Regulation of crassulacean acid metabolism (CAM) in the facultative CAM species *Talinum triangulare*

Inauguraldissertation

zur Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Eva Malečková

aus Bruntál

Düsseldorf, Juni 2020

aus dem
Institut für Biochemie der Pflanzen
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Referent: Prof. Dr. Andreas P. M. Weber
Korreferent: Prof. Dr. Matias Zurbriggen
Tag der mündlichen Prüfung:

Eidesstattliche Versicherung und Selbstständigkeitserklärung

Ich versichere an Eides Statt, dass die Dissertation von mir selbständig und ohne unzulässige fremde Hilfe unter Beachtung der „Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf“ erstellt worden ist.

Diese Dissertation habe ich in dieser oder ähnlicher Form noch bei keiner anderen Fakultät vorgelegt.

Ich habe bisher keine erfolglosen Promotionsversuche unternommen.

Düsseldorf, den _____

_____

Eva Malečková

Table of contents

Summary ..... *v*

Introduction *6*

CAM evolution **7**

Diversity of CAM 7

Molecular evolution of CAM 8

Biochemistry of CAM **10**

Interplay with carbohydrate metabolism 13

Regulation of the CAM pathway 14

CAM engineering **18**

The role of genomics 20

Abscisic acid signalling **22**

ABA signalling in CAM induction 23

References *24*

Abbreviations *33*

Aims of the thesis *34*

Manuscript I *35*

Supplementary figures **51**

Supplementary tables **53**

Supplementary protocols **64**

Manuscript II *73*

Supplementary figures **129**

Supplementary tables **137**

Acknowledgements *163*

Summary

Agriculture of the 21st century faces challenges in form of growing global population, increasing demands for food, feed and plant-based products, ongoing climate changes with more frequent and more severe drought periods, and shrinking arable land. To satisfy the production demands, plants relying on crassulacean acid metabolism (CAM) are an attractive alternative to C_3 and C_4 crops that require extensive irrigation. In parallel to increased reliance on CAM plants, CAM engineering (*i.e.* introduction of the CAM pathway to conventional crops) is a promising strategy to generate crops better suited to warmer and drier conditions.

CAM is an extremely plastic adaptation, including facultative CAM - a reversible induction of the CAM pathway in C_3 or rarely C_4 species. Facultative CAM species are an excellent system to dissect the genetic components of CAM, mechanisms of their regulation and to shed more light on the evolutionary trajectories of CAM. *Talinum triangulare* is one of many facultative CAM species, with well described drought-inducible CAM and first '-omics' resources available. The present thesis aims to establish *T. triangulare* as a model species for the field of CAM research.

Even though water withdrawal induces CAM within few days in *T. triangulare*, the timing and nature of the initial signalling events leading to CAM induction are not known. Moreover, the onset of CAM induction depends on external variables, such as soil moisture and air humidity. Following up on the findings from the drought experiment performed previously, the phytohormone abscisic acid (ABA) was applied as a foliar spray. The effect of the treatment was studied by the means of RNA-sequencing and metabolite measurements, revealing the rapid pace and extent of ABA-induced transcriptional changes. These were moreover accompanied by altered levels of several metabolites and most importantly titratable acidity at the end of the dark. Several candidate transcription factors were identified as possible mediators of the CAM induction based on their transcriptional responsiveness. An unexpected finding was pronounced and rapid protein degradation following the ABA treatment. It was hypothesized that degradation of proteins with activities competing with the CAM cycle is an essential to CAM induction alongside with synthetic processes (Manuscript I).

Draft genome assembly and reference transcriptome were generated with long read sequencing technologies. Temporal adjustment of gene expression likely gained on importance in CAM species and promoter sequences of *T. triangulare* were therefore analysed for presence of known transcription factor binding sites. Motifs associated with ABA response and the circadian clock were enriched among genes sharing similar expression profiles during CAM induction. Interspecies comparison identified several putative transcription factor binding sites enriched in promoters of *T. triangulare* but not in C_3 and obligate CAM species. Furthermore, protein-coding sequences were examined. Candidate genes shared by facultative CAM species, which included several putative transcription factors, an ABC transporter, and an ABA-inducible protein, were identified in an interspecies comparison. Examination of predicted protein sequences of phospho*enol*pyruvate carboxylase responsible for nocturnal CO_2 fixation revealed considerable diversity among isoforms of *T. triangulare* as well as across species. Further investigations of CAM regulation and evolution should embrace much of the diversity of CAM species and focus on both protein-coding sequences and regulatory elements (Manuscript II).

Introduction

Until 2050, human population on the Earth is expected to reach 9 billion, which together with rising standards of living will increase the demands on food production by up to 49 % (FAO, 2019). Besides growing need to produce more feed for livestock, plant-based products, such as fibre and biofuels gain on importance. Satisfying all those needs is challenged by ongoing and expected climate changes, including reduced soil moisture and higher frequencies of extreme drought events, by shrinking areas of arable land both due to climate changes and growing competition for land between agriculture and urban growth. Taken together, the agriculture has to change to use less resources, such as irrigation water and fertilizers (Trenberth, 2011; Dai, 2013; Maurino and Weber, 2013; Cushman *et al.*, 2015; Yang *et al.*, 2015). As of 2007, arid, semi-arid and dry sub-humid areas accounted for approximately 30 % of the world's global area. These lands are not only one of the world's most fragile ecosystems but also provide means of living to about one billion people, a large proportion of whom belongs to the poorest in the world. Given inadequate precipitation in these regions, they are at the same unsuitable for growing economically and dietary important C_3 and C_4 crops (Malagnoux, 2007).

One possible solution to face these challenges is to take advantage of CAM species with often remarkable water-use efficiency (WUE), which is defined as units of CO_2 fixed per a unit water lost (Niechayev *et al.*, 2019). By shifting all or part of the net CO_2 uptake to the dark phase of the 24-hr cycle, CAM species can perform up to 6- and 3-fold better than species relying on C_3 and C_4 photosynthesis, respectively. High WUE comes hand in hand with increased heat stress tolerance, meaning that CAM crops could serve as climate-resilient sources for food, feed, fibre, biofuel and even pharmaceuticals, expanding their production to semiarid, abandoned and/or marginal agricultural lands (Mason *et al.*, 2015; Liu *et al.*, 2018; Davis *et al.*, 2019). Indeed, this is the case for economically increasingly important crops, such as *Agave* spp., *Aloe vera* and *Opuntia ficus-indica* (summarized by Liu *et al.*, 2018). More challenging yet even more promising strategy is engineering of the CAM pathway into conventional crops, thus making them more adapted to lands with reduced soil moisture and reduced precipitation (Borland *et al.*, 2014; Yang *et al.*, 2015; Liu *et al.*, 2018). Here, two alternative strategies are imaginable: i) full reversion of C_3 (or C_4) photosynthesis into CAM, thus expanding cultivation of present-day crops to regions which are or will become too arid (Yang *et al.*, 2015; Liu *et al.*, 2018) or ii) implementation of an inducible CAM pathway, in which case the water saving strategy could be activated on demand (*e.g.* during extended heat and drought waves in the areas with moderate climate).

However, to make CAM engineering possible, a deeper understanding of this complex pathway and its regulation is required, while multiple points of view are desirable: ecological, anatomical, biochemical, genetic as well as evolutionary perspective (Goolsby *et al.*, 2018). Even though the CAM pathway has been studied biochemically starting in 1970s, most of our current knowledge is based on *Mesembryanthemum crystallinum* and *Kalanchoë* species. Moreover, there is a lack of knowledge regarding genomic features underlying this complex phenotype and its evolutionary history. The present thesis aims to establish the facultative CAM species *Talinum triangulare* as another CAM model. In this species, stress-induction of CAM can be replaced by exogenous ABA, thus making CAM induction more reproducible and faster. Time-course mRNA-sequencing experiment identified early responsive genes during the induction phase for future investigations (Manuscript I). Secondly, taking advantage of long-read sequencing technologies, the first draft genome assembly and an improved transcriptome assembly

were generated. Thanks to these resources, it was possible to identify *cis*-regulatory elements associated with ABA-induced transcriptional changes and to identify candidate genes that possibly evolved in facultative CAM species (Manuscript II).

CAM evolution

CAM emergence is a great example of convergent evolution. Up to date, CAM has been described in about 6% of flowering plants which are represented by at least 35 plant families of monocots, dicots as well as some basal plant lineages. The latter include submerged aquatic macrophytes *Isoëtes* spp., cycad *Dioon edule* and several ferns (Keeley, 1998; Holtum and Winter, 1999; Vovides *et al.*, 2002; Silvera *et al.*, 2010). Repeated and independent emergence of CAM on numerous occasions is supported by i) presence of CAM in phylogenetically taxa, and ii) diversity of decarboxylating enzymes as well as carbohydrate storage systems recruited in CAM (Christopher and Holtum, 1996; Silvera *et al.*, 2010; Christin *et al.*, 2014; Niechayev *et al.*, 2019). Independent CAM origins are common even among relatively closely related species as observed for example in *Agavoideae, Euphorbia, Bromeliaceae* and *Caryophyllales* (Christin *et al.*, 2014; Horn *et al.*, 2014; Crayn *et al.*, 2015; Heyduk *et al.*, 2016). Besides selection for increased WUE, the decline in atmospheric CO_2 concentration in the Miocene and early Pleistocene could impose a selection pressure for CAM evolution (Silvera *et al.*, 2010; Arakaki *et al.*, 2011). Limited CO_2 availability due to low diffusion coefficient of CO_2 in water was also the likely driver for emergence of CAM in primitive aquatic plants such as in the above-mentioned genus *Isoëtes* (Keeley, 1998).

In some families, CAM is much more frequent than in others. Families especially rich on CAM species include *Cactaceae, Agavaceae* and *Didiereaceae* with all genera likely performing CAM, and *Orchidaceae* and *Bromeliaceae*, with about 10% and 57% of species performing CAM, respectively. The high frequency of CAM in these families correlates with water- and nutrition-limited, terrestrial and epiphytic habitats these species thrive in (Silvera *et al.*, 2009, 2010; Bone *et al.*, 2015; Crayn *et al.*, 2015; Niechayev *et al.*, 2019). Among dicots, the order of *Caryophyllales* presents a true hotspot of photosynthetic transitions, which makes it a great system to study evolution of these complex traits. It is worth mentioning that six *Portulaca* species have been described up to date, which evolved Kranz anatomy typical of C_4 species but can express CAM in a facultative manner (Christin *et al.*, 2014; Holtum *et al.*, 2017*a*; Winter and Holtum, 2017).

Diversity of CAM

In most cases described up to date, CAM does not appear as a sole carbon assimilation strategy but comes together with C_3 and in rare cases with C_4 photosynthesis, with the relative proportion of carbon assimilated via either pathway changing (Winter, 2019). The factors shifting the preference to one or another photosynthetic type are developmental as well as environmental. For simplicity, only the terms relevant to current hypotheses about CAM evolution will be clarified here.

In obligate (constitutive) CAM species, CAM expression in photosynthetic tissue is under developmental control, with young C_3 tissue progressively transitioning to CAM (*e.g. K. pinnata*, *K. fedtschenkoi*, *Agave* spp.). In contrast, facultative CAM is induced in C_3 or C_4 tissue in response to abiotic stress, such as drought or high salinity. In truly facultative CAM species, the transition is reversible

upon stress removal. Besides *T. triangulare* and *M. crystallinum*, *Calandrinia polyandra*, an epiphytic fern *Vittaria lineata* and an orchid *Dendorbium catenatum* are examples of facultative CAM species (Minardi *et al.*, 2014; Winter and Holtum, 2014; Zou *et al.*, 2018; Winter, 2019). The flexibility of facultative CAM is its major advantage. C_3 assimilation is associated with faster growth, while stress-induced CAM brings the advantage of increased WUE when conditions become less favourable (Wai *et al.*, 2019).

Weak and strong CAM refer to the proportion of total carbon gain during the dark phase. Weak CAM can be expressed both constitutively and facultatively, but C_3 remains the primary mode of carbon assimilation. Nocturnal CO_2 uptake via the CAM pathway does not exceed 5% of total carbon gain in constitutive CAM species and in weak facultative CAM species, nocturnal CO_2 uptake contributes less than 5% of the daytime CO_2 fixation (Winter, 2019). Strong CAM, in contrast, is characterized by high rates of nocturnal assimilation and is typical for obligate CAM species (Winter, 2019).

Nocturnal stomal opening is a typical feature of CAM. However, this may not always be the case as well illustrated by CAM cycling and CAM-idling. Under CAM cycling conditions, atmospheric CO_2 is taken up during the light phase and the detectable fluctuations in malic acid content between the light and dark phases are due to nocturnal malate synthesis fuelled by re-fixation of respiratory CO_2. CAM cycling most often occurs in species expressing facultative or weak CAM. Even more extreme form is CAM-idling when stomata remain close throughout the 24-hr cycle and malate synthesis fully depends on re-fixation of respiratory CO_2. Strong CAM species switch to CAM-idling as a survival strategy under severe stress conditions (Harris and Martin, 1991; Silvera *et al.*, 2010; Winter, 2019).

Currently, CAM is viewed as a continuum of flexible modes of carbon assimilation rather than as a discrete trait as supported by plasticity of the CAM syndrome outlined above (Silvera *et al.*, 2010; Winter *et al.*, 2015). Yet, the evolutionary trajectory leading to CAM remains to be fully deciphered. Yang *et al.* (2019) proposed two alternative hypotheses. One possibility is evolution along a C_3 – facultative CAM – weak (obligate) CAM – strong (obligate) CAM continuum. Alternatively, facultative CAM could have evolved from C_3 ancestors independently of a C_3 – weak CAM – strong CAM trajectory. Regardless of the exact evolutionary path, CAM does not have to be a finite state and sliding back is possible. In *Bromelioideae*, a subfamily of bromelids, CAM was detected in 90% of species tested. CAM was determined to be the ancestral carbon assimilation strategy with reversal to C_3 in some species (Ming *et al.*, 2016). Strong CAM seems to be the ancestral carbon assimilation pathway in the *Agave* clade with reversion to weak CAM in the sister group of *Polianthes* (Heyduk *et al.*, 2018). The effect of loss of individual CAM enzymes was investigated in RNAi lines of *Kalanchoë* spp. Loss of NAD-ME1 or PPDK1 did not disadvantage the mutant lines unless exposed to extreme drought. Mutants lacking PPC1 activity resembled facultative CAM species when drought-treated and re-watered (Dever *et al.*, 2015; Boxall *et al.*, 2020). In all those cases, nocturnal acidification was abolished but the loss of CAM activities was not lethal. This supports that an evolutionary CAM to C_3 reversal may not come with detrimental costs.

Molecular evolution of CAM

Our understanding of the genetic blueprint underlying CAM photosynthesis remains incomplete but the findings of years of CAM research and high degree of similarity between C_4 and CAM pathways led to

the hypothesis that CAM developed through reorganization of metabolic processes present in C_3 ancestors (West-Eberhard *et al.*, 2011; Bräutigam *et al.*, 2017). This likely requires increased abundance of the (de)carboxylating enzymes, enzymes of carbohydrate metabolism, regulatory proteins and transporters necessary to accommodate the metabolic flow through the pathway.

Bräutigam *et al.* (2017) proposed that all necessary metabolite fluxes are already in place in C_3 species and their strengthening is thus sufficient to lay foundations for the CAM pathway even without any temporal rewiring. If this was the case, CAM engineering into C_3 hosts might require smaller adjustments to the host's metabolism than currently assumed. However, with the increasing number of transcriptomics studies, pronounced changes in transcript levels and temporal abundance patterns of core CAM enzymes but also genes of starch turnover were reported between obligate CAM and C_3 species as well as between C_3 and CAM modes of a single facultative CAM species (Cushman and Bohnert, 1999; Cushman *et al.*, 2008*b*; Abraham *et al.*, 2016; Brilhaus *et al.*, 2016). The extent of transcriptional changes underlying the transition from C_3 to CAM was revealed by two recent network analyses (Heyduk *et al.*, 2019*a*; Wai *et al.*, 2019).

Supportive of altered expression patterns in CAM species are the recent studies with pineapple and *Sedum album*, which identified *cis*-regulatory elements associated with promoters of core CAM and related genes (Wai *et al.*, 2017, 2019; Chen *et al.*, 2020). In both species, recruitment of clock-associated motifs was detected but it remains to be identified whether the *trans*-regulators are conserved as well or whether novel *trans*-regulators evolved in CAM species. It cannot be excluded that in parallel, novel transcription factor binding sites (TFBS) were acquired during CAM evolution (Boxall *et al.*, 2005; Wai *et al.*, 2019). These questions will be easier to address with genomics recourse being generated for more CAM species.

The best studied example of molecular changes associated with the CAM phenotype is *PHOSPHOENOLPYRUVATE CARBOXYLASE* (*PPC* gene, PEPC protein). *PPC* is a multimember gene family in green plants, which underwent numerous duplications in its evolutionary history, both ancestral ones and in individual plant lineages (Deng *et al.*, 2016). *PPC* isogenes belong to three clades, of which *PPC1* clade further diversified between dicots and monocots. Despite of many independent origins of CAM, isogenes from the *PPC1* lineage were recruited to perform C_4 or CAM function in all monocots and dicots examined up to date. Besides that, the emergence of CAM predated C_4 in both lineages (Christin *et al.*, 2014; Deng *et al.*, 2016). Comparative and expression studies support differentiation of amino acid sequences as well as expression patterns of *PPC* isogenes (Christin *et al.*, 2014; Brilhaus *et al.*, 2016; Deng *et al.*, 2016; Chen *et al.*, 2020).

Taken together, there is a general agreement that CAM evolutions acts on the genetic components already present in C_3 species. However, since CAM requires a large number of biochemical and anatomical adjustments, it is likely that numerous genetic changes are necessary for CAM evolution (Silvera *et al.*, 2010), the degree and nature of which remains a subject of further research. Besides core CAM genes, regulators of CAM and its induction in facultative species need to be identified. There are indications that stress-responsive genes of C_3 species could have evolved to fulfil this function (Yang *et al.*, 2019). Borland *et al.* (2016) proposed this was the case of chloroplastic *GLUCOSE-6-PHOSPHATE/PHOSPHATE TRANSLOCATOR (GPT)* and *α-GLUCAN PHOSPHORYLASE (PHS1).*

More recently, phase shift of *HEAT SHOCK PROTEIN 70 (HSP70)* expression was determined both in *K. fedtschenkoi* and pineapple as compared to C_3 Arabidopsis (Yang *et al.*, 2017). Facultative CAM species present a powerful tool to identify regulators of CAM, which can subsequently be examined in obligate CAM and C_3 species.

Biochemistry of CAM

Crassulacean acid metabolism (CAM) is a specialized mode of photosynthesis, relying on strict temporal regulation of primary CO_2 assimilation through a series of carboxylation reactions and its secondary fixation via Calvin-Benson-Bassham cycle (CBBC). CO_2 refixation during the light phase leads to increased CO_2 concentration around 1,5-bisphosphate carboxylase/oxygenase (RuBisCO), which was reported to achieve up to 60-fold increase compared to atmospheric [CO_2] (Lüttge, 2002). Thus, the carboxylase activity of RuBisCO is strongly favoured, and energetically wasteful process of photorespiration suppressed. Temporal separation of primary and secondary CO_2 fixation is possible thanks to a reversed stomatal behaviour compared to C_3 and C_4 plants. Stomata closure during the light is the key to remarkable WUE of CAM species (Osmond, 1978; Males and Griffiths, 2017).

