al.*, 2015).

Despite not targeting the carbohydrate metabolic process directly, the aforementioned RNAi experiments have provided insights into the regulatory processes associated with CAM carbohydrate metabolism. Firstly, starch accumulation, which normally precedes the dark period to provide carbon skeletons for carboxylation, was generally reduced in the transgenic line with perturbed carboxylation–decarboxylation processes (Dever *et al.*, 2015; Boxall *et al.*, 2017). Secondly, transcripts of genes associated with the partitioning of starch in C3 systems, such as the amylolytic pathway at night, were upregulated in the *K. laxiflora* transgenic lines with reduced PEPC activity (Boxall *et al.*, 2019) compared with the phosphorylytic route normally associated with this starch-storing CAM species (Borland *et al.*, 2009). Finally, there was also a distinct correlation between the peaks of sucrose and the expression profiles of CAM orthologues of circadian clock components in the *K. fedtschenkoi* line with reduced PEPC kinase activity, suggesting sucrose as a potential candidate for linking CAM carbohydrate metabolic status back to the circadian clock operation (Boxall *et al.*, 2017).

Experimental evidence: gene expression and transcriptomic approaches

Comparative transcriptomic studies have provided a list of genes with specific CAM expression patterns with timing either phase-shifted or inverted between C3 and CAM. These genes with specific CAM expression patterns can be considered 'candidates' that may play important roles in synchronising CAM mesophyll processes. Such large-scale comparative data are useful resources for bioinformatics and network analysis for grouping genes into expression modules under candidate transcription factors, although functional tests need to be conducted to infer the physiological influence on CAM behaviour.

Comparative transcriptomic studies have associated the evolution of CAM with specific regulation of transcript abundance patterns. Firstly, comparative studies within 13 orchid species suggested that regulation of the transcription level of carbon fixation pathway genes (*PEPC, PPCK* and *PPDK*) was associated with the evolution of CAM within this clade (Zhang *et al.*, 2016). Secondly, transcriptomic analysis of two cultivated pineapple varieties and one wild pineapple relative also showed that CAM-related genes demonstrated an inverse diel expression pattern in photosynthetic tissues (Ming *et al.*, 2015). With the available pineapple genome data, they also suggested that CAM evolved not by gene duplication but through the modified expression of existing C3 genes. In addition, a key mechanism that controls timing of the expression of CAM-specific genes has been suggested to be the circadian clock, because the differentially regulated CAM pathway genes were enriched with clock-associated *cis*-regulatory elements (Ming *et al.*, 2015). Although the presence of *cis*-regulatory elements suggested that a transcriptional control process is in place, the contribution of post-transcriptional control must not be neglected. Emerging evidence from studies of miRNA and lncRNA in pineapple has also offered additional control mechanisms for several CAM-related genes at the post-transcriptional level (Wai *et al.*, 2017; Bai *et al.*, 2019).

Candidate CAM regulatory points embedded in carbohydrate and malate metabolism can be as important as the control of the carboxylation module discussed above. Firstly, *Kalanchoe* genome analysis provided evidence for contrasting differential regulation of carbohydrate processing between CAM and C3 species. A network analysis compared families of carbohydrate 'active' gene families which were similar in number to those found in Arabidopsis but with altered expression profiles, with a possible role for a trehalose-6-phosphate synthase and an invertase, acting to alter carbohydrate partitioning between substrates for CAM and those needed for growth (Yang *et al.*, 2017). Secondly, putative orthologues of chloroplast and vacuolar sugar transporters of pineapple have adopted a defined diel expression pattern, and are hence implicated in the diel processing of carbohydrate in this soluble-sugar-processing CAM species (Borland *et al.*, 2016). Thirdly, genomic studies of transcriptional patterns have also supported that malate channel aluminium-activated malate

transporter (ALMT) could be one of the key candidate control points to allow for CAM evolution and function in both pineapple and *Talinum* (Brilhaus *et al.*, 2016; Wai and VanBuren, 2018). Finally, it is well established that CAM induction in *M. crystallinum* is associated with increased in-transcript abundance of a chloroplast carbohydrate transporter gene associated with a higher glucose-6-phosphate transport rate in its CAM mode (Cushman *et al.*, 2008).

Computational approaches needed to define metabolic and environmental regulation of CAM

Top-down modelling approaches in terms of network analysis have been adopted to group candidate genes into modules according to the specific diel expression patterns. A transcriptome module network analysis could be used to construct CAM-specific regulatory hubs on obligate CAM species under different environmental stress manipulations and/or on facultative CAM species, with the time series sampled through the CAM induction process (Ming *et al.*, 2015; Abraham *et al.*, 2016; Brilhaus *et al.*, 2016; Yang *et al.*, 2017; Maleckova *et al.*, 2019) To some extent, the pioneering study by Ming *et al.* (2015) has already demonstrated the power of comparative gene expression network analysis from a comparison of circadian clock motifs between pineapple, orchid, rice and maize. They also presented a gene regulatory network analysis for green tissues, highlighting genes and their interacting partners associated with CAM. Yang *et al.* (2017) also undertook a comparative gene regulatory network analysis, firstly for phase shifts in gene expression and also for heat shock proteins between *Kalanchoe*, pineapple and Arabidopsis. It is also possible to take a further step with a network inference algorithm to infer the distinct mechanisms of circadian control of CAM-related genes within each expression module (Moseley *et al.*, 2019a). Overall, these top-down network analysis approaches still suffer from the limitation of unannotated components within CAM genomes and unidentified regulatory processes. This suggests that the construction of CAM regulatory networks must be continuously updated as more experimental data become available. Nonetheless, network analysis alone would not address the functional significance of any regulatory process until functional tests have been conducted.