The CAM cycle operates via four temporally separated phases (Fig. 1A). At night (Phase I), when ambient temperature is lower and relative humidity higher, evapotranspiration is minimized and stomata open to enable primary atmospheric CO_2 assimilation (Osmond, 1978; Hartwell, 2005). CO_2 is converted, likely by carbonic anhydrase (CA), into bicarbonate (HCO_3^-), which is combined with phospho*enol*pyruvate (PEP) by the action of phospho*enol*pyruvate carboxylase (PEPC) to form oxaloacetate (OAA). Subsequently, malate dehydrogenase (MDH) reduces OAA to malate, the primary storage form of nocturnally fixed CO_2 (Cushman and Bohnert, 1997; Borland *et al.*, 2009). Following the H^+ electrochemical gradient, malate^{2-} anions cross the tonoplast together with two titratable protons in a charge-balancing, passive process. Additionally, the excess of H^+ inside the vacuole leads to malate accumulation in form of malic acid, concentration of which can reach up to 200 mM in *Clusia hilariana*. Vacuolar-type H^+-ATPase and H^+-PP_iase maintain the electrochemical gradient across the tonoplast (Lüttge *et al.*, 1981; Franco *et al.*, 1996; Cheffings *et al.*, 1997). Expression of vacuolar H^+-ATPase *c* subunit is light-dependent in *Kalachoë daigremontiana* and induced upon salt exposure in stress-inducible CAM species *M. crystalinum* (Bartholomew *et al.*, 1996; Tsiantis *et al.*, 1996).

While nocturnal accumulation of malic acid is a well described phenomenon and measurements of titratable acidity are an easy and widely used indicator of CAM, the nature of malate transporters has not been fully elucidated yet. Malate^{2-} influx occurs via anion-selective channels which are likely members of the aluminium-activated malate transporter (ALMT) family (Hafke *et al.*, 2003; Kovermann *et al.*, 2007). While low in abundance, *ALMT* transcript levels increase up to 28-fold in *T. triangulare* assimilating in the CAM mode (Brilhaus *et al.*, 2016). Similarly, *ALMT1* transcript accumulations overlaps with increased PEPC activity and nocturnal acidification in bromeliad *Guzmania monostachia* (Pereira *et al.*, 2018) but *ALMT* homologs have not been described functionally in any CAM species yet.

Upon illumination (Phase II), PEPC activity starts to decline and with a short delay, RuBisCO is activated by a dedicated activase. After the initial transient burst of carbon assimilation into both C_4 acids and C_3 intermediates, stomata close, PEPC is deactivated, and RuBisCO becomes the only active carboxylase (Fig. 1B; Cushman and Bohnert, 1997; Maxwell *et al.*, 1999). RuBisCO remains active

throughout most of the light period (Phase III), re-fixing CO_2 released from the nocturnally stored malate. When vacuolar malic acid concentration is still high, undissociated malic acid exists the vacuole in a passive, non-catalysed manner. With progressing time and vacuolar malic acid reserves declining, efflux changes to $Hmal^{1-}$ and/or mal^{2-} forms (Lüttge and Smith, 1984). Malate exporter(s) seem to be distinct from the malate channel(s) mediating its influx (Cheffings *et al.*, 1997) but they remain to be characterized at molecular and genetic level. Efflux in a stochiometric manner (*i.e.* 2 H^+:mal^{2-} or H^+:$Hmal^{1-}$) would explain net malic acid efflux but it could not be proven yet. More recently, tonoplast dicarboxylate transporter (tDT) identified as vacuolar importer and exporter of malate and other dicarboxylates in Arabidopsis was proposed to mediate malate efflux in CAM plants (Cheffings *et al.*, 1997; Emmerlich *et al.*, 2003; Hurth *et al.*, 2005; Borland *et al.*, 2009; Frei *et al.*, 2018). Indeed, *T. triangulare* homolog was among the most up-regulated genes during drought-induced CAM (Brilhaus *et al.*, 2016) but like for malate importers, functional characterisation is missing up to date in CAM species.

Once exported from the vacuole, malate undergoes decarboxylation, the biochemistry of which is species-specific. Malic enzyme (ME)-species rely primarily on one or more isoforms of ME, which are: mitochondrial NAD^+-ME, chloroplastic $NADP^+$-ME and cytosolic $NADP^+$-ME. Besides CO_2, pyruvate is generated and it subsequently undergoes enzymatic conversion to PEP, which is catalysed by pyruvate orthophosphate dikinase (PPDK) in ME-species (Cushman and Bohnert, 1999; Borland *et al.*, 2009). In CAM model species *M. crystallinum*, PPDK activity could be confirmed in chloroplasts only, requiring the presence of a dedicated pyruvate and PEP transporters in the chloroplast envelope for successful PEP regeneration (Winter *et al.*, 1982; Demmig and Winter, 1983; Neuhaus *et al.*, 1988; Holtum *et al.*, 2005). In a biochemically closely related C_4 pathway, bile-acid sodium symporter 2 (BASS2) mediates pyruvate transport into the chloroplast (Furumoto *et al.*, 2011) but up to date, it has not been functionally characterized in any CAM species. Brilhaus *et al.*, (2016) however showed *BASS2* transcript accumulation in *T. triangulare* in the CAM mode.

Presence of PEP phosphate translocator (PPT) acting as a PEP exporter and accumulation of its transcripts throughout the light phase was confirmed in chloroplasts of *M. crystallinum* (Häusler *et al.*, 2000). In contrast to *M. crystallinum*, facultative CAM species *K. blossfeldiana* accumulates PPDK both in cytosol and chloroplast, with both isoforms showing increased abundance in response to drought (Kondo *et al.*, 2001). This suggests that PEP regeneration can occur both in cytosol and chloroplast, but studies shedding more light on relative importance and regulation of the various PPDK and ME isoforms are missing.

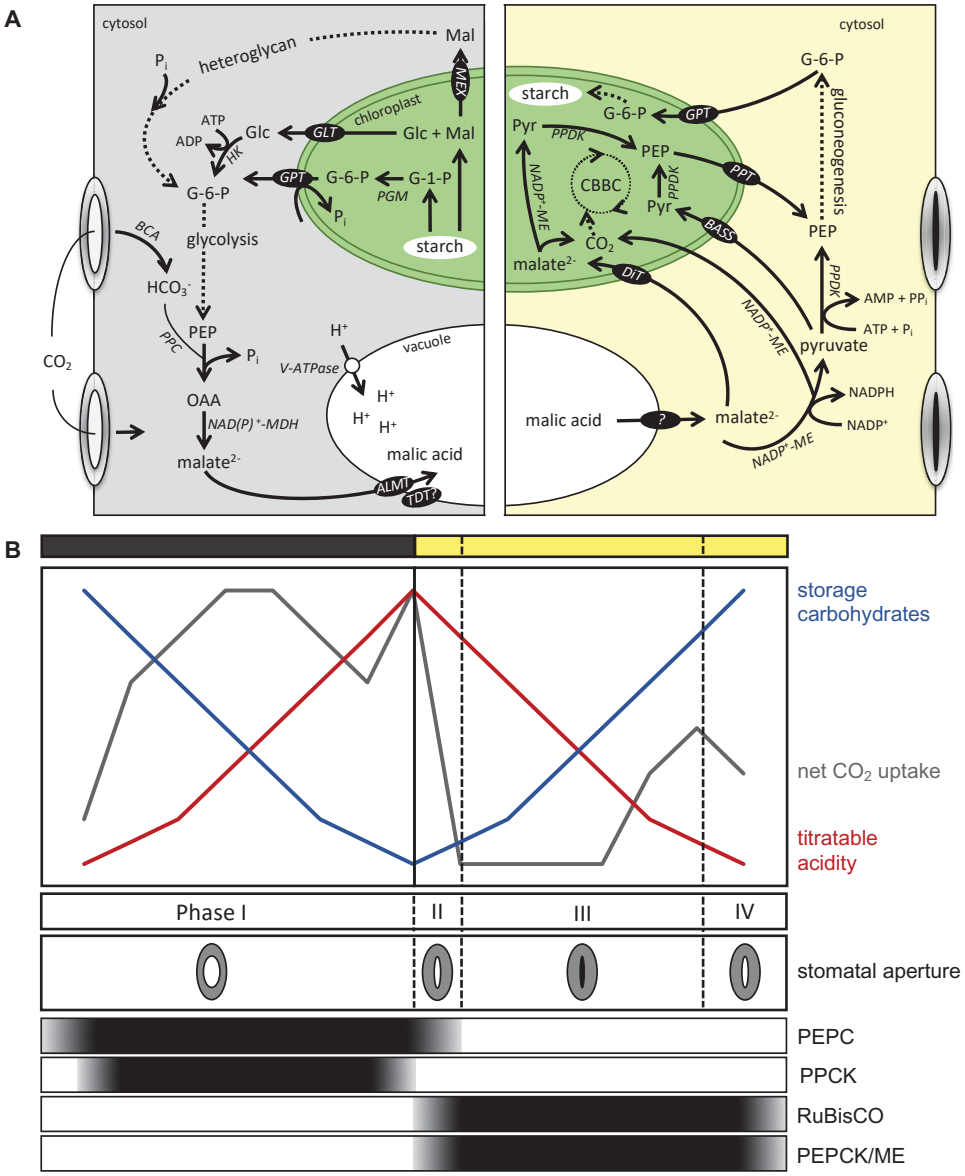


Figure 1. Crassulacean acid metabolism (CAM). (A) Key enzymatic reactions of an NADP-ME, starch-storing species such as *Talinum triangulare*. Reactions taking place in the dark phase of the diel cycle are highlighted with the grey background. Involvement of several transporters is expected but many are not characterized in CAM species yet (marked with "?"). (B) Diel patterns of key CAM metabolites and enzymatic activities. ALMT, aluminium-activate malate transporter; BASS, bile acid:sodium symporter protein; BCA, β-carbonic anhydrase; CBBC, Calvin-Benson-Bassham cycle; DiT, dicarboxylate transporter; G-1-P, glucose-1-phosphate; G-6-P, glucose-6-phosphate; Glc, glucose; Mal, maltose; GLT, glucose transporter; GPT, glucose-6-phosphate/phosphate translocator; MEX, maltose exporter; NAD(P)$^+$-MDH, NAD(P)$^+$-dependent malate dehydrogenase; NAD(P)$^+$-ME, NAD(P)$^+$-dependent malic enzyme; OAA, oxaloacetate; PEP, phosphoenolpyruvate; PEPC, phosphoenolpyruvate carboxylase; PEPCK, phosphoenolpyruvate carboxykinase; PGM; phosphoglucomutase; PPCK, phosphoenolpyruvate carboxylase kinase; PPDK, pyruvate, phosphate dikinase; PPT, phosphate/phosphoenolpyruvate translocator; Pyr, pyruvate; RuBisCO, ribulose-1,5-bisphosphate carboxylase/oxygenase; TDT, tonoplast dicarboxylase transporter. Modified after Christopher and Holtum (1996) and Borland *et al.* (2009).

An alternative decarboxylation route is by the action of cytosolic phospho*enol*pyruvate carboxykinase (PEPCK). In this scenario, $NAD(P)^+$-MDH converts malate to OAA first. At the costs of one ATP per an OAA molecule, PEPCK breaks OAA down to CO_2 and PEP directly in the cytosol (Christopher and Holtum, 1996). In a survey of 35 CAM species, no PEPCK activity could be measured for species with high ME activities. This is in contrast to PEPCK-type species showing minor measurable activities of ME even in presence of substantial PEPCK activity (Dittrich *et al.*, 1973).

Once malic acid reserves are depleted towards the end of the light phase (Phase IV), intracellular CO_2 concentration declines, causing opening of stomata to allow net CO_2 uptake and its assimilation primarily via RuBisCO until the onset of dark period (Fig. 1B; Cushman and Bohnert, 1997; Liu *et al.*, 2018). By the end of Phase IV, C_3-backbones released upon malate decarboxylation were converted to storage carbohydrates (*e.g.* starch in *T. triangulare*) via gluconeogenesis and readily available to re-generate the PEP pool needed for PEPC activity (Fig. 1).

Interplay with carbohydrate metabolism

The CAM cycle imposes substantial requirements on carbohydrate metabolism and the size of carbohydrate reserves is the main determinant of magnitude of CAM expression. For example, 40 to 60% of nocturnally degraded starch fuels PEP synthesis for nocturnal CO_2 fixation in *M. crystallinum* in the CAM mode. Besides providing carbon backbones for PEP, carbohydrate reserves must support other processes such as dark respiration and growth (Borland and Dodd, 2002; Borland *et al.*, 2009, 2016). Thus, detailed understanding of carbohydrate metabolism of CAM species and its interplay with the diurnal (de)carboxylation is crucial for a successful CAM biodesign.

Significant level of diversity in the preferred carbohydrate storage form exists among CAM species, ranging from starch to soluble sugars (fructose, glucose, sucrose), fructans, and combination of soluble sugars and fructans in some monocots (Christopher and Holtum, 1996; Niechayev *et al.*, 2019). Regardless of the preferred type of storage carbohydrates, the diurnal fluctuation of the carbohydrate pool is reciprocal to the fluctuation of organic acids (Fig. 1B).

Starch-storing species such as *M. crystallinum* and *T. triangulare* accumulate starch in their chloroplasts. In contrast to C_3 species, the primary route for nocturnal starch degradation in CAM species is via the phosphorylitic pathway leading to glucose-6-phosphate (G-6-P) production. Based on transcript accumulation as well as measurable transport rates in *M. crystallinum*, GPT was determined as the key transporter connecting carbohydrate metabolism to the core CAM activities. Besides mediating G-6-P export from the chloroplast, GPT also acts as an importer to support starch synthesis in the light phase (Neuhaus and Schulte, 1996; Häusler *et al.*, 2000). Altered *GPT* transcript levels and/or expression patterns were reported in recent transcriptomics studies in other species as well (Kore-eda *et al.*, 2005; Cushman *et al.*, 2008*b*; Brilhaus *et al.*, 2016; Wai *et al.*, 2019).

Utilisation of the phosphorylitic route of starch degradation, instead of hydrolytic night-time degradation as in C_3 species, provides two advantages to CAM species. Firstly, conversion of G-6-P to PEP yields one ATP, which is required in large amounts to energise nocturnal import of malate to the vacuole (Holtum *et al.*, 2005). Secondly, G-6-P activates PEPC allosterically and could thus contribute to regulation of the CAM pathway (Takahashi-Terada *et al.*, 2005).

In species relying on accumulation of soluble sugars, vacuole is the central compartment of carbohydrate metabolism, with import to and export from the vacuole being likely mediated by distinct transport proteins. Sucrose is imported to the vacuole during the light period and hydrolysed to glucose and fructose in a reaction catalysed by invertase such as in pineapple, or sucrose is converted to fructans enzymatically as described in *Agave* (Black *et al.*, 1996; Valluru and Van Den Ende, 2008; Cushman *et al.*, 2015). In pineapple, candidate *TONOPLAST SUGAR TRANSPORTERs (TST)* have been identified in pineapple (Ming *et al.*, 2015). In the case of Arabidopsis, TST proteins transport sucrose, glucose and fructose (Wormit *et al.*, 2006; Schulz *et al.*, 2011). Candidate sugar exporters localized in the tonoplast include *SUCROSE TRANSPORTER 4 (SUT/SUC4)* and glucose exporter *EARLY RESPONSE TO DEHYDRATION (ERDL6)*, orthologs of which were identified in the pineapple transcriptome (Borland *et al.*, 2016).

Besides variance in utilized transitory carbohydrates among CAM species, the CAM pathway is plastic enough to adapt to current conditions as demonstrated in bromelid *Aechmea* 'Maya'. Under conditions of prolonged water stress, which are accompanied by starch depletion, *Aechmea* switches flexibly to use sucrose for PEP generation instead of starch (Ceusters *et al.*, 2009). Maintenance of sufficient pool of transitory carbohydrates is crucial for sustained nocturnal carboxylation but at the same time, it also contributes to regulation of key enzymatic activities. In starch-deficient *pgm* mutant of *M. crystallinum*, transcript levels of *PPC* and *PHOSPHOENOLPYRUVATE CARBOXYLASE KINASE (PPCK)*, both believed to be under circadian control, were reduced compared to the wild type and increased upon sugar feeding (Taybi *et al.*, 2017). Control of CAM-related activities by the metabolic status is another aspect of CAM regulation, which just started to be explored.

Regulation of the CAM pathway

The CAM pathway can operate efficiently only under a strict control of primary and secondary CO_2 assimilation. The exact molecular mechanisms have just started being deciphered and emerging evidence shows how complex the regulation is, happening at the level of transcriptional, post-translational and metabolic control, with a contribution of the circadian clock.

The role of the circadian clock

The role of the endogenous circadian oscillator on the CAM cycle was suggested already more than 30 years ago based on the ability of *Kalanchoë (Bryophyllum) fedtschenkoi* to maintain rhythmical CO_2 uptake typical of the CAM pathway and periodical changes in PEPC phosphorylation status under various constant conditions, including constant light, constant darkness and CO_2-free air, upon a period of entrainment (Warren and Wilkins, 1961; Nimmo *et al.*, 1984, 1987; Wilkins, 1992). Early experiments with inhibitors of protein synthesis suggested endogenously controlled, rhythmical synthesis of the dedicated PPCK (Carter *et al.*, 1991), the nature and transcriptional control of which was revealed later (Hartwell *et al.*, 1999; Taybi *et al.*, 2000; Mallona *et al.*, 2011). Only recently, thanks to *Kalanchoë* RNAi lines, effect of loss of core CAM genes on rhythmical expression of the clock genes was demonstrated (Dever *et al.*, 2015; Boxall *et al.*, 2017, 2020).

Identification of orthologs of known Arabidopsis clock genes in *M. crystallinum, K. fedtschenkoi* and *Opuntia ficus-indica* confirmed common components of the circadian oscillator between the two species. While transcripts of the core clock components fluctuate periodically in all three species with a high degree of conservation, several exceptions were observed. In the facultative CAM species *M. crystallinum*, abundance patterns are the in both C_3 and CAM mode except for *ZEITLUPE (ZTL)*, transcript levels of which fluctuate in both CO_2 assimilation modes in *M. crystallinum* but remain constant in Arabidopsis (Boxall *et al.*, 2005). In obligate CAM species *O. focus-indica*, some clock genes adapted their expression to 24-hr periodicity and other to a period of 12 hrs, suggesting differentiated interaction of some clock components compared to Arabidopsis (Mallona *et al.*, 2011). In *K. fedtschenkoi*, a shift in gene expression of the evening complex was observed (Moseley *et al.*, 2018). Taken together, presence of orthologous clock genes and their periodical expression in both C_3 and CAM species provide evidence that the circadian clock already in place in C_3 species might be with certain modifications sufficient to control the timing of CAM activities (Boxall *et al.*, 2005; Mallona *et al.*, 2011).

On one hand, motif enrichment analysis of pineapple promoter sequences revealed presence of TFBS of several core clock components in genes co-expressed in the green leaf tissue (as compare to non-photosynthetic white tissue), which included also several core CAM genes (Wai *et al.*, 2017). On the other hand, loss of PPC, PPDK or NAD-ME activities perturbated periodical fluctuations in transcript levels of several clock genes in *Kalanchoë* spp, including the core components *CCA1* and *TOC1* (Dever *et al.*, 2015; Boxall *et al.*, 2017). In the light of the altered oscillation of several core components of the endogenous clock when CAM-related enzymatic activities were disturbed, Boxall *et al.* (2017, 2020) proposed that the resulting metabolic changes affect the circadian oscillator itself. The comparison of diel abundance patterns between C_3 Arabidopsis and obligate CAM species *Agave americana* revealed largely overlapping patterns between the two species except for *REVEILLE 1 (RVE1)*, peak expression of which was phase shifted from morning to midnight in *A. americana*. For its role as a "regulatory clock output" gene, *RVE1* would be a good candidate to connect regulation of the circadian oscillator to the metabolic status (Yin *et al.*, 2018).

Despite aspects of the circadian control of the CAM pathway still remaining unclear, it is likely that the circadian clock regulates the necessary build-up of CAM enzymes at transcriptional, post-transcriptional and translational level, while post-translation regulation would primarily regulate enzymatic activities over the 24-hr cycle (Cushman *et al.*, 2000). This is not the case only for core CAM enzymes but also for enzymes of starch turnover in starch-storing CAM species, vacuolar ATPases, and enzymes of glycolysis/gluconeogenesis as revealed thanks to transcriptional studies on the facultative CAM species *M. crystallinum* as well as on obligate CAM species pineapple and *A. americana* (Hartwell, 2005; Cushman *et al.*, 2008*b*; Abraham *et al.*, 2016). Number of glycolytic enzymes and enzymes of starch metabolism are subjected to circadian control in C_3 Arabidopsis as well (Harmer *et al.*, 2000; Smith *et al.*, 2004) and this is true for *M. crystallinum* in C_3 mode of carbon assimilation. When operating in stress-induced CAM mode, however, many genes display shifts in peak transcript accumulation and control by the circadian oscillator could explain such re-programming (Cushman *et al.*, 2008*b*). Several examples of time-dependent regulation of CAM enzymes are covered below.

(Post-)transcriptional regulation of PPCK

PPCK transcript levels (alongside with *e.g. NADP-ME* and *PPDK*) fluctuate in obligate CAM species *O. ficus-indica* and accumulate to higher levels in *T. triangulare*, *M. crystallinum* and *S. album* when assimilating in CAM mode upon exposure to stress (Michalowski *et al.*, 1989; Cushman, 1992; Mallona *et al.*, 2011; Brilhaus *et al.*, 2016). Up to date, PPCK remains the only well-characterized control point of the CAM cycle under the circadian control (Hartwell *et al.*, 2016).

PPCK is a Ca^{2+}-independent kinase regulating PPC activity post-translationally in response to light in C_3 and C_4 species. In CAM species, *PPCK* transcript abundance is under the control of the circadian clock, with progressive accumulation during the dark accompanied by *de novo* synthesis of PPCK each night (Nimmo *et al.*, 1984, 1987; Jiao and Chollet, 1991; Duff and Chollet, 1995; Hartwell *et al.*, 1996, 1999; Taybi *et al.*, 2017). The more surprising was the finding that PPCK activity is not essential for nocturnal CO_2 assimilation as demonstrated in a *K. fedtschenkoi* RNAi line. It is however critical for fine-tuning of the PEPC activity, thus enabling malate accumulation to higher level and ultimately faster growth (Boxall *et al.*, 2017). Besides periodical expression in both obligate and facultative CAM species (Hartwell *et al.*, 1996, 1999; Taybi *et al.*, 2000, 2004; Mallona *et al.*, 2011; Abraham *et al.*, 2016), *PPCK* transcript abundance increases in facultative CAM species when assimilating in the CAM mode (Li and Chollet, 1994; Brilhaus *et al.*, 2016).

With increasing knowledge about molecular regulatory mechanism in general and new genomic resources being generated for CAM species, the contribution of post-transcriptional regulation to control the CAM cycle has started to emerge. Expression profiling in pineapple revealed numerous microRNAs (miRNAs) exhibiting strong diel cycling and candidate miRNAs contributing to regulation of *PPCK1* and *MDH* were identified (Wai *et al.*, 2017). Bai *et al.* (2019) focused on expression of long non-coding RNAs (lncRNA) in the same species and observed tissue-specific expression profiles of lncRNAs as well as diurnal fluctuation in transcript abundance for nearly half of the annotated lncRNAs. Furthermore, lncRNAs acting as competing endogenous RNAs (ceRNAs) to *PPC* and *PPCK* mRNAs were identified and competition of ceRNAs and mRNAs for binding of the regulatory miRNAs identified earlier was proposed in this work. While that many layers to regulate CAM-specific activities may not be necessary for a successful CAM biodesign, they may provide a fine-tuning mechanism (Bai *et al.*, 2019).

Post-translational regulation of PEPC and PPDK

Observations regarding periodical fluctuations of *PPC* transcript levels are largely unambiguous with no major changes reported for *O. ficus-indica* over the diel cycle, while *PPC* transcripts accumulated 25-fold at midnight in *T. triangulare* and 14-fold in *M. crystallinum* when assimilating in the CAM mode (Li and Chollet, 1994; Mallona *et al.*, 2011; Brilhaus *et al.*, 2016). In *M. crystallinum*, PEPC fluctuation was also detected at the protein level (Taybi *et al.*, 2017), while constant protein levels were observed in *K. fedtschenkoi* (Dever *et al.*, 2015; Boxall *et al.*, 2017).

Post-translational regulation by phosphorylation is however described in detail. (De)phosphorylation of PEPC is catalysed by two distinct enzymes: a PROTEIN PHOSPHATASE TYPE 2A (PP2A) and PPCK. Since PP2A extractable activity remains nearly constant over the diel cycle, PPCK – under a strict circadian control at the transcriptional level itself – was determined as a key component of

the temporal regulation of PEPC activity (Hartwell *et al.*, 1999; Taybi *et al.*, 2000, 2004; Boxall *et al.*, 2017). In dephosphorylated state, PEPC is sensitive to malate inhibition, while PPCK-mediated phosphorylation at an N-terminal Ser dramatically decreases enzyme's sensitivity to malate and thus contributes to sustained nocturnal CO_2 fixation (Nimmo *et al.*, 1984; Taybi *et al.*, 2000; Boxall *et al.*, 2017).

PPDK is essential for gluconeogenic recovery of storage carbohydrate reserves in ME species. At the protein level, nearly steady abundance was detected in *K. fedtschenkoi* but with major changes in the phosphorylation status over the 24-hr cycle (Dever *et al.*, 2015). These are achieved by a dedicated bifunctional REGULATORY PROTEN (RP), which deactivates PPDK by phosphorylation and its kinase activity ensures activation of PPDK activity. In chloroplast of C_3 species as well as in mesophyll cells of C_4 species, PPDK is phosphorylated upon illumination (Slack, 1968; Burnell and Hatch, 1983; Chastain *et al.*, 2002). However, in obligate CAM species, such as *K. laxiflora*, PPDK is present in the phosphorylated form primarily during the dark phase, while in the light, PPDK is activated by dephosphorylation (Dever *et al.*, 2015). Contribution of transcriptional regulation cannot be excluded as *PPDK* transcripts accumulated both in *M. crystallinum* and *T. triangulare* assimilating in the CAM mode (Michalowski *et al.*, 1989; Brilhaus *et al.*, 2016).

Regulation of stomatal behaviour

Stomata opening in C_3 plants is regulation by light. Activation of the light signalling cascade leads to activation of H^+-ATPase in plasma membrane of the guard cells, generating the driving force for ion accumulation needed to increase the cell volume leading to stomata opening (Ogawa *et al.*, 1978; Zeiger, 1990; Kinoshita and Shimazaki, 1999; Shimazaki *et al.*, 2007). Since the pattern of stomatal opening is reversed in CAM species, understanding of stomatal regulation in CAM species is crucial for a successful CAM biodesign.

In the past, the effect of CO_2 concentration as well as of blue light (BL) was examined in *Kalanchoë* spp. and while interesting insights were obtained, many aspects of stomatal regulation in CAM species remain unknown. Regulation of stomatal aperture by internal CO_2 concentration is an intriguing opportunity as it can reach 1% or more, a concentration which induces stomatal closure in C_3 species (Cockburn *et al.*, 1979; Gotoh *et al.*, 2019). Exposure of *Kalanchoë daigremontiana* and *K. pinnata* to low ambient CO_2 revealed induction of stomatal opening specifically in the latter half of the light period and in the dark, and a presence of CO_2 sensor was proposed to initiate stomatal opening (von Caemmerer and Griffiths, 2009). Such a sensor does not necessarily have to be a dedicated cellular receptor, but the internal CO_2 concentration might be sensed at the level of changing metabolite levels. Mesophyll-derived apoplastic [$malate^{2-}$] fulfils such a function in C_3 species and would enable adjustment of stomatal behaviour to altered (de)carboxylation biochemistry not only during CAM induction in facultative CAM species but also along the evolutionary trajectory from C_3 to CAM (Borland *et al.*, 2014)

Investigations of the effect of blue light on stomatal opening led to contradicting findings (Tallman *et al.*, 1997; Grams and Thiel, 2002). Loss *PHOT2* in *K. fedtschenkoi* led to reduced stomatal conductance in phase IV of the CAM cycle (*i.e.* stomata re-opening in late afternoon to allow CO_2 fixation by RuBisCO) and stomata responsiveness to BL was restricted to the phase IV only (Gotoh *et al.*, 2019; Liu *et al.*, 2019). BL-dependent modulation of stomatal aperture via a conserved signalling cascade is likely

important towards the end of the light period but does not explain the reversed pattern of stomatal opening in CAM species (Gotoh *et al.*, 2019). Regulation by the circadian oscillator might play a role instead.

Insights into the circadian regulation of stomatal behaviour were obtained thanks to transcriptomic studies with *A. americana* and *M. crystallinum*. Comparison of diel expression patterns of CO_2-sensing and BL-responsive genes between *A. americana* and Arabidopsis revealed largely overlapping expression patterns. In contrast, temporal patterns of several components of ABA signalling, including *HIGH LEAF TEMPERATURE 1 (HT1)* acting as a central regulator of stomatal CO_2 signalling, and K^+ and Cl^- channels were phase shifted in *A. americana* or even reciprocal to patterns in Arabidopsis. This suggests temporal re-programming of some genes regulating stomatal behaviour in obligate CAM species (Abraham *et al.*, 2016). Facultative CAM species present a special case as the pattern of stomatal opening changes during CAM induction. Kong *et al.* (2019) investigated underlying transcriptional changes by the means of guard cell transcriptome profiling. They observed altered transcript abundance and temporal patterns of several core CAM genes in *M. crystallinum* upon exposure to salt stress, which indicates that guard cells themselves undergo a transition from C_3 to CAM during CAM induction. It will be interesting to investigate, whether alternations in metabolic status of the guard cells are sufficient to regulate stomatal opening.

CAM engineering

CAM engineering promises an opportunity to improve WUE of conventional C_3 (and even C_4) crops, thus reducing irrigation requirements for food and biomass production and/or enabling expansion of agricultural production to areas with limited water availability. There are several pieces of evidence supporting the feasibility of this strategy: i) multiple independent evolutionary trajectories from ancestral C_3 species to CAM [*e.g.* Silvera *et al.* (2009), Heyduk *et al.* (2019)], ii) existence of facultative CAM species, demonstrating compatibility of the CAM pathway with C_3 (and C_4) genetic background(s), and iii) operation of the entire pathway within a single cell, unlike in C_4 photosynthesis relying on separation of primary and secondary carbon assimilation between two specialized cell types (Borland *et al.*, 2014; Yang *et al.*, 2015).

In contrast to general agreement on biochemical characteristics of a typical CAM cycle, a detailed genetic blueprint enabling this complex trait to arise remains unknown. Current understanding of genes (and biological functions where candidate genes are still missing) involved in the CAM pathway is summarized in Tab. 1. However, there are two major limitations to keep in mind. Firstly, an introduction of a single gene to the target crop or altered regulation thereof may not be sufficient to induce a trait as complex as CAM and thus, when testing candidate genes functionally, these should be introduced in functional modules (*e.g.* carboxylation, decarboxylation, stomatal regulation and leaf anatomy) to speed up identification of necessary genes and isoforms (Borland *et al.*, 2014; DePaoli *et al.*, 2014; Liu *et al.*, 2018). Secondly, the list of candidate genes for a successful CAM biodesing is not complete. For several features of the CAM trait (*e.g.* malate import to and export from the vacuole), presence of respective genes in expected and Arabidopsis orthologs are known, but their confirmation and functional characterization in CAM species is missing so far. It is also likely that molecular players with regulatory functions, such as kinases, phosphatases and transcription factors, have not been fully identified yet. (Cushman and Bohnert, 1997; Borland *et al.*, 2014; Yang *et al.*, 2015; Goolsby *et al.*, 2018).

The modular strategy outlined above promises faster identification of molecular players required for CAM biodesign but so far, *PPC* remains the only described example of a gene where its coding sequence has evolved into CAM-specific lineages (Christin *et al.*, 2014; Deng *et al.*, 2016). It remains to be answered whether this is also the case for other components of the CAM cycle or whether regulatory changes are the major driver of CAM evolution. Once the minimal set of "CAM building blocks" is identified, their appropriate expression over the 24-hr cycle must be ensured by employment of suitable promoters, *cis*- and *trans*-regulators. Furthermore, corresponding post-translational regulation will likely be required for at least some enzymes as described for PEPC and PPDK (Hartwell *et al.*, 1996; Chastain and Chollet, 2003). To avoid, futile cycling which would lead to a wasteful consumption of large amounts of ATP and NADPH, it will likely be necessary to modify activities of the host's endogenous genes, thus requiring parallel research at the side of promising host species as well This might be for example the case for the enzymes of CBBC and their regulators such as RuBisCO activase (Portis Jr, 1990; Borland *et al.*, 2014).

Besides avoiding competing activities of C_3- and C_4-carboxylases, sufficient and phase-dependent metabolite flow through the (de)carboxylation reactions and metabolite exchange between organelles, adequate storage capacities – both for malic acid and storage carbohydrates – have to be ensured (Silvera *et al.*, 2010; Borland *et al.*, 2014). Enlarged vacuoles accommodating accumulation of nocturnally synthesized malic acid translate into succulent leaves and stems, especially in obligate CAM species (Nelson and Sage, 2008). Therefore, the aspects of vacuolar size and leaf anatomy should not be omitted while developing strategies for CAM engineering. Recently, cell size of model species Arabidopsis and *Nicotiana sylvestris* could be increased by over-expression of fruit-specific basic helix-loop-helix (bHLH) transcription factor of *Vitis vinifera*, providing evidence that it is feasible to modify anatomical features through a single gene change (Lim *et al.*, 2018). Besides enlarged vacuoles, leaves of CAM species have reduced internal air space as well as reduced surface area of chloroplast-containing mesophyll cells, thus limiting leaf internal conductance to CO_2. In CAM biodesign, it will therefore be essential to find a balance between CO_2 diffusion into the leaf and its concentrating in the proximity of RuBisCO (Maxwell *et al.*, 1997; Borland *et al.*, 2014).

Another challenge in the CAM biodesign is localization of CAM-specific isoforms, both at sub-cellular level (see Tab. 1 for examples) and in specific cell types. Most of our knowledge about enzyme localization is based on laborious measurements of enzymatic activities in fractions enriched for a specific cell type [*e.g.* CA by Tsuzuki *et al.* (1982)] or immunoelectron microscopy [*e.g.* PPDK by Kondo *et al.* (2001)] and only recently, several GFP-tagged *M. crystallinum* CAM-cycle enzymes and regulatory proteins were heterologously expressed in Arabidopsis (Lim *et al.*, 2019). Cell type-specific localization gains on importance in the context of stomatal regulation, which is mediated via guard cell-specific transporters and kinases.

Taken together, successful CAM biodesign requires temporal regulation of numerous enzymes and regulatory proteins as well as metabolite flows, which also need to be sufficiently strong. A critical factor of CAM biodesign is to ensure sufficient storage capacity of the vacuole, which might also require anatomical modifications (Borland *et al.*, 2014; DePaoli *et al.*, 2014; Sweetlove *et al.*, 2017). Established plant models could be used to characterize candidate genes for CAM engineering. In the last decade, photosynthetic performance of Arabidopsis was improved by overexpression of individual CAM genes.

Nocturnal expression of exogenous PEPC resulted into a higher proportion of opened stomata in the dark together with increased CO_2 assimilation rates (Kebeish *et al.*, 2012). Similarly, shoot biomass was increased upon overexpression of single CAM genes (Lim *et al.*, 2019). While these pioneering works provide the very first evidence of the feasibility of CAM engineering, deeper understanding of CAM regulation is needed, including identification of CAM-specific isoforms and regulatory mechanism ensuring proper timing of enzymatic activities. This is the area where genomics can contribute largely.

The role of genomics

Even though the field of CAM research has over 30 years long history, the studies were often restricted to phenotypical characterisation, such as measurements of titratable acidity, gas exchange and carbon isotope composition analyses (*e.g.* Holtum *et al.*, 2016, 2017; Winter and Holtum, 2017; Winter, 2019). Only a limited number of species was investigated biochemically and at the genetic level, with *M. crystallinum* and *Kalanchoë* spp. being the primary experimental models, and first mutant collections have been created only recently (Cushman *et al.*, 2008*a*; Liu *et al.*, 2019). Combining well-established biochemical and physiological methods together with rapidly developing sequencing technologies and methods of genome editing have a potential to expand our knowledge at a higher pace but also for more species, covering the CAM diversity.

Among monocots, transcriptome assemblies of *A. deserti* and *A. tequilana* (Gross *et al.*, 2013) as well as genome assemblies of orchids *Phalaenopsis equestris* (Cai *et al.*, 2014) and *D. catenatum* (Zhang *et al.*, 2016) were obtained, RNA-seq datasets also being available for the latter (Zou *et al.*, 2018). The presence of pineapple genome (Ming *et al.*, 2015) enabled genome-wide identification and classification of protein kinases (Zhu *et al.*, 2018) and it also remains the only CAM species up to date, for which temporal expression of microRNAs was examined (Wai *et al.*, 2017). Abraham *et al.* (2016) performed a comprehensive study of transcript, protein and metabolite abundance of *A. americana* over the diel cycle. Comparative studies include gene expression analysis of four *Agavoideae* species, which revealed a correlation between transcript abundance of CAM-related genes and strength of the CAM phenotype (Heyduk *et al.*, 2018), and co-expression gene network analysis of C_3 and CAM *Erycina* species, suggesting differentiation at the level of transcriptional cascades depending on carbon assimilation strategy (Heyduk *et al.*, 2019*a*). Fewer resources exist for dicots but include genome assembly of *K. fedtschenkoi* (Yang *et al.*, 2017), RNA-seq datasets of facultative CAM species *M. crystallinum* (Tsukagoshi *et al.*, 2015) and *T. triangulare* (Brilhaus *et al.*, 2016). Especially facultative CAM species are useful experimental models to understand CAM regulation as they enable direct comparison of C_3 and CAM states (Yang *et al.*, 2019).

Table 1. Likely targets for CAM engineering and assignment of genes to modules depending on their function in the context of the CAM pathway. It remains to be determined whether temporal adjustment of protein activities would be sufficient or whether specific isoforms must be introduced to the host.