A bottom-up approach builds on existing mechanistic models (Owen and Griffiths, 2013a; Hartzell *et al.*, 2018) and metabolic network models (Maurice Cheung *et al.*, 2014; Shameer *et al.*, 2018). Such models posit a null state, in which the only a minimal circadian and/or metabolic regulation is imposed on selective model components. At this stage, the experiments reporting manipulated gene expression and associated physiological outputs are fruitful resources for developing improved versions of the existing CAM mesophyll metabolic models (Dever *et al.*, 2015; Boxall *et al.*, 2017). These gene manipulation experiments can be regarded as 'test' datasets, whereas the models have previously been 'trained' against the unperturbed condition. With this approach, we are looking for the simplest model that captures CAM mesophyll metabolic behaviour based on existing physiological and enzyme kinetic data, visualised as the goodness of fit to diverse gas exchange phase profiles. If the models can capture CAM behaviour of these genetically manipulated data, it would suggest that the minimal control processes already built in to these models are robust enough for CAM to function. In contrast, if any metabolic module of the models fails to capture modified CAM outputs, candidate control loops can be modified or augmented with the information drawn from comparative genomic data and biochemistry of allosteric effects of metabolites. The advantage of this bottom-up approach is that it does not require prior knowledge of an entire transcription-gene regulatory network to capture CAM functions. Instead, it allows plant physiologists to focus on the functional layer of the CAM system in terms of enzyme/transporter activities.

An example of the simplest CAM control loop is the interplay between the carboxylation–decarboxylation pattern and carbohydrate storage. Firstly, mesophyll CAM activity is dominated by PEPC and RubisCO that influence Ci, which in turn alters stomatal responses, the potential carboxylation rate (by day or night) and carbohydrate production. Subsequently, carbohydrate availability then determines the amount of starting material for the next round of CO_2 fixation, and so on (Borland *et al.*, 2016, 2018). If this simplest loop is not enough to explain CAM behaviour, then a more complex loop needs to be introduced, based on the regulatory gene expression described above which controls circadian clock gene expression, carboxylase activity or carbohydrate partitioning. Nevertheless, one must be aware that there are multiple levels of control. Genes that have not adopted a CAM-specific diel expression pattern can have a crucial role in CAM function if the protein activities are modified at other levels of regulatory processes, including but not limited to post-translational modifications (Yang *et al.*, 2015; Zhu *et al.*, 2018).

Ultimately the top-down and bottom-up approaches converge, in allowing identification of specific gene knockdown/knockout targets which could be introduced into tractable transformation systems, such as *Kalanchoe* (Hartwell *et al.*, 2016; Liu *et al.*, 2019). Firstly, the refinement of the pre-existing mesophyll mechanistic models would enable systematic *in silico* quantification of the effects that multiple candidate control points have on CAM physiological outputs. Subsequently, experiments must be carefully designed to represent a subset of *in silico* scenarios which significantly affect CAM behaviour in terms of

carbon assimilation pattern, malate accumulation or carbohydrate metabolism. With additional information about the gene regulatory network, the most effective genetic manipulation strategy may be achieved by targeting a specific set of key CAM regulatory process (e.g. transcription factors, metabolic signal integration point or specific CAM circadian genes). More importantly, a better understanding of regulatory networks would ensure that the targeted genetic manipulation points are selected to not interfere with the fundamental cellular processes that may lead to an unfavourable pleiotropic effect. Finally, genetic perturbations can be introduced into a transformation system to perform functional tests. The responses to the genetic perturbation can be used to validate the models and potentially be informative for a future attempt at CAM biodesign (Lim *et al.*, 2019). In the end, the combination of network analysis, mechanistic models and functional tests would reveal how a given component of the CAM system is wired with the central CAM regulatory network which would inform subsequent genetic manipulations and mitigate against biased component selection or undetectable network readjustments.

THE COMPROMISE BETWEEN CARBON UPTAKE AND WATER LOSS: HYDRAULIC CONSTRAINTS WITHIN CAM SHOOT AND ROOT SYSTEMS

If understanding the biochemical integration of CAM carbon assimilation can be addressed by integrating molecular and systems approaches (see above), control of water uptake and use will require bottom-up models to capture the biophysical elements of water supply, storage and evaporative demand in the succulent tissues (Bartlett *et al.*, 2014; Hartzell *et al.*, 2015, 2018). The relevance and context for this section are provided by two contrasting themes evident in recent major reviews of CAM processes – one being the evolutionary trajectory associated with the diversification of CAM phylogenetically over the past 5 million years (Edwards, 2019) and the second the notion that CAM could be engineered into C3 crops or trees to engender resilience to water deficits under the climate change (Yang *et al.*, 2015; Borland *et al.*, 2018).

There are three competing views regarding the evolution of succulent tissues and CAM metabolic activities. The key points have been captured by Edwards (2019), who questions whether strong CAM was needed as a response to photosynthetic limitation (low mesophyll conductance) in dense, succulent tissues which had evolved to enhance water storage and hence drought tolerance. Alternatively, the enhanced water-use efficiency associated with CAM would need large vacuolate cells to provide a nocturnal storage reservoir for the accumulated malic acid, so succulence could have been a secondary trait. Finally, she suggests that water storage in succulent tissues and high water-use efficiency in CAM were co-selected (Edwards, 2019).

The answer to these questions may not lie in a single linear progression from a typical C3 leaf (high air space, low succulence, facultative CAM) to strong CAM in dense, succulent tissues, particularly given recent findings on the extent of intermediate C3–CAM forms (Winter *et al.*, 2015; Earles *et al.*, 2018; Edwards, 2019). Such evidence might be compelling within single clades of annual or perennial herbs with similar growth forms, or when inferred from indirect proxies such as air spaces and vasculature (Earles *et al.*, 2018; Males and Griffiths, 2018) or carbon isotope compositions (Edwards, 2019). However, different evolutionary trajectories could have been associated with the origins of CAM in contrasting leaf succulence traits, which range across annual herbs, perennial leaves (planar or massive) or stems, with differing arrangements of chlorenchyma and water storage parenchyma (hydrenchyma), as well as their associated two- or three-dimensional (3D) vasculatures (Ogburn and Edwards, 2013; Griffiths, 2013; Griffiths and Males, 2017). Currently, we do not have enough comparative studies between C3 and CAM members within a single family, comparing both gas exchange and hydraulic traits, to infer cause and effect in this evolutionary progression (Borland *et al.*, 2018; Males and Griffiths, 2018), particularly for the 'massive' leaf and stem succulent families (e.g. Agavaceae, Euphorbiaceae and Cactaceae). Hence, in parallel with the rapid progress in generating comparative transcriptomic studies, comparative hydraulic trait measurements from similar clades would provide invaluable information.