CAM module	Gene	Protein localisation	Function	Reference
Carboxylation	*β-CA*	cytosol and/or chloroplast	CO_2 hydration to HCO_3^-	Borland *et al.* (2014)
	PPC	cytosol	nocturnal carboxylation	Borland *et al.* (2014), Goolsby *et al.* (2018)
	dark-phase active *PPCK*	cytosol	phosphorylation of PEPC	Borland *et al.* (2014)
	malate importer (*ALMT*?)	tonoplast	import to the vacuole	
Decarboxylation *	malate exporter (*tDT*?)	tonoplast	export from the vacuole	Borland *et al.* (2014), Goolsby *et al.* (2018)
	NADP-ME	cytosol and/or chloroplast	malate decarboxylation	Borland *et al.* (2014)
	PPDK	cytosol and/or chloroplast	PEP regeneration	Borland *et al.* (2014)
	PPDK-RP	cytosol and/or chloroplast	time-dependent regulation of PPDK activity	Borland *et al.* (2014)
	DiT1/DiT2	chloroplast	malate import to chloroplast	Goolsby *et al.* (2018)
	BASS2	chloroplast	pyruvate import to chloroplast (co-transport with Na^+)	Goolsby *et al.* (2018)
	PPT	chloroplast	PEP export from chloroplast	Goolsby *et al.* (2018)
	NHD	chloroplast	Na^+/H^+ antiporter establishing Na^+ gradient across the membrane	Goolsby *et al.* (2018)
	NAD(P)-ME	mitochondrion	malate decarboxylation	Borland *et al.* (2014)
	DiC2	mitochondrion	malate import to mitochondrion	Goolsby *et al.* (2018)
Stomatal regulation	guard cell-specific *β-CA*		regulation of CO_2-controlled stomatal movement	Borland *et al.* (2014)
	high leaf temperature 1 (HT1)		ABA-dependent stomatal regulation	Borland *et al.* (2014), Abraham *et al.* (2016)
	slow anion channel associated 1 (SLAC1)		regulation of stomatal aperture	Borland *et al.* (2014)
	unknown		regulation of malate transport (vacuole-cytosol-apoplast)	Borland *et al.* (2014)
Carbohydrate metabolism	glycolytic enzymes		feeding into carboxylation	Borland *et al.* (2014)
	enzymes of gluconeogenesis		utilisation of decarboxylation products	Borland *et al.* (2014)
	enzymes of starch degradation		feeding into glycolysis	
	enzymes of starch synthesis: *APL*		feeding into gluconeogenesis	Goolsby *et al.* (2018)
	sugar transporters	chloroplast, tonoplast**	Transport to support turnover of transitory storage carbohydrates	Borland *et al.* (2014)
Stomatal movement	unknown		control of stomatal regulation	Borland *et al.* (2014)
Anatomy	cell elongation bHLH protein (TF)		increased cell size	Nicolas *et al.* (2013), Borland *et al.* (2014)
	xyloglucan endotransglucosylase hydrolase		increased leaf water storage and succulence	Han *et al.* (2013), Borland *et al.* (2014)

* Depending on carboxylation strategy, either a suitable malic enzyme isoform in combination with PPDK or PEPCK only.
** Vacuolar sugar gain on importance in species relying on accumulation of soluble sugars

Despite expanding genomic resources, a finite list of genes required for successful CAM biodesign is lacking. Firstly, a possibility exists that target genes to be transferred belong to multimember families as described for *PPC* (Deng *et al.*, 2016), opening a room for neo- or sub-functionalization (Emms *et al.*, 2016; Heyduk *et al.*, 2018) or even dosage effect on the strength of CAM as observed for examined orchids in *Oncidiinae* subtribe (Silvera *et al.*, 2014). It is therefore desirable to identify the orthologs recruited in different CAM linages or depending on CAM biochemistry employed (Yang *et al.*, 2015). Interspecies comparisons covering phylogenetic and CAM diversity will be useful in this respect to shed more light on distinct gene lineages (*e.g.* Christin *et al.*, 2014), identify highly expressed isoforms (*e.g.* Heyduk *et al.*, 2018) or those with altered temporal pattern (*e.g.* Yang *et al.*, 2017) and/or responsiveness to external stimuli leading to CAM induction in facultative CAM species (*e.g.* Cushman *et al.*, 2008*b*). It cannot be excluded that with increased number of sequenced CAM species, genes specific to certain CAM biochemistry (*e.g.* obligate *vs* facultative or depending on decarboxylation strategy) will be discovered and tools of functional genomics will be useful to predict the function of encoded proteins (Yang *et al.*, 2015).

Secondly, all known CAM genes were identified in ancestral C_3 species (Heyduk *et al.*, 2019*a*), which raises the question, whether successful CAM biodesign necessarily requires introduction of new genes into the host species or whether a 'mere' time- or cell type-dependent adjustment of existing protein activities and enhancement of metabolite flows present in C_3 species – in other words mimicking the evolutionary trajectory from C_3 to CAM – will be sufficient for successful CAM engineering (Taybi *et al.*, 2002; West-Eberhard *et al.*, 2011; Yang *et al.*, 2015; Bräutigam *et al.*, 2017). With regard to gene expression regulation, genomics can contribute by analysing *cis*-regulatory elements and regulatory non-coding RNAs and how they affect pattern and magnitude of gene expression (Hibberd and Covshoff, 2010; DePaoli *et al.*, 2014; Mihailescu, 2015).

Abscisic acid signalling

The phytohormone abscisic acid (ABA) contributes to regulation of numerous developmental processes throughout a plant's life – from embryo and seed development, seed dormancy, to germination and early seedling development (Finkelstein, 2013). ABA signalling also plays a critical role in adjustment of physiological processes to the changing environment (Christmann *et al.*, 2006). One of the earliest and best described responses includes induction of stomatal closure (Jones and Mansfield, 1970; Mittelheuser and van Steveninck, 1971). Besides response to abiotic stress, ABA is also involved in response to pathogen attacks (Finkelstein, 2013). At the molecular level, ABA signalling depends on ABA synthesis, transport from the site of synthesis and induction of the ABA signalling pathway, ultimately leading to changes in gene expression of the target genes (Endo *et al.*, 2014).

The ABA signalling cascade begins with ABA binding to devoted receptors, of which several classes were identified: soluble PYRABACTIN RESISTANCE/PYRABACTIN RESISTANCE-LIKE/REGULATORY COMPONENTS OF ABA RECEPTOR (PYR/PYL/RCAR), plasma membrane-localized G-PROTEIN COUPLED RECEPTORs, and H SUBUNIT OF CHLOROPLAST Mg^{2+}-CHELATASE (ChlH) integrated in the chloroplast envelope. The steps of downstream signal transmission are best understood for the soluble PYR/PYL/RCAR receptors, which comprise a family of 14 members in Arabidopsis (Shen *et al.*, 2006; Liu *et al.*, 2007; Park *et al.*, 2009; Cutler *et al.*, 2010; Guo *et al.*, 2011).

ABA binding to PYR/PYL/RCAR receptors induces a conformational change increasing affinity of the receptors to bind clade A PROTEIN PHOSPHATASEs 2C (PP2CA), acting as negative regulators of ABA signalling (*e.g.* ABA-INSENSITIVE 1 (ABI1), ABI2, and HYPERSENSITIVE TO ABA 1 (HAB1) in Arabidopsis). As a result of this protein-protein interaction, SNF1-RELATED PROTEIN KINASEs 2 (SnRK2), comprising ten members in Arabidopsis, are released free and can maintain their auto-phosphorylated status. In their active form, SnRK2s post-translationally modulate the activity of their downstream targets, which besides a variety of transcription factors (TFs) also include guard cell-localized anion channels, components of reactive oxygen species homeostasis and chloroplast processes (Park *et al.*, 2009; Cutler *et al.*, 2010; Guo *et al.*, 2011; Kulik *et al.*, 2011; Wang *et al.*, 2013).

Expression of nearly one tenth of Arabidopsis genes is regulated by ABA via a complex regulatory network (Nemhauser *et al.*, 2006; Song *et al.*, 2016). Many components of ABA-dependent stress signalling overlap between different stressor, such as *cis*-elements designated as ABA-responsive elements (ABREs) in promoter sequences of numerous stress-responsive genes. These serve as TFBS for bZIP-type ABA-RESPONSIVE ELEMENT-BINDING PROTEINs/ABA-BINDING FACTORs (AREB/ABF) induced in response to drought, osmotic and salt stress. Besides numerous structural genes, the target genes under the control of AREB/ABF TFs include a variety of other transcription factors (Uno *et al.*, 2000; Fujita *et al.*, 2005, 2013; Yoshida *et al.*, 2010).

ABA signalling in CAM induction

In facultative CAM species, the CAM pathway is induced upon exposure to stress, which makes ABA a suitable signalling molecule to activate this alternative mode of carbon assimilation in response to altered environmental conditions. From first physiological experiments with *M. crystallinum* to recent transcriptome studies in *T. triangulare*, *A. americana*, and C_3 and CAM *Erycina* orchids, there is an accumulating evidence for the role of ABA in CAM induction and for involvement of components of ABA signalling in regulation of the CAM pathway in obligate CAM species.

Exogenous ABA could mimic the effect of salt or drought stress in *M. crystallinum* as supported by increased nocturnal acidification and PEPC and NADP-ME enzymatic activities. In the tested range of ABA concentrations, the strength of CAM was concentration dependent when applied as a foliar spray. In addition, exogenous ABA could trigger CAM expression also when applied via roots, even though to a lesser extent than salt treatment (Chu *et al.*, 1990; Dai *et al.*, 1994). These observations are in line with increased concentration of endogenous ABA in response to increased salinity and other abiotic stresses (Thomas *et al.*, 1992; Taybi and Cushman, 2002). ABA treatment induced nocturnal acidification also in an epiphytic fern *V. lineata* (Minardi *et al.*, 2014), suggesting that the role ABA in CAM induction is conserved across plant lineages. Evidence for transcriptional regulation of ABA-mediated CAM induction in *T. triangulare* was provided by Brilhaus *et al.* (2016). With progressing water limitation, transcriptional changes indicative to general stress response were accompanied by co-expression of core CAM enzymes, enzymes of carbohydrate metabolism and components of ABA signalling. Remarkably, transcripts of both CAM and ABA-responsive genes returned to non-stress (*i.e.* C_3) levels rapidly and co-ordinately upon re-watering.

Besides facultative CAM species, a role of ABA signalling in CAM was also proposed for obligate CAM species *Erycina pusilla* end *A. americana*. Gene network analysis of a C_3 and an obligate CAM *Erycina (Orchidaceae)* species, revealed differences in the network connectivity depending on photosynthetic mode. Interestingly, the genes with changed connectivity included among other components of ABA signalling (Heyduk *et al.*, 2016). Comparison of expression patterns between Arabidopsis and *A. americana* revealed shifted expression of several components of ABA signalling, among them *SnRK2.6/OST1, SnRK2.10*, several *PP2C*s and *RCAR3* (Abraham *et al.*, 2016). In the context of altered diel expression patterns of several ion channels involved in regulation guard cell turgor in *A. americana* (Abraham *et al.*, 2016), it will be of interest to see, whether ABA regulates stomatal opening in CAM species as proposed by Heyduk *et al.* (2016) even in absence of stress. Additional evidence for the role of ABA in obligate CAM species comes from the transcriptome analysis of pineapple. Promoters of genes showing circadian fluctuation in transcript abundance specifically in the green leaf tissue, which included also several CAM genes such as *PPC* and *PPDK*, were besides clock-associated motifs also enriched for the G-box motif (Ming *et al.*, 2016; Wai *et al.*, 2017). G-box binding proteins represent a diverse group of TFs, which includes numerous ABA-responsive TFs (Menkens *et al.*, 1995; Shinozaki and Yamaguchi-Shinozaki, 1997). Based on the recent findings, it is intriguing to hypothesize that ABA could even control core enzymatic activities of the CAM pathway.

References

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

Arakaki M, Christin P-A, Nyffeler R, Lendel A, Eggli U, Ogburn RM, Spriggs E, Moore MJ, Edwards EJ. 2011. Contemporaneous and recent radiations of the world's major succulent plant lineages. Proceedings of the National Academy of Sciences of the United States of America **108**, 8379–8384.

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

Bartholomew DM, Rees DJG, Rambaut A, Smith JAC. 1996. Isolation and sequence analysis of a cDNA encoding the c subunit of a vacuolar-type H+-ATPase from the CAM plant *Kalanchoë daigremontiana*. Plant Molecular Biology **31**, 435–442.

Black CC, Chen JQ, Doong RL, Angelov MN, Sung SJS. 1996. Alternative Carbohydrate Reserves Used in the Daily Cycle of Crassulacean Acid Metabolism. In: Winter K, Smith JAC, eds. Crassulacean Acid Metabolism. Berlin, Heidelberg: Springer.

Bone RE, Smith JAC, Arrigo N, Buerki S. 2015. A macro-ecological perspective on crassulacean acid metabolism (CAM) photosynthesis evolution in Afro-Madagascan drylands: *Eulophiinae* orchids as a case study. New Phytologist **208**, 469–481.

Borland AM, Dodd AN. 2002. Carbohydrate partitioning in crassulacean acid metabolism plants: reconciling potential conflicts of interest. Functional Plant Biology **29**, 707–716.

Borland AM, Griffiths H, Hartwell J, Smith JAC. 2009. Exploiting the potential of plants with crassulacean acid metabolism for bioenergy production on marginal lands. Journal of Experimental Botany **60**, 2879–2896.

Borland AM, Guo HB, 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, Weston DJ, Schlauch KA, Tschaplinski TJ, Tuskan GA, Yang X, Cushman JC. 2014. Engineering crassulacean acid metabolism to improve water-use efficiency. Trends in Plant Science **19**, 327–338.

Boxall SF, Dever L V, Knerova J, Gould PD, Hartwell J. 2017. Phosphorylation of Phosphoenolpyruvate 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. The Plant Cell **29**, tpc.00301.2017.

Boxall SF, Foster JM, Bohnert HJ, Cushman JC, Nimmo HG, Hartwell J. 2005. Conservation and divergence of circadian clock operation in a stress-inducible Crassulacean acid metabolism species reveals clock compensation against stress. Plant Physiology **137**, 969–982.

Boxall SF, Kadu N, Dever L V, Kneřová 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.

Bräutigam A, Eisenhut M, Schlüter U, Gowik U. 2017. On the evolutionary origin of CAM photosynthesis. Plant Physiology **174**, 473–477.

Brilhaus D, Bräutigam A, Mettler-Altmann T, Winter K, Weber APM. 2016. Reversible Burst of Transcriptional Changes During Induction of Crassulacean Acid Metabolism (CAM) in *Talinum triangulare.* Plant Physiology **170**, 102–122.

Burnell JN, Hatch MD. 1983. Dark-light regulation of pyruvate, Pi dikinase in C_4 plants: Evidence that the same protein catalyses activation and inactivation. Biochemical and Biophysical Research Communications **111**, 288–293.

von Caemmerer S, Griffiths H. 2009. Stomatal responses to CO_2 during a diel Crassulacean acid metabolism cycle in *Kalanchoe daigremontiana* and *Kalanchoe pinnata*. Plant, Cell and Environment **32**, 567–576.

Cai J, Liu X, Vanneste K, *et al.* 2014. The genome sequence of the orchid *Phalaenopsis equestris*. Nature Genetics **47**, 65–72.

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 J, Borland AM, Londers E, Verdoodt V, Godts C, De Proft MP. 2009. Differential usage of storage carbohydrates in the CAM bromeliad *Aechmea* 'Maya' during acclimation to drought and recovery from dehydration. Physiologia Plantarum **135**, 174–184.

Chastain CJ, Chollet R. 2003. Regulation of pyruvate, orthophosphate dikinase by ADP-/Pi-dependent reversible phosphorylation in C_3 and C_4 plants. Plant Physiology and Biochemistry **41**, 523–532.

Chastain CJ, Fries JP, Vogel JA, *et al.* 2002. Pyruvate, orthophosphate dikinase in leaves and chloroplasts of C_3 plants undergoes light-/dark-induced reversible phosphorylation. Plant Physiology **128**, 1368–1378.

Cheffings CM, Pantoja O, Ashcroft FM, Smith J a. 1997. Malate transport and vacuolar ion channels in CAM plants. Journal of Experimental Botany **48 Spec No**, 623–631.

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**, 5.

Christin P-A, Arakaki M, Osborne CP, *et al.* 2014. Shared origins of a key enzyme during the evolution of C_4 and CAM metabolism. Journal of Experimental Botany **65**, 3609–3621.

Christmann A, Moes D, Himmelbach A, Yang Y, Tang Y, Grill E. 2006. Integration of abscisic acid signalling into plant responses. Plant Biology **8**, 314–325.

Christopher JT, Holtum J. 1996. Patterns of Carbon Partitioning in Leaves of Crassulacean Acid Metabolism Species during Deacidification. Plant physiology **112**, 393–399.

Chu C, Dai Z, Ku M, Edwards G. 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**, 1029–1032.

Crayn DM, Winter K, Schulte K, Smith JAC. 2015. Photosynthetic pathways in *Bromeliaceae*: Phylogenetic and ecological significance of CAM and C_3 based on carbon isotope ratios for 1893 species. Botanical Journal of the Linnean Society **178**, 169–221.

Cushman JC. 1992. Characterization and expression of a NADP-malic enzyme cDNA induced by salt stress from the facultative crassulacean acid metabolism plant, *Mesembryanthemum crystallinum*. European Journal of Biochemistry **208**, 259–266.

Cushman JC, Agarie S, Albion RL, Elliot SM, Taybi T, Borland AM. 2008*a*. Isolation and Characterization of Mutants of Common Ice Plant Deficient in Crassulacean Acid Metabolism. Plant

Physiology **147**, 228–238.

Cushman JC, Bohnert HJ. 1997. Molecular genetics of Crassulacean acid metabolism. Plant Physiology **113**, 667–676.

Cushman JJC, Bohnert HHJ. 1999. Crassulacean Acid Metabolism: Molecular Genetics. Annual Review of Plant Biology **50**, 305–332.

Cushman JC, Davis SC, Yang X, Borland AM. 2015. Development and use of bioenergy feedstocks for semi-arid and arid lands. Journal of Experimental Botany **66**, 4177–4193.

Cushman JC, Taybi T, Bohnert HJ. 2000. Induction of Crassulacean Acid Metabolism - Molecular Aspects. In: Leegood RC, Sharkey TD, von Caemmerer S, eds. Photosynthesis: Physiology and Metabolism. Kluwer Academic Publishers, 551–582.

Cushman JC, Tillett RL, Wood JA, Branco JM, Schlauch KA. 2008*b*. Large-scale mRNA expression profiling in the common ice plant, *Mesembryanthemum crystallinum*, performing C_3 photosynthesis and Crassulacean acid metabolism (CAM). Journal of Experimental Botany **59**, 1875–1894.

Cutler SR, Rodriguez PL, Finkelstein RR, Abrams SR. 2010. Abscisic acid: emergence of a core signaling network. Annual Review of Plant Biology **61**, 651–679.

Dai AG. 2013. Increasing drought under global warming in observations and models. Nature Climate Change **3**, 52–58.

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.

Davis ASC, Simpson J, Del K, Gil C, Nicholas A, Tongerlo E Van, Castano NH, Dever L V. 2019. Undervalued potential of crassulacean acid metabolism (CAM) for current and future agricultural production. Journal of Experimental Botany **70**, 6521–6537.

Demmig B, Winter K. 1983. Photosynthetic characteristics of chloroplasts isolated from *Mesembryanthemum crystallinum* L., a halophilic plant capable of Crassulacean acid metabolism. Planta **159**, 66–76.

Deng H, Zhang LS, Zhang GQ, Zheng BQ, Liu ZJ, Wang Y. 2016. Evolutionary history of PEPC genes in green plants: Implications for the evolution of CAM in orchids. Molecular Phylogenetics and Evolution **94**, 559–564.

DePaoli HC, Borland AM, Tuskan GA, Cushman JC, Yang X. 2014. Synthetic biology as it relates to CAM photosynthesis: Challenges and opportunities. Journal of Experimental Botany **65**, 3381–3393.

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.

Dittrich P, Campbell WH, Black CC. 1973. Phosphoenolpyruvate Carboxykinase in Plants Exhibiting Crassulacean Acid Metabolism. Plant Physiology **52**, 357–361.

Duff SM, Chollet R. 1995. In vivo regulation of wheat-leaf phosphoenolpyruvate carboxylase by reversible phosphorylation. Plant Physiology **107**, 775–782.

Emmerlich V, Linka N, Reinhold T, Hurth MA, Traub M, Martinoia E, Neuhaus HE. 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.

Emms DM, Covshoff S, Hibberd JM, Kelly S. 2016. Independent and Parallel Evolution of New Genes by Gene Duplication in Two Origins of C_4 Photosynthesis Provides New Insight into the Mechanism of Phloem Loading in C_4 Species. Molecular Biology and Evolution **33**, 1796–1806.

Endo A, Okamoto M, Koshiba T. 2014. ABA Biosynthetic and Catabolic Pathways. In: Zhang D-P, ed. Abscisic Acid: Metabolism, Transport and Signaling. Springer Netherlands, 191–223.

FAO. 2019. *FAO's Work on Climate Change*.

Finkelstein R. 2013. Abscisic Acid Synthesis and Response. The Arabidopsis Book **11**, e0166.

Franco AC, Haag-Kerwer A, Herzog B, *et al.* 1996. The effect of light levels on daily patterns of chlorophyll fluorescence and organic acid accumulation in the tropical CAM tree *Clusia hilariana*. Trees - Structure and Function **10**, 359–365.

Frei B, Eisenach C, Martinoia E, Hussein S, Chen XZ, Arrivault S, Neuhaus HE. 2018. Purification and

functional characterization of the vacuolar malate transporter tDT from Arabidopsis. Journal of Biological Chemistry **293**, 4180–4190.

Fujita Y, Fujita M, Satoh R, Kyonoshin Maruyama MMP, Seki M, Hiratsu K, Ohme-Takagi M, Shinozaki K, Yamaguchi-Shinozaki K. 2005. AREB1 Is a Transcription Activator of Novel ABRE-Dependent ABA Signaling That Enhances Drought Stress Tolerance in Arabidopsis. The Plant Cell **17**, 3470–3488.

Fujita Y, Yoshida T, Yamaguchi-Shinozaki K. 2013. Pivotal role of the AREB/ABF-SnRK2 pathway in ABRE-mediated transcription in response to osmotic stress in plants. Physiologia Plantarum **147**, 15–27.

Furumoto T, Yamaguchi T, Ohshima-Ichie Y, *et al.* 2011. A plastidial sodium-dependent pyruvate transporter. Nature **476**, 472–475.

Goolsby EW, Moore AJ, Hancock LP, De Vos JM, Edwards EJ. 2018. Molecular evolution of key metabolic genes during transitions to C_4 and CAM photosynthesis. American Journal of Botany **105**, 602–613.

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.

Grams TEE, Thiel S. 2002. High light-induced switch from C_3-photosynthesis to Crassulacean acid metabolism is mediated by UV-A/blue light. Journal of Experimental Botany **53**, 1475–1483.

Gross SM, Martin JA, Simpson J, Abraham-Juarez MJ, Wang Z, Visel A. 2013. De novo transcriptome assembly of drought tolerant CAM plants, *Agave deserti* and *Agave tequilana*. BMC Genomics **14**, 1–14.

Guo J, Yang X, Weston DJ, Chen JG. 2011. Abscisic Acid Receptors: Past, Present and Future. Journal of Integrative Plant Biology **53**, 469–479.

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

Han Y, Wang W, Sun J, *et al.* 2013. *Populus euphratica* XTH overexpression enhances salinity tolerance by the development of leaf succulence in transgenic tobacco plants. Journal of Experimental Botany **64**, 4225–4238.

Harmer SL, Hogenesch JB, Straume M, Chang HS, Han B, Zhu T, Wang X, Kreps JA, Kay SA. 2000. Orchestrated transcription of key pathways in Arabidopsis by the circadian clock. Science **290**, 2110–2113.

Harris FS, Martin CE. 1991. Plasticity in the degree of CAM-cycling and its relationship to drought stress in five species of *Talinum (Portulacaceae).* Oecologia **86**, 575–584.

Hartwell J. 2005. The co-ordination of central plant metabolism by the circadian clock. Biochemical Society Transactions **33**, 945–948.

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. PEP carboxylase kinase is a novel protein kinase controlled at the level of expression. New Phytologist **151**, 91–97.

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. Plant Journal **10**, 1071–1078.

Häusler RE, Baur B, Scharte J, Teichmann T, Eicks M, Fischer KL, Flügge UI, Schubert S, Weber A, Fischer K. 2000. Plastidic metabolite transporters and their physiological functions in the inducible crassulacean acid metabolism plant *Mesembryanthemum crystallinum*. Plant Journal **24**, 285–296.

Heyduk K, Hwang M, Albert V, Silvera K, Lan T, Farr K, Chang T, Chan M, Winter K, Leebens-mack J. 2019*a*. Altered Gene Regulatory Networks Are Associated With the Transition From C3 to Crassulacean Acid Metabolism in *Erycina* (*Oncidiinae: Orchidaceae*). Frontiers in Plant Science **9**, 2000.

Heyduk K, Mckain MR, Lalani F, Leebens-mack J. 2016. Molecular Phylogenetics and Evolution Evolution of a CAM anatomy predates the origins of Crassulacean acid metabolism in the *Agavoideae* (*Asparagaceae*). Molecular Phylogenetics and Evolution **105**, 102–113.

Heyduk K, Moreno-Villena JJ, Gilman IS, Christin P-A, Edwards EJ. 2019*b*. The genetics of convergent evolution: insights from plant photosynthesis. Nature Reviews Genetics **20**, 485–493.

Heyduk K, Ray JN, Ayyampalayam S, Leebens-Mack J. 2018. Shifts in gene expression profiles are associated with weak and strong Crassulacean acid metabolism. American Journal of Botany **105**, 587–601.

Hibberd JM, Covshoff S. 2010. The Regulation of Gene Expression Required for C_4 Photosynthesis. Annual Review of Plant Biology **61**, 181–207.

Holtum JAM, Hancock LP, Edwards EJ, Crisp MD, Crayn DM, Sage R, Winter K. 2016. Australia lacks stem succulents but is it depauperate in plants with crassulacean acid metabolism (CAM)? Current Opinion in Plant Biology **31**, 109–117.

Holtum JAM, Hancock LP, Edwards EJ, Winter K. 2017*a*. Optional use of CAM photosynthesis in two C_4 species, *Portulaca cyclophylla* and *Portulaca digyna*. Journal of Plant Physiology **214**, 91–96.

Holtum JAM, Hancock LP, Edwards EJ, Winter K. 2017*b*. Facultative CAM photosynthesis (crassulacean acid metabolism) in four species of *Calandrinia*, ephemeral succulents of arid Australia. Photosynthesis Research **134**, 17–25.

Holtum JAM, Smith JAC, Neuhaus HE. 2005. Intracellular transport and pathways of carbon flow in plants with crassulacean acid metabolism. Functional Plant Biology **32**, 429.

Holtum JAM, Winter K. 1999. Degrees of crassulacean acid metabolism in tropical epiphytic and lithophytic ferns. Australian Journal of Plant Physiology **26**, 749–757.

Horn JW, Xi Z, Riina R, Peirson JA, Yang Y, Dorsey BL, Berry PE, Davis CC, Wurdack KJ. 2014. Evolutionary bursts in *Euphorbia (Euphorbiaceae)* are linked with photosynthetic pathway. Evolution **68**, 3485–3504.

Hurth MA, Su JS, Kretzschmar T, Geis T, Bregante M, Gambale F, Martinoia E, Neuhaus HE. 2005. Impaired pH homeostasis in Arabidopsis lacking the vacuolar dicarboxylate transporter and analysis of carboxylic acid transport across the tonoplast. Plant Physiology **137**, 901–910.

Jiao JA, Chollet R. 1991. Posttranslational regulation of phosphoenolpyruvate carboxylase in C4 and Crassulacean Acid Metabolism plants. Plant Physiology **95**, 981–5.

Jones RJ, Mansfield TA. 1970. Suppression of stomatal opening in leaves treated with abscisic acid. Journal of Experimental Botany **21**, 714–719.

Kebeish R, Niessen M, Oksaksin M, Blume C, Peterhaensel C. 2012. Constitutive and Dark-Induced Expression of *Solanum tuberosum* Phosphoenolpyruvate Carboxylase Enhances Stomatal Opening and Photosynthetic Performance of Arabidopsis thaliana. **109**, 536–544.

Keeley JE. 1998. CAM Photosynthesis in Submerged Aquatic Plants. The Botanical Review **64**, 273–289.

Kinoshita T, Shimazaki KI. 1999. Blue light activates the plasma membrane H^+-ATPase by phosphorylation of the C-terminus in stomatal guard cells. EMBO Journal **18**, 5548–5558.

Kondo A, Nose A, Ueno O. 2001. Coordinated accumulation of the chloroplastic and cytosolic pyruvate, Pi dikinases with enhanced expression of CAM in *Kalanchoë blossfeldiana*. Physiologia Plantarum **111**, 116–122.

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

Kore-eda S, Noake C, Ohishi M, Ohnishi J, Cushman JC. 2005. Transcriptional profiles of organellar metabolite transporters during induction of crassulacean acid metabolism in *Mesembryanthemum crystallinum*. Functional Plant Biology **32**, 451–466.

Kovermann P, Meyer S, Hörtensteiner S, Picco C, Scholz-Starke J, Ravera S, Lee Y, Martinoia E. 2007. The Arabidopsis vacuolar malate channel is a member of the ALMT family. Plant Journal **52**, 1169–1180.

Kulik A, Wawer I, Krzywińska E, Bucholc M, Dobrowolska G. 2011. SnRK2 Protein Kinases—Key Regulators of Plant Response to Abiotic Stresses. OMICS: A Journal of Integrative Biology **15**, 859–872.

Li B, Chollet R. 1994. Salt Induction and the Partial Purification/Characterization of Phosphoenolpyruvate Carboxylase Protein-Serine Kinase from an Inducible Crassulacean-Acid-Metabolism (CAM) Plant, *Mesembryanthemum crystallinum* L. Archives of Biochemistry and Biophysics **314**, 247–254.

Lim SD, Lee S, Choi W, 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**, 101.

Lim SD, Yim WC, Liu D, Hu R, Yang X, Cushman JC. 2018. A *Vitis vinifera* basic helix–loop–helix transcription factor enhances plant cell size, vegetative biomass and reproductive yield. Plant Biotechnology Journal **16**, 1595–1615.

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

Liu D, Palla KJ, Hu R, *et al.* 2018. Perspectives on the basic and applied aspects of crassulacean acid metabolism (CAM) research. Plant Science **274**, 394–401.

Liu X, Yue Y, Li W, Ma L. 2007. A G Protein-Coupled Receptor Is a Plasma Membrane Receptor for the Plant Hormone Abscisic Acid. Science **318**, 17–19.

Lüttge U. 2002. CO_2-concentrating: consequences in crassulacean acid metabolism. Journal of Experimental Botany **53**, 2131–2142.

Lüttge U, Smith JAC. 1984. Mechanism of passive malic-acid efflux from vacuoles of the CAM plant *Kalanchoë daigremontiana*. The Journal of Membrane Biology **81**, 149–158.

Lüttge U, Smith JAC, Osmond CB. 1981. Energetics of malate accumulation in the vacuoles of *Kalanchoë tubiflora* cells. FEBS Letter **126**, 81–84.

Malagnoux M. 2007. *Arid Land Forests of the World - Global Environmental Perspectives*. Rome (Italy).

Males J, Griffiths H. 2017. Stomatal biology of CAM plants. Plant Physiology **174,** 550–560.

Mallona I, Egea-Cortines M, Weiss J. 2011. Conserved and Divergent Rhythms of Crassulacean Acid Metabolism-Related and Core Clock Gene Expression in the *Cactus Opuntia ficus-indica*. Plant Physiology **156**, 1978–1989.

Mason PM, Glover K, Smith JAC, Willis KJ, Woods J, Thompson IP. 2015. The potential of CAM crops as a globally significant bioenergy resource: Moving from 'fuel or food' to 'fuel and more food'. Energy and Environmental Science **8**, 2320–2329.

Maurino VG, Weber APM. 2013. Engineering photosynthesis in plants and synthetic microorganisms. Journal of Experimental Botany **64**, 743–751.

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 *Kalanchoë daigremontiana*. Plant Physiology **121**, 849–856.

Maxwell K, Von Caemmerer S, Evans JR. 1997. Is a low internal conductance to CO_2 diffusion a consequence of succulence in plants with crassulacean acid metabolism? Australian Journal of Plant Physiology **24**, 777–786.

Menkens AE, Schindler U, Cashmore AR. 1995. The G-box: a ubiquitous regulatory DNA element in plants bound by the GBF family of bZIP proteins. Trends in Biochemical Sciences **20**, 506–510.

Michalowski CB, Olson SW, Piepenbrock M, Schmitt JM, Bohnert HJ. 1989. Time course of mRNA induction elicited by salt stress in the common ice plant (*Mesembryanthemum crystallinum*). Plant Physiology **89**, 811.

Mihailescu R. 2015. Gene expression regulation: Lessons from noncoding RNAs. RNA **21**, 695–696.

Minardi BD, Voytena APL, Santos M, Randi ÁM. 2014. Water stress and abscisic acid treatments induce the CAM pathway in the epiphytic fern *Vittaria lineata* (L.) Smith. Photosynthetica **52**, 404–412.

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

Ming R, Wai CM, Guyot R. 2016. Pineapple Genome: A Reference for Monocots and CAM Photosynthesis. Trends in Genetics **32**, 690–696.

Mittelheuser CJ, van Steveninck RFM. 1971. Rapid action of abscisic acid on photosynthesis and stomatal resistance. Planta **97**, 83–86.

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 C_3 Photosynthesis Plant Arabidopsis thaliana. Frontiers in Plant Science **9**.

Nelson EA, Sage RF. 2008. Functional constraints of CAM leaf anatomy: Tight cell packing is associated with increased CAM function across a gradient of CAM expression. Journal of Experimental Botany **59**, 1841–1850.

Nemhauser JL, Hong F, Chory J. 2006. Different Plant Hormones Regulate Similar Processes through Largely Nonoverlapping Transcriptional Responses. Cell **126**, 467–475.

Neuhaus HE, Holtum JAM, Latzko E. 1988. Transport of Phosphoenolpyruvate by Chloroplasts from *Mesembryanthemum crystallinum* L. Exhibiting Crassulacean Acid Metabolism. plant Physiology **87**, 64–68.

Neuhaus HE, Schulte N. 1996. Starch degradation in chloroplasts isolated from C_3 or CAM (crassulacean acid metabolism)-induced *Mesembryanthemum crystallinum* L. The Biochemical Journal **318**, 945–953.

Nicolas P, Lecourieux D, Gomès E, Delrot S, Lecourieux F. 2013. The grape berry-specific basic helix – loop – helix transcription factor VvCEB1 affects cell size. **64**, 991–1003.

Niechayev NA, Pereira PN, Cushman JC. 2019. Understanding trait diversity associated with crassulacean acid metabolism (CAM). Current Opinion in Plant Biology **49**, 74–85.

Nimmo GA, Nimmo HG, Fewson CA, Wilkins MB. 1984. Diurnal changes in the properties of phosphoenolpyruvate carboxylase in *Bryophyllum* leaves: a possible covalent modification. FEBS Letter **178**, 199–203.

Nimmo GA, Wilkins MB, Fewson CA, Nimmo HG. 1987. Persistent circadian rhythms in the phosphorylation state of phosphoenolpyruvate carboxylase from *Bryophyllum fedtschenkoi* leaves and in its sensitivity to inhibition by malate. Planta **170**, 408–415.

Ogawa T, Inshikawa H, Shimada K, Shibata K. 1978. Synergic Action of Red and Blue Light and Action Spectra for Malate Formation in Guard Cells of *Vicia faba* L. Planta **142**, 61–65.

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

Park S, Fung P, Nishimura N, *et al.* 2009. Abscisic acid inhibits type 2C protein phosphatases via the PYR/PYL family of START proteins. Science **324**, 1068–1071.

Pereira PN, Gaspar M, Smith JAC, Mercier H. 2018. Ammonium intensifies CAM photosynthesis and counteracts drought effects by increasing malate transport and antioxidant capacity in *Guzmania monostachia*. Journal of Experimental Botany **69**, 1993–2003.

Portis Jr AR. 1990. Rubisco activase. Biochimica et Biophysica Acta **1015**, 15–28.

Schulz A, Beyhl D, Marten I, Wormit A, Neuhaus E, Poschet G, Büttner M, Schneider S, Sauer N, Hedrich R. 2011. Proton-driven sucrose symport and antiport are provided by the vacuolar transporters SUC4 and TMT1/2. Plant Journal **68**, 129–136.

Shen Y-Y, Wang X-F, Wu F-Q, *et al.* 2006. The Mg-chelatase H subunit is an abscisic acid receptor. Nature **443**, 823–826.

Shimazaki K, Doi M, Assmann SM, Kinoshita T. 2007. Light Regulation of Stomatal Movement. Annual Review of Plant Biology **58**, 219–247.

Shinozaki K, Yamaguchi-Shinozaki K. 1997. Gene expression and signal transduction in water-stress response. Plant Physiology **115**, 327–334.

Silvera K, Neubig KM, Whitten WM, Williams NH, Winter K, Cushman JC. 2010. Evolution along the crassulacean acid metabolism continuum. Functional Plant Biology **37**, 995–1010.

Silvera K, Santiago LS, Cushman JC, Winter K. 2009. Crassulacean acid metabolism and epiphytism linked to adaptive radiations in the Orchidaceae. Plant Physiology **149**, 1838–1847.

Silvera K, Winter K, Rodriguez BL, Albion RL, Cushman JC. 2014. Multiple isoforms of phosphoenolpyruvate carboxylase in the *Orchidaceae* (subtribe *Oncidiinae*): implications for the evolution of crassulacean acid metabolism. Journal of Experimental Botany **65**, 3623–3636.

Slack CR. 1968. The photoactivation of a phosphopyruvate synthase in leaves of *Amaranthus palmeri*. Biochemical and Biophysical Research Communications **30**, 483–488.

Smith SM, Fulton DC, Chia T, Thorneycroft D, Chapple A, Dunstan H, Hylton C, Zeeman SC, Smith AM. 2004. Diurnal Changes in the Transcriptome Encoding Enzymes of Starch Metabolism Provide Evidence for Both Transcriptional and Posttranscriptional Regulation of Starch Metabolism in Arabidopsis Leaves. Plant Physiology **136**, 2687–2699.

Song L, Huang SC, Wise A, Castanon R, Nery JR, Chen H, Watanabe M, Thomas J, Bar-Joseph Z, Ecker JR. 2016. A transcription factor hierarchy defines an environmental stress response network.