Modelling approaches have been developed to capture hydraulic aspects of CAM. Firstly, the systems dynamic modelling of Owen and Griffiths (2013a) demonstrated that the plasticity in the expression of CAM phases could be captured in a single model when applied to both extreme forms of leaf succulence. For massive leaf succulents such as *Agave* transient gas exchange during Phases II and IV is typically limited in laboratory and field studies (Nobel and Valenzuela, 1987; Owen *et al.*, 2016). In contrast, for the genus *Kalanchoe*, where leaves consist of homogeneous chlorenchyma, the expression of these phases is more extensive and has been related to the degree of succulence (tissue density and airspaces) (Griffiths *et al.*, 2008). However, the Owen and Griffiths (2013a) model was able to capture the transition from *Agave* to *Kalanchoe* gas exchange profiles by altering mesophyll conductance and associated biochemical constraints (in terms of PEPC and RubisCO activities and their activation timing). Alternatively, a coupled carbon–water flux model has incorporated the classical Farquhar–von Caemmerer–Berry photosynthetic model with soil–plant–atmospheric continuum (SPAC) components (Farquhar *et al.*, 1980; Bartlett *et al.*, 2014; Hartzell *et al.*, 2018). This shows promise for integrating the carbon–water balance in CAM plants and successfully captured CAM phases and their responses under

progressive droughting conditions (Bartlett *et al.*, 2014). The latest iteration is able to compare C3, C4 and CAM systems using the Photo3 modelling framework (Hartzell *et al.*, 2018).

More aspects of specialised CAM hydraulic traits still need to be incorporated into the model or functionally tested. Firstly, tissue differentiation in succulent CAM systems has allowed a range of chlorenchyma–hydrenchyma associations to develop, in which the hydrenchyma may be integrated more or less directly above, within or between chlorenchyma cell layers (Griffiths and Males, 2017). Both cell types with large vacuoles can contribute to overall leaf water storage (capacitance) but there may be a differential regulation of cellular biophysical traits such as the bulk modulus of elasticity and half-life for water exchange (as reviewed recently by Borland *et al.*, 2018). For some systems, the distinct hydrenchyma layers may preferentially lose water to support overall leaf or cladode transpiration, which in *Agave* was demonstrated by an electrical analogue model (Smith *et al.*, 1987). Secondly, it is also becoming evident that we need to partition the overall hydraulic conductance into xylem (Kx) and mesophyll (Kox) components to sites of evaporation (Scoffoni *et al.*, 2018; Males and Griffiths, 2018). Whilst even in C3 leaves it is increasingly found that Kox represents a major limitation in the transport of water (Scoffoni *et al.*, 2018), this is also likely to be true for succulent CAM leaves given the relatively high whole-leaf water potentials (about −1 MPa) retained even during an extensive drought (Borland *et al.*, 2009; Griffiths, 2013; Males and Griffiths, 2018; North *et al.*, 2019). Finally, the likely susceptibility to xylem cavitation (i.e. embolisms) under more extreme water potentials in CAM has recently been linked to the absence of secondary lignification generally across CAM systems (Borland *et al.*, 2018).

The more recent application of 3D tomographic scanning techniques, used to define intracellular air spaces, as well as the distribution of primary and secondary vasculature (Earles *et al.*, 2018; North *et al.*, 2019) can also be coupled to reveal the proportion of embolised conduits (Nolf *et al.*, 2017; North *et al.*, 2019). Empirical observations of this sort will help to compartmentalise the hydraulic limitations across CAM systems, from rain roots through to stem and leaf xylem to mesophyll tissues, which will then allow models to capture the co-regulation of these stages in water transfer. Additional molecular evidence on the role of aquaporins in regulating water transport in roots or shoots and any associated ABA signalling (North *et al.*, 2004; Sade *et al.*, 2014) would also allow the dynamics of water transport to be reflected in future tissue-specific transcriptomic studies.

To conclude this consideration of hydraulic limitations associated with succulent tissues, it now seems generally accepted that CAM represents a drought avoidance mechanism, with mature, sclerified roots and their shrinkage being the accepted mechanism to prevent loss of water to drying soils (Nobel and Valenzuela, 1987; Griffiths, 2013; Borland *et al.*, 2018). One further question remains as yet untested: whether the development of 3D vasculature in many succulent systems permits more rapid water uptake to commence following intermittent precipitation events (Griffiths, 2013). Coupling the molecular expression of aquaporins between xylem tissues, mesophyll hydrenchyma and chlorenchyma would maximise the recharge of capacitance for stem or leaf tissues via the 3D vein network. The evidence that *Agave* and *Opuntia* can take up nearly 50% of a major precipitation event (Nobel and Valenzuela, 1987) is consistent with succulent plants being restricted to semi-arid areas where rainfall is seasonally predictable. Therefore, CAM, in the broadest sense, has perhaps necessitated the development of stomatal, mesophyll and xylem hydraulic processes associated with optimising both carbon assimilation and water uptake for succulent tissues.

CONCLUSIONS

How variations upon CAM themes are regulated has continued to engender fascination in terms of fundamental research and more translational applications for bioenergy or biomass production (Borland *et al.*, 2018; Edwards, 2019). There has been a dramatic growth in datasets defining molecular genetics, phylogenetics and mechanistic processes associated with the CAM cycle across varying degrees of succulence and varying levels of commitment to the CAM pathway (be it strong or constitutive CAM or C3–CAM intermediate states) (Edwards, 2019; Winter, 2019). Such approaches within specialised CAM groups are important, but increasingly we need more generalised approaches to capture the key regulatory networks and signalling effectors associated with the expression and maintenance of CAM (Dever *et al.*, 2015; Ming *et al.*, 2015; Abraham *et al.*, 2016; Boxall *et al.*, 2017, 2019).

In this review, we have identified the need to develop more rigorous computational modelling approaches which can capture the information within huge transcriptomic datasets. Careful comparative approaches are needed to sift and distil this information into gene regulatory networks, and thence identify targets for manipulation and assessment of resultant phenotypes. Furthermore, we recognise the need for the additional development of systems models and/or metabolic network models which could then predict or capture the novel phenotypic traits arising from molecular transformation and more traditional physiological manipulations (Owen and Griffiths, 2013a; Maurice Cheung *et al.*, 2014; Hartzell *et al.*, 2018). Ultimately, key questions which still endure in terms of the effect of interaction between circadian and environmental effectors on the guard cell and mesophyll processes may also be resolved in the long run.

Similar approaches could also be used to model phylogenetic transitions within major CAM families and address the 'cause or effect' of tissue succulence in driving the commitment to CAM shown across the CAM trait space (Males and Griffiths, 2017; Edwards, 2019). Here, we stress the need for CAM syndromes to be assessed in their entirety: the regulation of CO_2 supply and evaporative demand as expressed by stomata, the mesophyll metabolic processes and the limitations imposed by low mesophyll conductance in dense tissues with low airspaces. These aspects need to be integrated with how water supply is monitored, metered and recharged when intermittent external sources are available to roots or leaf rosettes (Griffiths, 2013; Males and Griffiths, 2018; North *et al.*, 2019). We hope these challenges will continue to attract future generations of researchers who can identify the peculiar attraction required to 'work the night shift' with CAM (Black and Osmond, 2003) and determine the extent that CAM systems could contribute to more sustainable biomass production in a changing world.

ACKNOWLEDGEMENTS

We are grateful for the perceptive promptings of C. Barry Osmond, who in his 80th year continues to inspire CAM researchers. We are thankful for the helpful comments of two reviewers. This research was supported by a Thai government scholarship to MC.

AUTHOR CONTRIBUTIONS

Both authors contributed equally to the work.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interests.

DATA AVAILABILITY STATEMENT

All relevant data can be found within this manuscript with the citations to primary literature.

REFERENCES

Abraham, P.E., Yin, H., Borland, A.M. *et al.* (2016) Transcript, protein and metabolite temporal dynamics in the CAM plant Agave. *Nat. Plants*, **2**, 16178.

Adams, P., Nelson, D.E., Yamada, S., Chmara, W., Jensen, R.G., Bohnert, H.J. and Griffiths, H. (1998) Growth and development of *Mesembryanthemum crystallinum* (Aizoaceae). *New Phytol.* **138**, 171–190.

Amin, A.B., Rathnayake, K.N., Yim, W.C., Garcia, T.M., Wone, B., Cushman, J.C. and Wone, B.W.M. (2019) Crassulacean acid metabolism abiotic stress-responsive transcription factors: a potential genetic engineering approach for improving crop tolerance to abiotic stress. *Front. Plant Sci.* **10**, 129.

Asai, N., Nakajima, N., Tamaoki, M., Kamada, H. and Kondo, N. (2000) Role of malate synthesis mediated by phosphoenolpyruvate carboxylase in guard cells in the regulation of stomatal movement. *Plant Cell Physiol.* **41**, 10–15.

Assmann, S.M. and Jegla, T. (2016) Guard cell sensory systems: recent insights on stomatal responses to light, abscisic acid, and CO_2. *Curr. Opin. Plant Biol.* **33**, 157–167.

Bai, Y., Dai, X., Li, Y., Wang, L., Li, W., Liu, Y., Cheng, Y. and Qin, Y. (2019) Identification and characterization of pineapple leaf lncRNAs in crassulacean acid metabolism (CAM) photosynthesis pathway. *Sci. Rep.* **9**, 6658.

Bartlett, M.S., Vico, G. and Porporato, A. (2014) Coupled carbon and water fluxes in CAM photosynthesis: modeling quantification of water use efficiency and productivity. *Plant Soil*, **383**, 111–138.

Bastide, B., Sipes, D., Hann, J. and Ting, I.P. (1993) Effect of severe water stress on aspects of crassulacean acid metabolism in Xerosicyos. *Plant Physiol.* **103**, 1089–1096.

Black, C.C. and Osmond, C.B. (2003) Crassulacean acid metabolism photosynthesis: 'working the night shift'. *Photosynth. Res.* **76**, 329–341.

Bohn, A., Geist, A., Rascher, U. and Luttge, U. (2001) Responses to different external light rhythms by the circadian rhythm of crassulacean acid metabolism in *Kalanchoe daigremontiana*. *Plant Cell Environ.* **24**, 811–820.

Borland, A.M. and Griffiths, H. (1990) The regulation of CAM and respiratory recycling by water supply and light regime in the C3-CAM intermediate *Sedum telephium*. *Funct. Ecol.* **4**, 33.

Borland, A.M. and Griffiths, H. (1997) A comparative study on the regulation of C3 and C4 carboxylation processes in the constitutive crassulacean acid metabolism (CAM) plant *Kalanchoe daigremontiana* and the C3-CAM intermediate *Clusia minor*. *Planta*, **201**, 368–378.

Borland, A.M. and Taybi, T. (2004) Synchronization of metabolic processes in plants with crassulacean acid metabolism. *J. Exp. Bot.* **55**, 1255–1265.