Science **354**, aag15510-1.

Sweetlove LJ, Nielsen J, Fernie AR. 2017. Engineering central metabolism – a grand challenge for plant biologists. Plant Journal **90**, 749–763.

Takahashi-Terada A, Kotera M, Ohshima K, Furumoto T, Matsumura H, Kai Y, Izui K. 2005. Maize phosphoenolpyruvate carboxylase: Mutations at the putative binding site for glucose-6-phosphate caused desensitization and abolished responsiveness to regulatory phosphorylation. Journal of Biological Chemistry **280**, 11798–11806.

Tallman G, Zhu J, Mawson BT, Amodeo G, Nouhi Z, Levy K, 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 and Cell Physiology **38**, 236–242.

Taybi T, Cushman JC. 2002. Abscisic acid signaling and protein synthesis requirements for phosphoenolpyruvate carboxylase transcript induction in the common ice plant. Journal of Plant Physiology **159**, 1235–1243.

Taybi T, Cushman JC, Borland AM. 2002. Environmental, hormonal and circadian regulation of crassulacean acid metabolism expression. Functional Plant Biology **29**, 669–678.

Taybi T, Cushman JC, Borland AM. 2017. Leaf carbohydrates influence transcriptional and post-transcriptional regulation of nocturnal carboxylation and starch degradation in the facultative CAM plant, *Mesembryanthemum crystallinum*. Journal of Plant Physiology **218**, 144–154.

Taybi T, Nimmo HG, Borland AM. 2004. Expression of Phosphoenolpyruvate Carboxylase and Phosphoenolpyruvate Carboxylase Kinase Genes. Implications for Genotypic Capacity and Phenotypic Plasticity in the Expression of Crassulacean Acid Metabolism. Plant Physiology **135**, 587–598.

Taybi T, Patil S, Chollet R, Cushman JC. 2000. A minimal serine/threonine protein kinase circadianly regulates phosphoenolpyruvate carboxylase activity in crassulacean acid metabolism-induced leaves of the common ice plant. Plant Physiology **123**, 1471–82.

Thomas JC, McElwain EF, Bohnert HJ. 1992. Convergent Induction of Osmotic Stress-Responses. Abscisic Acid, Cytokinin, and the Effects of NaCl. Plant Physiology **100**, 416–23.

Trenberth KE. 2011. Changes in precipitation with climate change. Climate Research **47**, 123–138.

Tsiantis MS, Bartholomew DM, Smith JAC. 1996. Salt regulation of transcript levels for the *c* subunit of a leaf vacuolar H^+-ATPase in the halophyte *Mesembryanthemum crystallinum*. Plant Journal **9**, 729–736.

Tsukagoshi H, Suzuki T, Nishikawa K, Agarie S, Ishiguro S, Higashiyama T. 2015. RNA-Seq analysis of the response of the halophyte, *Mesembryanthemum crystallinum* (ice plant) to high salinity. PLoS ONE **10**, 1–16.

Tsuzuki M, Miyachi S, Winter K, Edwards GE. 1982. Localization of Carbonic Anhydrase in Crassulacean Acid Metabolism Plants. Plant Science Letters **24**, 211–218.

Uno Y, Furihata T, Abe H, Yoshida R, Shinozaki K, Yamaguchi-Shinozaki K. 2000. Arabidopsis basic leucine zipper transcription factors involved in an abscisic acid-dependent signal transduction pathway under drought and high-salinity conditions. Proceedings of the National Academy of Sciences of the United States of America **97**, 11632–11637.

Valluru R, Van Den Ende W. 2008. Plant fructans in stress environments: Emerging concepts and future prospects. Journal of Experimental Botany **59**, 2905–2916.

Vovides AP, Etherington JR, Dresser PQ, Groenhof A, Ramirez CIJF. 2002. CAM-cycling in the cycad *Dioon edule* Lindl. in its natural tropical deciduous forest habitat in central Veracruz, Mexico. Botanical Journal of the Linnean Society **138**, 155–162.

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

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*. PLoS Genetics **15**, e1008209.

Wang P, Xue L, Batelli G, Lee S, Hou Y-J, Van Oosten MJ, Zhang H, Tao WA, Zhu J-K. 2013. Quantitative phosphoproteomics identifies SnRK2 protein kinase substrates and reveals the effectors of abscisic acid action. Proceedings of the National Academy of Sciences **110**, 11205–11210.

Warren DM, Wilkins MB. 1961. An endogenous rhythm in the rate of carbon dioxide output of *Bryophyllum fedtschenkoi*. Nature **191**, 686–688.

West-Eberhard MJ, Smith JAC, Winter K. 2011. Photosynthesis, Reorganized. Science **332**, 311–312.

Wilkins MB. 1992. Tansley Review No. 37 Circadian rhythms: their origin and control. New Phytologist **121**, 347–375.

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

Winter K, Foster JG, Edwards GE, Holtum JAM. 1982. Intracellular Localization of Enzymes of Carbon Metabolism in *Mesembryanthemum crystallinum* Exhibiting C_3 Photosynthetic Characteristics or Performing Crassulacean Acid Metabolism. Plant Physiology **69**, 300–307.

Winter K, Holtum JAM. 2014. Facultative crassulacean acid metabolism (CAM) plants: Powerful tools for unravelling the functional elements of CAM photosynthesis. Journal of Experimental Botany **65**, 3425–3441.

Winter K, Holtum JAM. 2017. Facultative crassulacean acid metabolism (CAM) in four small C_3 and C_4 leaf-succulents. Australian Journal of Botany **65**, 103–108.

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

Wormit A, Trentmann O, Feifer I, Lohr C, Tjaden J, Meyer S, Schmidt U, Martinoia E, Neuhaus HE. 2006. Molecular identification and physiological characterization of a novel monosaccharide transporter from Arabidopsis involved in vacuolar sugar transport. Plant Cell **18**, 3476–3490.

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, *et al.* 2017. The *Kalanchoë* genome provides insights into convergent evolution and building blocks of crassulacean acid metabolism. Nature Communications **8**, 1899.

Yang X, Liu D, Tschaplinski TJ, Tuskan GA. 2019. Comparative genomics can provide new insights into the evolutionary mechanisms and gene function in CAM plants. Journal of Experimental Botany **70**, 6539–6547.

Yin H, Guo H-B, Weston DJ, *et al.* 2018. Diel rewiring and positive selection of ancient plant proteins enabled evolution of CAM photosynthesis in *Agave*. BMC Genomics **19**, 588.

Yoshida T, Fujita Y, Sayama H, Kidokoro S, Maruyama K, Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K. 2010. AREB1, AREB2, and ABF3 are master transcription factors that cooperatively regulate ABRE-dependent ABA signaling involved in drought stress tolerance and require ABA for full activation. Plant Journal **61**, 672–685.

Zeiger E. 1990. Light perception in guard cells. Plant, Cell & Environment **13**, 739–744.

Zhang GQ, Xu Q, Bian C, *et al.* 2016. The *Dendrobium catenatum* Lindl. genome sequence provides insights into polysaccharide synthase, floral development and adaptive evolution. Scientific Reports **6**, 1–10.

Zhu K, Liu H, Chen X, Cheng Q, Cheng ZMM. 2018. The kinome of pineapple: Catalog and insights into functions in crassulacean acid metabolism plants. BMC Plant Biology **18**, 1–16.

Zou L-H, Wan X, Deng H, Zheng B-Q, Li B-J, Wang Y. 2018. RNA-seq transcriptomic profiling of crassulacean acid metabolism pathway in *Dendrobium catenatum*. Scientific Data **5**, 180252.

Abbreviations

ABA	abscisic acid
CAM	crassulacean acid metabolism
BL	blue light
CBBC	Calvin-Benson-Bassham cycle
G-6-P	glucose-6-phosphate
ME	malic enzyme
OAA	oxaloacetate
PEPC	phosphoenolpyruvate carboxylase (enzyme)
PPC	phosphoenolpyruvate carboxylase (gene)
PPCK	phosphoenolpyruvate carboxylase kinase
PPDK	pyruvate, phosphate dikinase
RuBisCO	ribulose-1,5-bisphosphate carboxylase/oxygenase
TF	transcription factor
TFBS	transcription factor binding site(s)
WUE	water-use efficiency

Aims of the thesis

Talinum triangulare (Caryophyllales), a facultative CAM species, was proposed as an experimental model to study the CAM pathway and its induction. Based on previous work, it was hypothesized that the phytohormone abscisic acid (ABA) is not only a mediator of stress signalling but gained a new function specifically in facultative CAM species, acting as a signal towards CAM induction (Brilhaus *et al.*, 2016). However, whether ABA is sufficient as the sole signal for CAM induction remained unknown. The point of divergence between ABA-dependent stress signalling and signalling towards CAM induction needs to be determined as well. The present doctoral thesis aimed at addressing these points by:

1) Testing the effect of exogenously applied ABA on CAM induction. The pace of CAM induction and changes to both transcriptome and metabolite levels were of interest.

2) Generating a reference transcriptome of *T. triangulare*. A representative transcript collection is desirable to avoid cross-species mapping of RNA-sequencing reads, to characterize the diversity of transcript variants and to distinguish them from each other unambiguously. It can also assist genome annotation.

3) Generating a draft genome assembly of *T. triangulare*. Availability of a reference genome is one of the prerequisites to establish *T. triangulare* as a model species for the field of CAM research. A genome reference will enable a better characterisation of *T. triangulare* (*e.g.* gene copy number and regulatory elements). Together with transcriptional data, it will be a powerful tool to identify molecular components required for CAM induction in this species. When used in comparative analyses, it will enable identification of conserved changes in protein-coding regions and shared regulatory sequences, thus contributing new pieces of knowledge about CAM regulation and evolution.

Manuscript I

Maleckova, E., Brilhaus, D., Wrobel, T. J., & 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*. *Journal of Experimental Botany*, *70*(22), 6581–6596. https://doi.org/10.1093/jxb/erz189

Own contribution: Experiment design and conductance, RNA extraction, data analysis and graphical presentation except for non-negative matrix factorization (T. J. W.), manuscript preparation

Journal of Experimental Botany
doi:10.1093/jxb/erz189 Advance Access Publication 24 April 2019
This paper is available online free of all access charges (see https://academic.oup.com/jxb/pages/openaccess for further details)

RESEARCH PAPER

Transcript and metabolite changes during the early phase of abscisic acid-mediated induction of crassulacean acid metabolism in *Talinum triangulare*

Eva Maleckova, Dominik Brilhaus, Thomas J. Wrobel, and Andreas P. M. Weber*,

Institute of Plant Biochemistry, Cluster of Excellence on Plant Sciences (CEPLAS), Heinrich-Heine-University, Düsseldorf, Germany

* Correspondence: andreas.weber@hhu.de

Received 17 February 2019; Editorial decision 3 April 2019; Accepted 4 April 2019

Editor: John Cushman, University of Nevada, USA

Abstract

Crassulacean acid metabolism (CAM) has evolved as a water-saving strategy, and its engineering into crops offers an opportunity to improve their water use efficiency. This requires a comprehensive understanding of the regulation of the CAM pathway. Here, we use the facultative CAM species *Talinum triangulare* as a model in which CAM can be induced rapidly by exogenous abscisic acid. RNA sequencing and metabolite measurements were employed to analyse the changes underlying CAM induction and identify potential CAM regulators. Non-negative matrix factorization followed by *k*-means clustering identified an early CAM-specific cluster and a late one, which was specific for the early light phase. Enrichment analysis revealed abscisic acid metabolism, WRKY-regulated transcription, sugar and nutrient transport, and protein degradation in these clusters. Activation of the CAM pathway was supported by up-regulation of phosphoenolpyruvate carboxylase, cytosolic and chloroplastic malic enzymes, and several transport proteins, as well as by increased end-of-night titratable acidity and malate accumulation. The transcription factors *HSFA2*, *NF-YA9*, and *JMJ27* were identified as candidate regulators of CAM induction. With this study we promote the model species *T. triangulare*, in which CAM can be induced in a controlled way, enabling further deciphering of CAM regulation.

Keywords: Abscisic acid, crassulacean acid metabolism, metabolome, *Talinum triangulare*, time course, transcriptome.

Introduction

Crassulacean acid metabolism (CAM) has independently evolved as a carbon-concentrating and water-saving strategy in approximately 6% of vascular plants (Winter *et al.*, 2008; Silvera *et al.*, 2010). CAM species rely on the temporal separation of primary carbon assimilation and its subsequent incorporation into carbohydrates through the Calvin–Benson–Bassham cycle (CBBC). Stomata open towards the end of the light period, enabling CO_2 uptake and its assimilation in the form of HCO_3^- by the concerted actions of phospohoenolpyruvate carboxylase (PPC) and malate dehydrogenase (MDH). Malate is the main organic acid stored in the vacuole, but accumulation of citrate and isocitrate has also been observed (Lüttge, 1988; Medina *et al.*, 1993; Herppich *et al.*, 1995; Chen *et al.*, 2002; Winter and Holtum, 2014). During the light period, stomata are closed and, depending on the type of CAM species, one of three decarboxylating enzymes releases nocturnally fixed CO_2 from malic acid, thus providing Rubisco with its substrate (Holtum *et al.*, 2005). The diurnal cycling of acidification and organic

Abbreviations:ABA, abscisic acid; CA, carbonic anhydrase; CAM, crassulacean acid metabolism; CBBC, Calvin–Benson–Bassham cycle; DEG, differentially expressed gene; ME, malic enzyme; NMF, non-negative matrix factorization; PEP, phosphoenolpyruvate; TCA, tricarboxylic acid; TF, transcription factor.

acid degradation is tightly interconnected with carbohydrate metabolism. Soluble sugars (fructose, glucose, or sucrose) or polysaccharides [starch, (poly)fructans] are degraded in the dark to provide carbon skeletons in the form of PEP for nocturnal carboxylation. The sugar/polysaccharide pool is then regenerated in the light via gluconeogenesis (Borland and Taybi, 2004; Holtum *et al.*, 2005; Borland *et al.*, 2016; Taybi *et al.*, 2017). To mitigate futile cycling, temporal control of enzyme activities is essential and occurs at both transcriptional and post-transcriptional levels, including diurnally regulated expression of *PHOSPHOENOLPYRUVATE CARBOXYLASE* (*PPC*) and *PHOSPHOENOLPYRUVATE CARBOXYLASE KINASE* (*PPCK*) (Nimmo *et al.*, 1999; Cushman *et al.*, 2008).

Despite the strict temporal regulation, CAM remains a highly plastic adaptation that has been defined in different modes, including obligate and facultative CAM, CAM-idling, and CAM-cycling (Lüttge, 2004). In particular, it has been proposed that facultative CAM could serve as a model to allow the identification of the minimal gene set required for a fully functional CAM pathway and to shed light on the regulation of these genes, thus allowing CAM engineering into C_3 crops (Borland *et al.*, 2014; Hartwell *et al.*, 2016). In facultative CAM species, CAM is induced in response to environmental stresses (e.g. drought) and abscisic acid (ABA) has been identified as an important signal in CAM induction (Chu *et al.*, 1990; Dai *et al.*, 1994; Taybi and Cushman, 2002). Characteristics of CAM, such as nocturnal acidification, accumulation of *PPC* transcript, and increased extractable activities of central CAM enzyme, have been induced by exogenous ABA in *Portulacaria afra* (Ting, 1981), *Mesembryanthemum crystallinum* (Holtum and Winter, 1982; Chu *et al.*, 1990; Dai *et al.*, 1994), or *Kalanchoë blossfeldiana* (Taybi *et al.*, 1995). In *Talinum triangulare*, increased amounts of transcripts encoding proteins of both CAM and ABA signalling pathways were detected in response to water withdrawal, returning to non-stress levels upon rewatering (Brilhaus *et al.*, 2016).

In the plant, ABA is perceived by soluble receptors of the PYRABACTIN RESISTANCE/PYRABACTIN RESISTANCE-LIKE/REGULATORY COMPONENTS OF ABA RECEPTOR group, which upon ABA binding physically interact with clade A PP2C phosphatases (PP2CA), preventing dephosphorylation of SNF1-RELATED PROTEIN KINASEs 2 (SnRK2) (Park *et al.*, 2009). Autophosphorylated SnRK2s subsequently modulate the activity of their downstream targets, ranging from other kinases to anion channels and numerous transcription factors (TFs), many of them recognizing ABA-responsive elements in the promoter sequences of target genes (Yoshida *et al.*, 2010; Antoni *et al.*, 2011).

Despite detailed understanding of ABA signalling and accumulating evidence of the ABA inducibility of CAM, a study further elucidating the role of ABA in CAM induction is lacking. We therefore conducted this study with the following aims: (i) to discover how rapidly the CAM pathway can be activated by exogenous ABA; (ii) to determine whether it is possible to distinguish drought-induced and ABA-induced facultative CAM; and (iii) to identify the CAM regulators downstream of ABA signalling.

Materials and methods

Plant material and growth conditions

Talinum triangulare plants used in this study originated from two subsequent, controlled self-pollination events, increasing the homogeneity of the plant material. Seeds were germinated in multiplication substrate (Floraton 3, Floragard) and 4-week-old seedlings were transferred to pots with D 400 soil and Cocopor (Stender). Two weeks before the treatment, the plants were placed into a controlled-environment plant chamber (MobyLux GroBanks, CLF Plant Climatics) with the following growth conditions: 12 h light/12 h dark at 25 °C/23 °C. The light intensity at the leaf level was 150–200 µmol s^{-1} m^{-2}. To avoid unintended induction of CAM by drought, all plants were watered as needed, with the same amount of water per pot.

Treatments and harvest of leaf material

To determine a suitable ABA concentration, a range from 10 to 600 µM was tested (see Supplementary Fig. S1 at *JXB* online) prior to the work presented here. Using 2-month-old plants, two mature and fully illuminated leaves of each plant were treated either with (+/–) ABA [Sigma-Aldrich; 200 µM solution in 0.095% (v/v) methanol] or mock solution (0.095% (v/v) methanol). Both solutions contained 0.02% (v/v) Tween-20 (PanReac AppliChem). The first treatment took place after 4 hours in the light and the second treatment followed 4 hours later (Fig. 1). Treated leaves were harvested and snap-frozen in liquid nitrogen 40, 80, 160, 320, 640, and 1280 min after the first treatment. Two biological replicates originating from two independent plants were harvested per time point. The treated leaves from the same plant were pooled prior to all analyses.

Titratable acidity

Approximately 30 mg of ground leaf material was incubated in 500 µl 50% (v/v) methanol at 90 °C while shaking. After centrifugation at 13,000 *g* for 5 min at room temperature, 450 µl of the supernatant was collected and the remaining pellet was extracted once more. Both fractions were pooled and stored at –20 °C until use. Titratable acidity was measured using bromothymol blue (Carl Roth) as a pH indicator: 60 µl of extract was mixed with 40 µl distilled H_2O and 4 µl bromothymol blue (1 mg/ml stock). Using the microinjector of a Synergy H1 (BioTek) plate reader, samples were titrated with 1 mM KOH in 5 µl steps. After each step, absorbance at 445 nm and 615 nm was recorded, and the ratio of the absorbances at 615/445 nm was calculated. Based on the 615/445 ratio of buffer standards (pH 4.6–7.5), the pH of analysed extracts was determined. From the volume of KOH required to titrate extracts to pH 7.0, the titratable acidity was calculated. Two extractions were performed from each sample, and each extract was titrated in three replicates.