Borland, A.M., Griffiths, H., Maxwell, C., Broadmeadow, M.S.J., Griffiths, N.M. and Barnes, J.D. (1992) On the ecophysiology of the Clusiaceae in Trinidad: expression of CAM in *Clusia minor* L. during the transition from wet to dry season and characterization of three endemic species. *New Phytol.* **122**, 349–357.

Borland, A.M., Hartwell, J., Jenkins, G.I., Wilkins, M.B., Nimmo, H.G. and Lüttge, U. (1999) Metabolite control overrides circadian regulation of phosphoenolpyruvate carboxylase kinase and CO(2) fixation in crassulacean acid metabolism. *Plant Physiol.* **121**, 889–896.

Borland, A.M., Griffiths, H., Hartwell, J. and Smith, J.A.C. (2009) Exploiting the potential of plants with crassulacean acid metabolism for bioenergy production on marginal lands. *J. Exp. Bot.* **60**, 2879–2896.

Borland, Anne M., Guo, H.-B., Yang, X. and Cushman, J.C. (2016) Orchestration of carbohydrate processing for crassulacean acid metabolism. *Curr. Opin. Plant Biol.* **31**, 118–124.

Borland, A.M., Leverett, A., Hurtado-Castano, N., Hu, R. and Yang, X. (2018) Functional anatomical traits of the photosynthetic organs of plants with crassulacean acid metabolism. In *The leaf: A platform for performing photosynthesis. Advances in photosynthesis and respiration (including bioenergy and related processes)* (Adams III, W. and Terashima, I. eds.), Cham: Springer, **44**, 281–305.

Boxall, S.F., Dever, L.V., Kneřová, J., Gould, P.D. and Hartwell, J. (2017) Phosphorylation of phospho*enol*pyruvate carboxylase is essential for maximal and sustained dark CO_2 fixation and core circadian clock operation in the obligate crassulacean acid metabolism species *Kalanchoë fedtschenkoi*. *Plant Cell*, **29**, 2519–2536.

Boxall, S.F., Kadu, N., Dever, L.V., Kneřová, J., Waller, J.L., Gould, P.J.D. and Hartwell, J. (2019) Silencing PHOSPHOENOLPYRUVATE CARBOXYLASE1 in the obligate crassulacean acid metabolism species *Kalanchoë laxiflora* causes reversion to C3-like metabolism and amplifies rhythmicity in a subset of core circadian clock genes. *bioRxiv*, 684050.

Brilhaus, D., Bräutigam, A., Mettler-Altmann, T., Winter, K. and Weber, A.P.M. (2016) Reversible burst of transcriptional changes during induction of crassulacean acid metabolism in *Talinum triangulare*1[OPEN]. *Plant Physiol.* **170**, 102–122.

Caemmerer, S.Von and Griffiths, H. (2009) Stomatal responses to CO_2 during a diel crassulacean acid metabolism cycle in *Kalanchoe daigremontiana* and *Kalanchoe pinnata*. *Plant Cell Environ.* **32**, 567–576.

Carter, P.J., Nimmo, H.G., Fewson, C.A. and Wilkins, M.B. (1991) Circadian rhythms in the activity of a plant protein kinase. *EMBO J.* **10**, 2063–2068.

Ceusters, J., Borland, A.M., Londers, E., Verdoodt, V., Godts, C. and De Proft, M.P. (2008) Diel shifts in carboxylation pathway and metabolite dynamics in the CAM bromeliad Aechmea "Maya" in response to elevated CO_2. *Ann. Bot.* **102**, 389–397.

Chiang, C.P., Yim, W.C., Sun, Y.H., Ohnishi, M., Mimura, T., Cushman, J.C. and Yen, H.E. (2016) Identification of ice plant (*Mesembryanthemum crystallinum* L.) microRNAs using RNA-seq and their putative roles in high salinity responses in seedlings. *Front. Plant Sci.* **7**, 1143.

Chu, C., Dai, Z., Ku, M.S.B. and Edwards, G.E. (1990) Induction of crassulacean acid metabolism in the facultative halophyte *Mesembryanthemum crystallinum* by abscisic acid. *Plant Physiol.* **93**, 1253–1260.

Cushman, J.C. (2001) Crassulacean acid metabolism. A plastic photosynthetic adaptation to arid environments. *Plant Physiol.* **127**, 1439–1448.

Cushman, J.C., Tillett, R.L., Wood, J.A., Branco, J.M. and Schlauch, K.A. (2008) Large-scale mRNA expression profiling in the common ice plant, *Mesembryanthemum crystallinum*, performing C3 photosynthesis and Crassulacean acid metabolism (CAM). *J. Exp. Bot.* **59**, 1875–1894.

Dai, Z., Ku, M.S.B., Zhang, D. and Edwards, G.E. (1994) Effects of growth regulators on the induction of crassulacean acid metabolism in the facultative halophyte *Mesembryanthemum crystallinum* L. *Planta*, **192**, 287–294.

Davies, B.N. and Griffiths, H. (2012) Competing carboxylases: circadian and metabolic regulation of Rubisco in C3 and CAM *Mesembryanthemum crystallinum* L. *Plant Cell Environ.* **35**, 1211–1220.

Dever, L.V., Boxall, S.F., Kne Rová, J. and Hartwell, J. (2015) Transgenic perturbation of the decarboxylation phase of crassulacean acid metabolism alters physiology and metabolism but has only a small effect on growth. *Plant Physiol.* **167**, 44–59.

Dodd, A.N., Borland, A.M., Haslam, R.P., Griffiths, H. and Maxwell, K. (2002) Crassulacean acid metabolism: plastic, fantastic. *J. Exp. Bot.* **53**, 569–580.

Drennan, P.M. and Nobel, P.S. (2000) Responses of CAM species to increasing atmospheric CO_2 concentrations. *Plant Cell Environ.* **23**, 767–781.

Earles, J.M., Theroux-Rancourt, G., Roddy, A.B., Gilbert, M.E., McElrone, A.J. and Brodersen, C.R. (2018) Beyond porosity: 3D leaf intercellular airspace traits that impact mesophyll conductance. *Plant Physiol.* **178**, 148–162.

Edwards, E.J. (2019) Evolutionary trajectories, accessibility and other metaphors: the case of C4 and CAM photosynthesis. *New Phytol.* **223**, 1742–1755.

Engineer, C.B., Hashimoto-Sugimoto, M., Negi, J., Israelsson-Nordström, M., Azoulay-Shemer, T., Rappel, W.-J., Iba, K. and Schroeder, J.I. (2016) CO_2 sensing and CO_2 regulation of stomatal conductance: advances and open questions. *Trends Plant Sci.* **21**, 16–30.

Farquhar, G.D., von Caemmerer, S. and Berry, J.A. (1980) A biochemical model of photosynthetic CO_2 assimilation in leaves of C3 species. *Planta*, **149**, 78–90.

Fernie, A.R. (2016) Systems biology: a new CAM era. *Nat. Plants*, **2**, 16190.

Gotoh, E., Oiwamoto, K., Inoue, S., Shimazaki, K. and Doi, M. (2019) Stomatal response to blue light in crassulacean acid metabolism plants *Kalanchoe pinnata* and *Kalanchoe daigremontiana*. *J. Exp. Bot.* **70**, 1367–1374.

Griffiths, H. (2013) Plant venation: from succulence to succulents. *Curr. Biol.* **23**, R340–R341.

Griffiths, H. and Males, J. (2017) Succulent plants. *Curr. Biol.* **27**, R890–R896.

Griffiths, H., Helliker Bc, B., Roberts, A., Haslam, R.P., Girnus, J., Robe, W.E., Borland, A.M. and Maxwell, K. (2002) Regulation of Rubisco activity in crassulacean acid metabolism plants: better late than never. *Funct. Plant Biol.* **29**, 689–696.

Griffiths, H., Cousins, A.B., Badger, M.R. and Von Caemmerer, S. (2007) Discrimination in the dark. Resolving the interplay between metabolic and physical constraints to phosphoenolpyruvate carboxylase activity during the crassulacean acid metabolism cycle. *Plant Physiol.* **143**, 1055–1067.

Griffiths, H., Robe, W.E., Girnus, J. and Maxwell, K. (2008) Leaf succulence determines the interplay between carboxylase systems and light use during crassulacean acid metabolism in Kalanchoë species*. *J. Exp. Bot.* **59**, 1851–1861.

Haag-kerwer, A., Franco, A.C. and Luttge, U. (1992) The effect of temperature and light on gas exchange and acid accumulation in the C3-CAM plant *Clusia minor* L. *J. Exp. Bot.* **43**, 345–352.

Hartwell, J., Smith, L.H., Wilkins, M.B., Jenkins, G.I. and Nimmo, H.G. (1996) Higher plant phosphoenolpyruvate carboxylase kinase is regulated at the level of translatable mRNA in response to light or a circadian rhythm. *Plant J.* **10**, 1071–1078.

Hartwell, J., Gill, A., Nimmo, G.A., Wilkins, M.B., Jenkins, G.I. and Nimmo, H.G. (1999) Phosphoenolpyruvate carboxylase kinase is a novel protein kinase regulated at the level of expression. *Plant J.* **20**, 333–342.

Hartwell, J., Dever, L.V. and Boxall, S.F. (2016) Emerging model systems for functional genomics analysis of crassulacean acid metabolism. *Curr. Opin. Plant Biol.* **31**, 100–108.

Hartzell, S., Bartlett, M.S., Virgin, L. and Porporato, A. (2015) Nonlinear dynamics of the CAM circadian rhythm in response to environmental forcing. *J. Theor. Biol.* **368**, 83–94.

Hartzell, S., Bartlett, M.S. and Porporato, A. (2018) Unified representation of the C3, C4, and CAM photosynthetic pathways with the Photo3 model. *Ecol. Modell.* **384**, 173–187.

Hassidim, M., Dakhiya, Y., Turjeman, A., Hussien, D., Shor, E., Anidjar, A., Goldberg, K. and Green, R.M. (2017) CIRCADIAN CLOCK ASSOCIATED1 (CCA1) and the circadian control of stomatal aperture. *Plant Physiol.* **175**, 1864–1877.

Henry, C., John, G.P., Pan, R., Bartlett, M.K., Fletcher, L.R., Scoffoni, C. and Sack, L. (2019) A stomatal safety-efficiency trade-off constrains responses to leaf dehydration. *Nat. Commun.* **10**, 3398.

Heyduk, K., Ray, J.N., Ayyampalayam, S., Moledina, N., Borland, A., Harding, S.A., Tsai, C.-J. and Leebens-Mack, J. (2019) Shared expression of crassulacean acid metabolism (CAM) genes pre-dates the origin of CAM in the genus Yucca. *J. Exp. Bot.* **70**, 6597–6609.

Hills, A., Chen, Z.-H., Amtmann, A., Blatt, M.R. and Lew, V.L. (2012) OnGuard, a computational platform for quantitative kinetic modeling of guard cell physiology. *Plant Physiol.* **159**, 1026–1042.

Holtum, J.A.M. and Winter, K. (1982) Activity of enzymes of carbon metabolism during the induction of crassulacean acid metabolism in *Mesembryanthemum crystallinum* L. *Planta*, **155**, 8–16.

Hsu, P.-K., Takahashi, Y., Munemasa, S., Merilo, E., Laanemets, K., Waadt, R., Pater, D., Kollist, H. and Schroeder, J.I. (2018) Abscisic acid-independent stomatal CO_2 signal transduction pathway and convergence of CO_2 and ABA signaling downstream of OST1 kinase. *Proc. Natl Acad. Sci. USA*, **115**, E9971–E9980.