Protein and starch measurements

Approximately 25 mg of ground leaf material was used for extraction in 80% (v/v) ethanol with 10 mM 2-(*N*-morpholino)ethanesulfonic acid (pH 6). Extraction was performed three times with 500 µl extraction buffer while shaking for 1 hour at 90 °C. After each extraction step, samples were centrifuged at 12,000 *g* for 10 min and the supernatant was collected and pooled with previous fractions. From each sample, two extracts were prepared and stored at –20 °C until analysed. The protein concentration was measured using a Pierce BCA Protein Assay Kit (Thermo Fisher Scientific).

Pellets obtained during protein extraction were gelatinized with 300 µl 0.2 M NaOH at 90 °C for 40 minutes. After adjustment to pH 5–6, 200 µl hydrolysis buffer was added [20 mM acetate buffer, pH 4.8; 0.5 U α-amylase (10102814001, Roche), and 4 U amyloglucosidase (A7095, Roche)] and the mixture was incubated overnight at 37 °C. Glucose resulting from starch degradation was measured enzymatically: 10 µl of digest was added to 100 mM HEPES-NaOH (pH 7.5) with 10 mM $MgCl_2$, 1 mM $NADP^+$, 2 mM ATP, 0.1 U glucose-6-phosphate dehydrogenase (G6PDHII-RO, Roche), and 4 U hexokinase (11426362001, Roche).

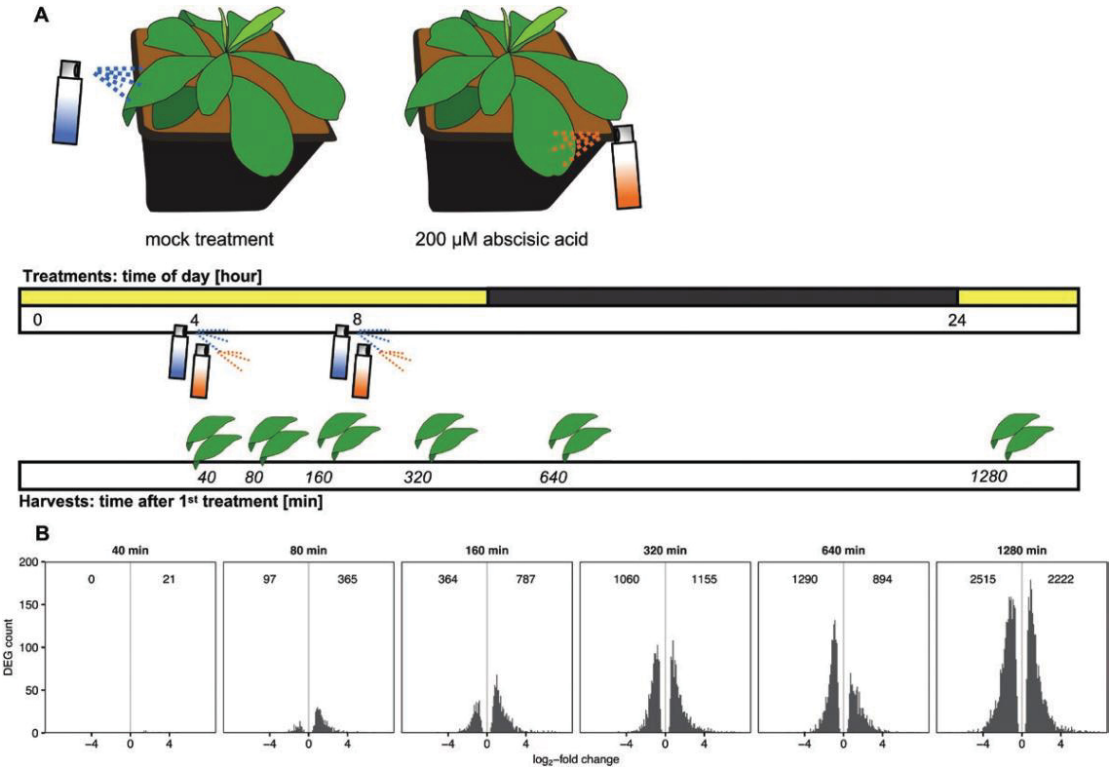


Fig. 1. Early response of *Talinum triangulare* to exogenous abscisic acid. (A) Experimental design. Prior to the treatments, *T. triangulare* plants were adapted to a 12 h/12 h light/dark, 25 °C/23 °C cycle. At 2 months of age, two mature leaves per plant were sprayed with 200 µM ABA or mock solution [0.095% (v/v) methanol]. The first treatment was applied 4 hours into the light period and was followed by a second treatment 4 hours later. Treated leaves were harvested and snap-frozen 40, 80, 160, 320, 640, and 1280 min after the first treatment. (B) Number of significantly ($q \leq 0.01$) down- and up-regulated genes (differentially expressed genes; DEGs) in ABA-treated leaves compared with mock-treated leaves at each time point.

The final reaction volume was 200 µl and the increase of absorbance at 340 nm was monitored.

Metabolite profiling

Approximately 50 mg of ground leaf material was used to obtain extracts for metabolite profiling by gas chromatography–mass spectrometry (GC–MS) as described by Fiehn *et al.* (2000). Data analysis was performed using MassHunter Software, version B.07.00 (Agilent). For relative quantification, all metabolite peak areas were normalized to the peak area of the internal standard ribitol added prior to extraction and the amount of leaf material. From each sample, two extracts were prepared and analysed.

Amino acids were measured by liquid chromatography–mass spectrometry (LC–MS). Extracts were prepared in 80% (v/v) ethanol (Fisher Scientific) as described by Di Martino *et al.* (2003), and data analysis was conducted using MassHunter Software, version B.07.00 (Agilent). From each sample, two extracts were prepared and analysed. Metabolites analysed by both methods are listed in Supplementary Dataset S3.

Total RNA extraction, preparation of Illumina libraries, and sequencing

Total RNA was extracted from ground leaf tissue using the Monarch® Total RNA Miniprep Kit (New England Biolabs Inc.) following the manufacturer's instructions. Contaminating DNA was digested with DNase I (New England Biolabs Inc.). The integrity of input RNA was analysed on a 2100 Bioanalyzer (Agilent). Libraries were prepared using the TruSeq Stranded mRNA Library Prep Kit (Illumina) following the Illumina TruSeq Stranded mRNA Sample Preparation Guide #15031047 Rev. E, with the following adaptations: 300 ng total RNA as the basis for library preparation, adapter index dilution to 50% with RNase-free water, and an additional purification step with AMPure XP (Beckman Coulter).

The concentration of the resulting libraries was estimated using a NanoDrop 8000 (Thermo Fisher Scientific) and quality control was performed with a High Sensitivity NGS Fragment Analysis Kit (1 bp–6000 bp) DNF-474 (Advanced Analytical Technologies, Inc.). Based on quantification with a Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific), libraries were adjusted to 2 nM concentration before cluster generation with a HiSeq 3000/4000 SR Cluster Kit (Illumina). Twelve samples were pooled per lane and sequenced in 150 bp single-end mode on a HiSeq 3000 platform (Illumina). On average, over 32 million reads per library were obtained (Supplementary Table S1).

Read mapping and annotation

Illumina reads were trimmed with Trimmomatic version 0.33 (Bolger *et al.*, 2014), keeping only the reads longer than 36 bases. Subsequently, BLAT (Kent, 2002) in protein space (parameters: -t=dnax -q=dnax) was employed to map the trimmed reads against the reference transcriptome of *Beta vulgaris*, containing 28 721 transcripts in total (RefBeet-1.2.2; Dohm *et al.*, 2014). The best BLAT hits (the lowest *e*-value and the highest bit score) were counted. On average, 70% mapping efficiency (i.e. *T. triangulare* reads matching a *B. vulgaris* target) was achieved, and 18 657 reference transcripts with at least one count were detected (Supplementary Table S1). Reference transcripts were annotated based

on their homology to representative Arabidopsis peptides (Araport11, https://www.araport.org) via BlastX (BLAST 2.6.0+ suite, parameters: -evalue 1e-5 -max_target_seqs 1) (Camacho *et al.* 2009). Identified homologies were also used for assignment of MapMan functional categories (version X4_R1.0; Thimm *et al.*, 2004) and TF families based on PlantTFDB (Jin *et al.*, 2017). Putative protein localization was predicted with TargetP (Emanuelsson *et al.*, 2000) based on *B. vulgaris* reference sequences. Selected transcripts were independently validated by quantitative real-time PCR (Supplementary Protocol S1).

Non-negative matrix factorization

Reads normalized to reads per kilobase million were used for non-negative matrix factorization (NMF), which was performed with altered variants of the Lee and Seung (euclidean distance), the Brunet, and the Kullback–Leibler (KL) algorithms (KL divergence) from the NMF package for R (Gaujoux and Seoighe, 2010). To facilitate interpretation, the coefficient matrix was scaled, so that the sum of all coefficients in a sample would sum to 1 in each iteration. Factorization was carried out in 50 iterations with random seeding and the factorization rank was determined using the elbow within the residual sum of squares and the average regression coefficient (R^2). With all three algorithms, five factors were identified, with the KL and Brunet algorithms having a higher average R^2 than the Lee and Seung method (Supplementary Fig. S4). The final factorization was performed 500 times and the result with the lowest residual sum of squares was used for further analysis. The factorized data were z-scored and clustered via the Hartigan and Wong algorithm for k-means clustering (Hartigan and Wong, 1979). The enrichment of biological functions was analysed using Fisher's exact test on a set of reduced MapMan categories (Supplementary Table S2). All P-values were adjusted for multiple hypothesis testing by Benjamini–Hochberg correction (Benjamini and Hochberg, 1995).

Data analysis

Downstream analyses were performed in R version 3.5.1 (https://www.R-project.org/) and required packages. Raw counts were used for analysis of differential gene expression with the DESeq2 package version 1.22.2 (Love *et al.*, 2014) in the default mode, with Benjamini–Hochberg P-value adjustment and a 0.01 significance threshold. Under- and over-represented biochemical pathways were identified using MapMan categorization and the Wilcoxon rank sum test with false discovery rate adjustment according to Benjamini and Hochberg (1995) and q-value≤0.05 as a significance threshold. Reads per million were used to identify genes with a significantly different temporal pattern between the mock and ABA treatments using the maSigPro package (Nueda *et al.*, 2014) (see Supplementary Protocol S2 for more details).

Accession numbers

The read data as well as processed files are deposited at the National Center for Biotechnology Information Gene Expression Omnibus under accession number GSE116590 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE116590).

Results

Exogenous ABA induces rapid changes at the level of the transcriptome

To assess the immediate response of the facultative CAM species *T. triangulare* to foliar application of ABA, we performed an RNA-seq time-course experiment spanning 40–1280 min after the treatment (Fig. 1A). DESeq analysis of transcript abundances in ABA-treated compared with mock-treated leaves revealed that the number of differentially expressed genes (DEGs) generally increased with time. While the effect of ABA on the number of DEGs was negligible after 40 min, 25% of mapped genes were differentially expressed after 1280 min (Fig. 1B).

For a closer analysis of the observed changes and their biological effect, the following approaches were taken. First, analysis of steady-state transcript abundances was performed, focusing especially on CAM genes and components of the ABA signalling pathway. Overall patterns of transcript abundance of all mapped genes were analysed to identify genes whose temporal patterns changed in response to ABA. Combining both analyses, genes with significantly altered steady-state transcript abundances (q-value ≤0.01) and temporal pattern were identified. NMF combined with k-means clustering was applied to identify genes and biological processes with similar time-dependent co-expression. Finally, we aimed to select candidates relevant for CAM induction by comparing ABA- and drought-induced responses of *T. triangulare* with ABA treatment of the C_3 model plant Arabidopsis.

Altered transcript levels of CAM genes

Out of 23 core CAM genes (Supplementary Table S2; Supplementary Dataset S1), a homologous gene for 20 of them was found in the *B. vulgaris* reference, of which five were up-regulated after 160 min and six after 320 min, (Supplementary Fig. S5A). Transcript levels of *PHOSPHOENOLPYRUVATE CARBOXYLASE 3* (*PPC3*) accumulated 2.1-fold in ABA-treated leaves after 160 min and remained increased at the remaining time points (5.1-fold after 1280 min). Transcripts of the primary PPC isoform, *PPC1*, accumulated likewise (Supplementary Fig. S6), but the overall abundance pattern did not change significantly in response to ABA (maSigPro analysis). *PYRUVATE ORTHOPHOSPHATE DIKINASE* (*PPDK*) transcripts started to accumulate after 80 min (3.6-fold), reaching a 12.8-fold higher level after 640 min. In addition, transcript levels of genes encoding carbonic anhydrases (CAs) were altered after ABA treatment. While *ALPHA CARBONIC ANHYDRASE 1* (*ACA1*) transcript diminished after 160 min (3.5-fold decrease) and remained reduced throughout the remaining sampling points (5.3-fold after 1280 min), transcripts of chloroplastic *BETA CARBONIC ANHYDRASE 5* (*BCA5*) accumulated up to 2.9-fold after 320 min (Fig. 2). To avoid futile cycling, PPC activity is post-translationally regulated by PPCK. In *T. triangulare*, *PPCK1* transcript levels were reduced 2.6-fold in ABA-treated leaves after the dark period (Fig. 2). Transcripts of decarboxylating *MALIC ENZYME*s (*ME*s), chloroplastic *NADP-ME4* and cytosolic *NADP-ME1*, accumulated in response to ABA. *NADP-ME4* was the faster-responding isoform, while *NADP-ME1* transcripts accumulated strongly after 320 min, reaching 2.8-fold up-regulation after 1280 min (Fig. 2).

Flow of metabolites through the CAM pathway requires numerous known and putative transporters. The chloroplastic pyruvate transporter *BILE ACID:SODIUM SYMPORTER FAMILY PROTEIN 2* (*BASS2*) was up-regulated, especially after 640 min (5.2-fold), while transcript levels of malate-transporting *DICARBOXYLATE TRANSPORTER 1* (*DiT1*) declined at 320 and 1280 min (2.1-fold and 2.9-fold, respectively)

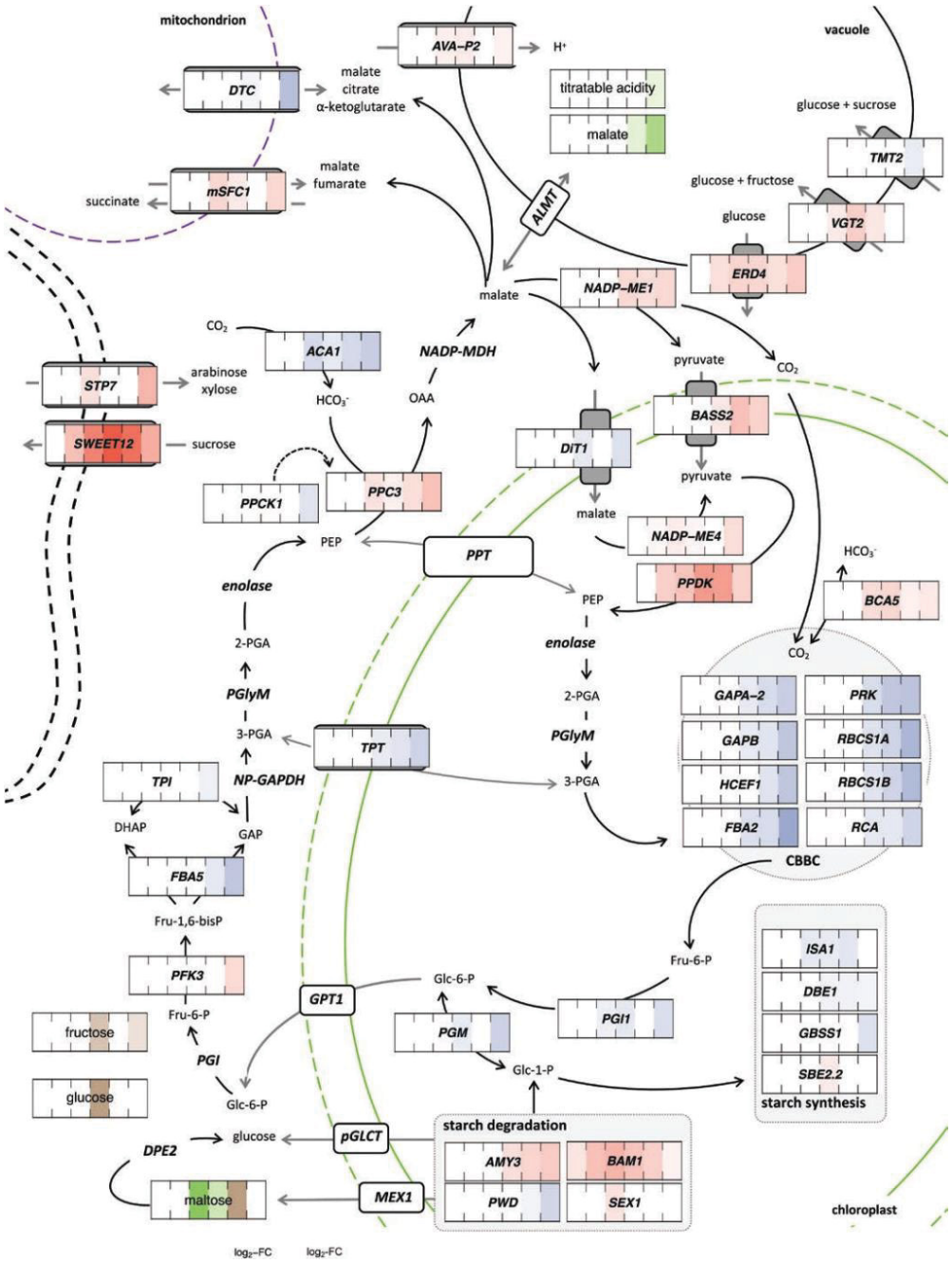


Fig. 2. Exogenous ABA altered the transcript levels of core CAM genes and genes of related pathways, correlating with altered levels of selected metabolites in *Talinum triangulare*. Transcript abundances and metabolite amounts are expressed as $\log_2$-fold changes of ABA-treated compared with mock-treated leaves on blue–red and yellow–green scales, respectively (expression, *n*=2; metabolites, *n*=4). Only genes with significantly different transcript abundances (DESeq with Benjamini–Hochberg correction, *q*-value ≤0.01) and significantly altered temporal patterns (maSigPro with Benjamini–Hochberg correction, *q*-value ≤0.01 and R^2>0.85) are shown. The significance level for metabolites was 0.05 after Benjamini–Hochberg correction. Substrate conversions are depicted with solid lines, post-translation regulations with dashed lines, and transport processes are shown in grey. Genes whose involvement in the depicted pathways is expected are also included. Protein subcellular localization is based on prediction from *Beta vulgaris* (TargetP). 1,3-BPG, 1,3-bisphosphoglycerate; 2-PGA, 2-phosphoglycerate; 3-PGA, 3-phosphoglycerate; CBBC, Calvin–Benson–Bassham cycle; DHAP, dihydroxyacetone phosphate; Fru-6-P, fructose-6-phosphate; Fru-1,6-bisP, fructose-1,6-bisphosphate; GAP, glyceraldehyde-3-phosphate; Glc-1-P, glucose-1-phosphate; Glc-6-P, glucose-6-phosphate; OAA, oxaloacetate; PEP, phosphoenolpyruvate; *PGlyM*, phosphoglycerate mutase.