Kim, T.-H., Böhmer, M., Hu, H., Nishimura, N. and Schroeder, J.I. (2010) Guard cell signal transduction network: advances in understanding abscisic acid, CO_2, and Ca^{2+} Signaling. *Annu. Rev. Plant Biol.* **61**, 561–591.

Kong, W., Yoo, M.-J., Noble, J.D., Kelley, T.M., Li, J., Kirst, M., Assmann, S.M. and Chen, S. (2019) Molecular changes in Mesembryanthemum crystallinum guard cells underlying the C3 to CAM transition. *bioRxiv*, 607333.

Lee, D.M. and Assmann, S.M. (1992) Stomatal responses to light in the facultative crassulacean acid metabolism species, *Portulacaria afra*. *Physiol. Plant*, **85**, 35–42.

Lian, C., Li, Q., Yao, K., Zhang, Y., Meng, S., Yin, W. and Xia, X. (2018) Populus trichocarpa PTNF-YA9, a multifunctional transcription factor, regulates seed germination, abiotic stress, plant growth and development in arabidopsis. *Front. Plant Sci.* **9**, 1–15.

Lim, S.D., Lee, S., Choi, W.G., Yim, W.C. and Cushman, J.C. (2019) Laying the foundation for crassulacean acid metabolism (CAM) biodesign: expression of the C4 metabolism cycle genes of CAM in arabidopsis. *Front. Plant Sci.* **10**, 101.

Liu, D., Chen, M., Mendoza, B., Cheng, H., Hu, R., Li, L., Trinh, C.T., Tuskan, G.A. and Yang, X. (2019) CRISPR/Cas9-mediated targeted mutagenesis for functional genomics research of crassulacean acid metabolism plants. *J. Exp. Bot.* **70**, 6621–6629.

Maleckova, E., Brilhaus, D., Wrobel, T.J. and Weber, A.P.M. (2019) Transcript and metabolite changes during the early phase of abscisic acid-mediated induction of crassulacean acid metabolism in *Talinum triangulare*. *J. Exp. Bot.* **70**, 6581–6596.

Males, J. and Griffiths, H. (2017) Stomatal biology of CAM plants. *Plant Physiol.* **174**, 550–560.

Males, J. and Griffiths, H. (2018) Economic and hydraulic divergences underpin ecological differentiation in the Bromeliaceae. *Plant Cell Environ.* **41**, 64–78.

Maurice Cheung, C.Y., Poolman, M.G., Fell, D.A., George Ratcliffe, R. and Sweetlove, L.J. (2014) A diel flux balance model captures interactions between light and dark metabolism during day-night cycles in C3 and crassulacean acid metabolism leaves. *Plant Physiol.* **165**, 917–929.

Maxwell, K., Griffiths, H., Helliker, B., Roberts, A., Haslam, R.P., Girnus, J., Robe, W.E. and Borland, A.M. (2002) Regulation of Rubisco activity in crassulacean acid metabolism plants: better late than never. *Funct. Plant Biol.* **29**, 689.

Ming, R., VanBuren, R., Wai, C.M. *et al.* (2015) The pineapple genome and the evolution of CAM photosynthesis. *Nat. Genet.* **47**, 1435–1442.

Moseley, R.C., Mewalal, R., Motta, F., Tuskan, G.A., Haase, S. and Yang, X. (2018) Conservation and diversification of circadian rhythmicity between a model crassulacean acid metabolism plant *Kalanchoë fedtschenkoi* and a model C3 photosynthesis plant *Arabidopsis thaliana*. *Front. Plant Sci.* **9**, 1757.

Moseley, Robert C., Motta, F., Tuskan, G.A., Haase, S. and Yang, X. (2019a) Inference of gene regulatory network uncovers the linkage between

circadian clock and crassulacean acid metabolism in *Kalanchoë fedtschenkoi*. *bioRxiv*. https://doi.org/10.1101/745893 [Accessed December 22].

Moseley, Robert C., Tuskan, G.A. and Yang, X. (2019b) Comparative genomics analysis provides new insight into molecular basis of stomatal movement in *Kalanchoë fedtschenkoi*. *Front. Plant Sci.* **10**, 292.

Nimmo, H.G. (2000) The regulation of phosphoenolpyruvate carboxylase in CAM plants. *Trends Plant Sci.* **5**, 75–80.

Nobel, P.S. and Valenzuela, A.G. (1987) Environmental responses and productivity of the CAM plant, *Agave tequilana*. *Agric. For. Meteorol.* **39**, 319–334.

Nolf, M., Lopez, R., Peters, J.M.R., Flavel, R.J., Koloadin, L.S., Young, I.M. and Choat, B. (2017) Visualization of xylem embolism by X-ray microtomography: a direct test against hydraulic measurements. *New Phytol.* **214**, 890–898.

North, G.B., Martre, P. and Nobel, P.S. (2004) Aquaporins account for variations in hydraulic conductance for metabolically active root regions of *Agave deserti* in wet, dry, and rewetted soil. *Plant Cell Environ.* **27**, 219–228.

North, G.B., Brinton, E.K., Browne, M.G., Gillman, M.G., Roddy, A.B., Kho, T.L., Wang, E., Fung, V.A. and Brodersen, C.R. (2019) Hydraulic conductance, resistance, and resilience: how leaves of a tropical epiphyte respond to drought. *Am. J. Bot.* **106**, 943–957.

Ogburn, R.M. and Edwards, E.J. (2013) Repeated origin of three-dimensional leaf venation releases constraints on the evolution of succulence in plants. *Curr. Biol.* **23**, 722–726.

Osmond, C.B. (1978) Crassulacean acid metabolism: a curiosity in context. *Annu. Rev. Plant Physiol.* **29**, 379–414.

Owen, N.A. and Griffiths, H. (2013a) A system dynamics model integrating physiology and biochemical regulation predicts extent of crassulacean acid metabolism (CAM) phases. *New Phytol.* **200**, 1116–1131.

Owen, N.A., Choncubhair, Ó.N., Males, J., del Real Laborde, J.I., Rubio-Cortés, R., Griffiths, H. and Lanigan, G. (2016) Eddy covariance captures four-phase crassulacean acid metabolism (CAM) gas exchange signature in *Agave*. *Plant Cell Environ.* **39**, 295–309.

Sade, N., Shatil-Cohen, A., Attia, Z. *et al.* (2014) The role of plasma membrane aquaporins in regulating the bundle sheath-mesophyll continuum and leaf hydraulics. *Plant Physiol.* **166**, 1609–1620.

Sakamoto, R., Matoba, C., Suemitsu, H. and Matsuo, T. (2015) Feedback control of biological rhythm in crassulacean acid metabolism by CO_2-uptake signal. *IFAC-PapersOnLine*, **48**, 59–64.

Santelia, D. and Lawson, T. (2016) Rethinking guard cell metabolism. *Plant Physiol.* **172**, 1371–1392.

Schatz, M.C. (2012) Computational thinking in the era of big data biology. *Genome Biol.* **13**, 177.

Scoffoni, C., Albuquerque, C., Cochard, H. *et al.* (2018) The causes of leaf hydraulic vulnerability and its influence on gas exchange in *Arabidopsis thaliana*. *Plant Physiol.* **178**, 1584–1601.

Shameer, S., Baghalian, K., Cheung, C.Y.M., Ratcliffe, R.G. and Sweetlove, L.J. (2018) Computational analysis of the productivity potential of CAM. *Nat. Plants*, **4**, 165–171.

Smita, S., Katiyar, A., Lenka, S.K. *et al.* (2019) Gene network modules associated with abiotic stress response in tolerant rice genotypes identified by transcriptome meta-analysis. *Funct. Integr. Genomics*, **20**, 29–49.

Smith, J.A.C., Schulte, P.J. and Nobel, P.S. (1987) Water flow and water storage in Agave Deserti: osmotic implications of crassulacean acid metabolism. *Plant Cell Environ.* **10**, 639–648.

Tallman, G., Zhu, J., Mawson, B.T., Amodeo, G., Nouhi, Z., Levy, K. and Zeiger, E. (1997) Induction of CAM in *Mesembryanthemum crystallinum* abolishes the stomatal response to blue light and light-dependent zeaxanthin formation in guard cell chloroplasts. *Plant Cell Physiol.* **38**, 236–242.

Taybi, T. and Cushman, J.C. (1999) Signaling events leading to crassulacean acid metabolism induction in the common ice plant. *Plant Physiol.* **121**, 545–555.

Taybi, T., Sotta, B., Gehrig, H., Güçlü, S., Kluge, M. and Brulfert, J. (1995) Differential effects of abscisic acid on phosphoenolpyruvate carboxylase and CAM operation in *Kalanchoë blossfeldiana*. *Bot. Acta*, **108**, 240–246.

Wai, C.M. and VanBuren, R. (2018) *Circadian Regulation of Pineapple CAM Photosynthesis*. Cham: In Springer, pp. 247–258.

Wai, C.M., VanBuren, R., Zhang, J. *et al.* (2017) Temporal and spatial transcriptomic and microRNA dynamics of CAM photosynthesis in pineapple. *Plant J.* **92**, 19–30.

Wai, C.M., Weise, S.E., Ozersky, P., Mockler, T.C., Michael, T.P. and VanBuren, R. (2019) Time of day and network reprogramming during drought induced CAM photosynthesis in *Sedum album*. *PLOS Genet.* **15**, e1008209.

Winter, K. (2019) Ecophysiology of constitutive and facultative CAM photosynthesis. *J. Exp. Bot.* **70**, 6495–6508.

Winter, K. and Holtum, J.A.M. (2007) Environment or development? Lifetime net CO_2 exchange and control of the expression of crassulacean acid metabolism in *Mesembryanthemum crystallinum*. *Plant Physiol.* **143**, 98–107.

Winter, K., Holtum, J.A.M. and Smith, J.A.C. (2015) Crassulacean acid metabolism: a continuous or discrete trait? *New Phytol.* **208**, 73–78.

Wyka, T.P., Duarte, H.M. and Lüttge, U.E. (2005) Redundancy of stomatal control for the circadian photosynthetic rhythm in *Kalanchoë daigremontiana* Hamet et Perrier. *Plant Biol.* **7**, 176–181.

Yang, X., Cushman, J.C., Borland, A.M. *et al.* (2015) A roadmap for research on crassulacean acid metabolism (CAM) to enhance sustainable food and bioenergy production in a hotter, drier world. *New Phytol.* **207**, 491–504.

Yang, X., Hu, R., Yin, H. *et al.* (2017) The Kalanchoë genome provides insights into convergent evolution and building blocks of crassulacean acid metabolism. *Nat. Commun.* **8**, 1899.

Yang, X., Liu, D., Tschaplinski, T.J. and Tuskan, G.A. (2019) Comparative genomics can provide new insights into the evolutionary mechanisms and gene function in CAM plants. *J. Exp. Bot.* **70**, 6539–6547.

Yin, H., Guo, H.-B., Weston, D.J. *et al.* (2018) Diel rewiring and positive selection of ancient plant proteins enabled evolution of CAM photosynthesis in Agave. *BMC Genom.* **19**, 588.

Zhang, L., Chen, F., Zhang, G.-Q., Zhang, Y.-Q., Niu, S., Xiong, J.-S., Lin, Z., Cheng, Z.-M.M. and Liu, Z.-J. (2016) Origin and mechanism of crassulacean acid metabolism in orchids as implied by comparative transcriptomics and genomics of the carbon fixation pathway. *Plant J.* **86**, 175–185.

Zhu, K., Liu, H., Chen, X., Cheng, Q. and Cheng, Z.M.M. (2018) The kinome of pineapple: catalog and insights into functions in crassulacean acid metabolism plants. *BMC Plant Biol.* **18**, 199.