(Fig. 2). Mitochondrial organic acid transporters were affected by ABA as well: *DICARBOXYLATE/TRICARBOXYLATE TRANSPORTER* (*DTC*) transcripts declined up to 8.1-fold at 1280 min, while *MITOCHONDRIAL SUCCINATE-FUMARATE CARRIER 1* (*mSFC1*) transcripts accumulated at 160, 320, and 1280 min (2.8-, 1.8-, and 2.9-fold, respectively) (Fig. 2).

Carbohydrate metabolism and transport

ABA treatment also affected the levels of transcripts encoding enzymes of starch and sucrose metabolism. Compared with the CAM genes, these responded later (after 640 and 1280 min) and they were generally down-regulated (Supplementary Fig. S5B). Of the genes encoding starch-metabolism enzymes, transcripts of *ISOAMYLASE 1* (*ISA1*) declined 2.2-fold after 160 min, followed by 1.6-fold depletion of *DEBRANCHING ENZYME 1* (*DBE1*) transcripts after 320 min and finally by 2.8-fold depletion of *GRANULE BOUND STARCH SYNTHASE 1* (*GBSS1*) transcripts, the depletion of which was restricted to the early hours of the light period (Fig. 2). In contrast, *BETA-AMYLASE 1* (*BAM1*) transcripts accumulated 2-fold in ABA-treated leaves as soon as after 40 min and remained increased throughout the remaining sampling points. Additionally, *ISOAMYLASE 3* (*AMY3*) transcripts accumulated 2.3-fold after 320 min (Fig. 2).

Multiple genes encoding glycolytic/CBBC enzymes were down-regulated, including at 1280 min (i.e. early in the dark period), among them chloroplastic *PHOSPHOGLYCERATE/BISPHOSPHOGLYCERATE MUTASE* (*PGM*) (4.7-fold), *PHOSPHOGLUCOSE ISOMERASE 1* (*PGI1*) (3.2-fold), *HIGH CYCLIC ELECTRON FLOW 1* (*HCEF1*) (6.7-fold), and *FRUCTOSE-BISPHOSPHATE ALDOLASE 2* (*FBA2*) (3.9-fold) (Fig. 2). Additionally, *TRIOSE-PHOSPHATE/PHOSPHATE TRANSLOCATOR* (*TPT*) transcripts declined 4.5-fold after 320 min and remained at a reduced level also at the last sampling (Fig. 2).

In contrast to the pronounced down-regulation of enzyme-coding transcripts, there was up-regulation of sugar transporters localized in both the plasma membrane and the tonoplast. Transcript accumulation of *SUGAR TRANSPORTER PROTEIN 7* (*STP7*) was found only at time points 160 and 1280 min (2.3- and 6.7-fold, respectively). Transcript levels of *HOMOLOG OF MEDICAGO TRUNCATULA MTN3* (*SWEET12*) accumulated 5.4-fold after 80 min, reaching an 80-fold increase late in the light period (320 min after the treatment) (Fig. 2). Transcripts of *EARLY-RESPONSIVE TO DEHYDRATION 4* (*ERD4*) and *VACUOLAR GLUCOSE TRANSPORTER 2* (*VGT2*) accumulated in ABA-treated leaves 2.4-fold after 80 min and 1.8-fold after 160 min, respectively (Fig. 2).

Changes in metabolite levels in response to exogenous ABA

The above-described changes in transcript levels correlated with altered levels of several metabolites. End-of-night titratable acidity increased in ABA-treated leaves to 52.7±4.1 μmol H^+ g^{-1} FW, representing a 1.6-fold increase compared with mock-treated leaves (Fig. 2). GC-MS analysis of leaf extracts revealed malate and citrate as the primary organic acids that accumulated during the dark period. At the last time point (1280 min), relative amounts of malate and citrate were 4- and 16-fold higher than levels in mock-treated leaves, respectively. Additionally, α-ketoglutarate, succinate, and fumarate accumulated 11.3-, 5.7-, and 3-fold, respectively (Supplementary Dataset S4; Supplementary Fig. S7). Besides organic acids, levels of several carbohydrates were affected. Maltose accumulated transiently at 320 min, followed by a decline at 640 min. At 640 min, glucose was reduced as well (Fig. 2). A reduction of fructose levels occurred specifically after the dark period. A trend towards a time-dependent decline in starch and sucrose levels was observed (Supplementary Fig. S8). Total protein levels also declined up to 1.3-fold in response to ABA (at 1280 min) (Supplementary Fig. S7).

As ABA plays a major role in stress signalling, the levels of stress-related compounds were assessed. γ-aminobutyric acid (GABA), myo-inositol, and raffinose levels declined transiently after 640 min, or after 320 min in the case of myo-inositol (Supplementary Dataset S4). In contrast, levels of amino acids were not affected by ABA treatment (Supplementary Dataset S4).

Exogenous ABA alters ABA signalling

Transcript levels of *REGULATORY COMPONENT OF ABA RECEPTOR 1* (*RCAR1/PYL9*) and *RCAR3/PYL8* showed altered temporal patterns in response to ABA, and their abundance declined 5.9- and 4-fold, respectively, after 1280 min (Fig. 3). In contrast, transcripts encoding downstream components of the ABA signalling pathway were affected more rapidly. Transcripts of the phosphatases *PROTEIN PHOSPHATASE 2CA* (*PP2CA*), *ABA INSENSITIVE 1* (*ABI1*), and *HIGHLY ABA-INDUCED PP2C GEN3* (*HAI3*) accumulated 3.1- 2.9-, and 6.5-fold, respectively, starting at 40 min, and remained up-regulated as time progressed. Transcript levels of *SUCROSE NONFERMENTING 1-RELATED PROTEIN KINASEs 2* (*SnRK2*) were influenced by ABA with similar kinetics, but the temporal patterns of only *SnRK2.5, SnRK2.6/OST1*, and *SnRK2.8* were altered. While *SnRKs 2.6* and *2.8* were up-regulated 4.5- and 2.1-fold, respectively, in response to ABA, there was 1.5-fold down-regulation of *SnRK2.5* after 160 min (Fig. 3).

The temporal patterns of numerous TFs acting downstream were influenced by ABA. Transcripts of *ABSCISIC ACID RESPONSIVE ELEMENT-BINDING FACTOR 2* (*ABF2*), *ABI FIVE BINDING PROTEIN 2* (*AFP2*), *EID1-LIKE 3* (*EDL3*), *HOMEOBOX PROTEIN 7* (*HB7*), and *RESPONSIVE TO DESICCATION* (*RD26*) accumulated 2.5-, 4.8-, 28.6-, 7.7-, and 12.7-fold, respectively, as soon as after 40 min, and remained increased at the remaining sampling times (Fig. 3). Levels of several transcripts encoding ABA synthesis and transport proteins were also affected by the ABA treatment. Transcripts of *ABA DEFICIENT 1* (*ABA1*) increased 1.8-fold after 80 min, but transcripts of *ABSCISIC ALDEHYDE OXIDASE 1* (*AAO1*) accumulated only after 1280 min (1.9-fold) (Fig. 3). Transcript levels of *ATP-BINDING CASETTE*

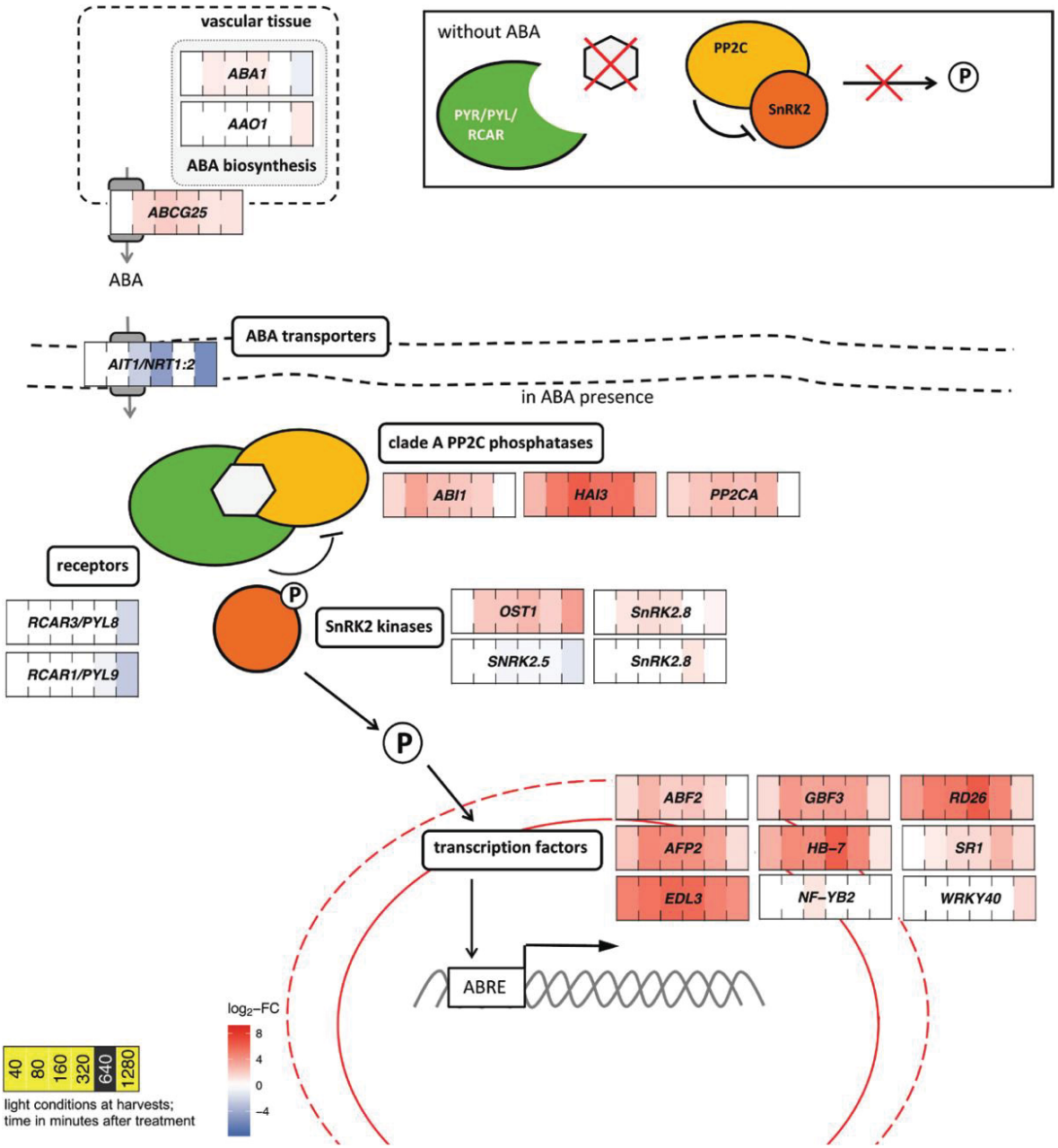


Fig. 3. Transcript levels of genes involved in ABA signalling and biosynthesis were influenced by exogenous ABA in *Talinum triangulare*. Transcript abundances are expressed as $\log_2$-fold changes of ABA-treated compared with mock-treated leaves (n=2). Only genes with significantly different transcript abundances (DESeq with Benjamini–Hochberg correction, q-value≤0.01) and significantly altered temporal patterns (maSigPro with Benjamini–Hochberg correction, q-value≤0.01 and R^2>0.85) are shown. The core signalling pathway comprises PYR/PYL/RCAR receptors, clade A phosphatases PP2C, and SnRK2 kinases. In the absence of ABA, PP2C phosphatases bind SnRK2s and prevent them from phosphorylating downstream targets. Upon binding of ABA to the receptors, they capture PP2Cs, releasing phosphorylated SnRK2s. Targets of SnRK2s include transcription factors in particular, many of which regulate gene expression through binding to ABA-responsive element (ABRE) motifs in promoter sequences of target genes. OST1 is involved in the control of stomatal movement. P, phosphorylation.

G25 (*ABCG25*), an ABA exporter from the site of synthesis, accumulated 2.9-fold starting 80 min after the treatment (Fig. 3). In contrast, *ABA-IMPORTING TRANSPORTER 1* (*AIT1/NRT1:2*) was 5.2-, 25.2-, and 45.9-fold down-regulated after 160, 320, and 1280 min.

Factorization and k-means clustering of temporal transcript patterns

NMF was performed to obtain insights into time-dependent abundance patterns. Based on the transcriptional contributions of the mapped genes, five factors were identified (Fig. 4A). In factor 1, most of the transcriptional contribution occurred at 320 and 640 min in the mock treatment. Similarly, factor 2 was most pronounced at 320 and 640 min, but was specific to the ABA treatment. Factor 3 was characterized by basal transcriptional contribution over time with major expression in the ABA treatment specifically after 1280 min. In factor 4, the contribution declined with time until 640 min, followed by a pronounced contribution only in the mock treatment after 1280 min. The transcriptional contribution in factor 5 changed over time but did not differ between treatments (Fig. 4A).

Based on these factors, subsequent *k*-means clustering identified nine clusters, three of them comprising significant contributions of either of the ABA-specific factors. The early ABA factor contributed to clusters 6 and 7, while the late ABA factor was present in cluster 5 (Fig. 4B). Enrichment analysis of MapMan categories identified ABA metabolism, WRKY TFs, phosphate transport, and ubiquitin-mediated protein degradation as over-represented in the early cluster 6. In the early cluster 7, vesicle transport, protein glycosylation, and reductive-oxidative thioredoxin reactions were over-represented. The late cluster 5 was enriched for sugar and nutrient signalling, ABC transporters, and glutathione *S*-transferases, but also heat stress response, gibberellin metabolism, and raffinose and phenylpropanoid metabolism pathways (Fig. 4B). Besides their specific enrichments, all three clusters shared

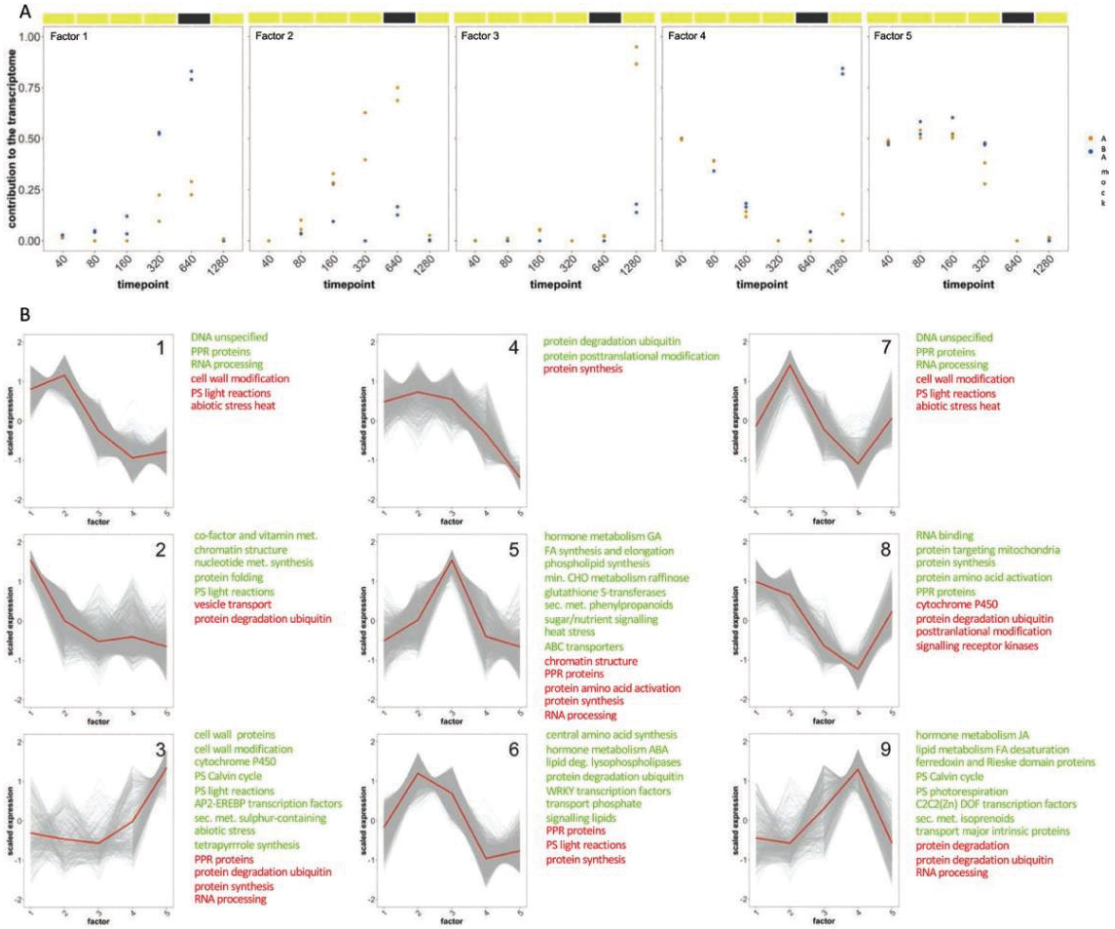


Fig. 4. Non-negative matrix factorization and *k*-means clustering of the *Talinum triangulare* transcriptome. (A) Five factors were identified based on time-dependent expression of individual genes (for details, see Materials and Methods). (B) Based on the transcriptional contribution of each factor, mapped genes were assigned to nine clusters (for details, see Materials and Methods). Enrichment of MapMan categories in each cluster is shown in green (over-represented categories) and red (under-represented categories). deg, Degradation; FA, fatty acids; GA, gibberellic acid; JA, jasmonic acid; met, metabolism; min, CHO, minor carbohydrate; misc, miscellaneous; PPR, pentatricopeptide repeat; PS, photosynthesis; sec. met., secondary metabolism.

under-representation of protein synthesis, and clusters 6 and 7 both had under-representation of photosynthesis.

Due to the large proportion of factor 5 in cluster 3, this cluster could explain time-dependent changes to the transcriptome in both treatments. This cluster was enriched for CBBC, light photosynthetic reactions, and *AP2-EREBP* TFs (Fig. 4B). Over-representation of photosynthetic processes was shared by clusters 1, 2, and 9, which comprised large contributions of mock-treatment-specific factors 1 and 4, as well as early ABA factor 2 in cluster 1 (Fig. 4B).

Candidate regulators of CAM induction

ABA-induced changes in both the transcriptome and the metabolome hinted at activation of the CAM pathway. This prompted the question of which gene products can translate the exogenous ABA signal into CAM induction. To answer this question, we performed a DESeq analysis, comparing a pool of all ABA-treated samples against a pool of mock-treated samples, thus reducing the changes in transcript abundance down to the direction and degree of the changes. In this way, 773 DEGs were obtained (Supplementary Dataset S3).

For further filtering, temporal transcript abundances were considered, limiting the search only to clusters 5, 6, and 7 (i.e. the clusters of major ABA responsiveness; Fig. 4B). Additionally, to be considered a CAM regulator, a gene had to show an altered temporal pattern in the maSigPro analysis (Supplementary Dataset S2). Finally, the transcript abundance of such a gene had to be altered with progressing time during drought induction of CAM and reversible to non-stress levels 2 days after rewatering in the same experiment (Brilhaus *et al.*, 2016). As we searched for genes responsive to ABA specifically in the facultative CAM species *T. triangulare*, we compared the DEGs identified here to ABA-responsive genes in Arabidopsis (Song *et al.*, 2016). Genes responsive in both species were not considered to be CAM regulators. When all these filtering criteria were applied, 32 candidate genes were identified; among them were nine transcription factors, six genes of amino acid metabolism, and five transporters (Table 1). Among the genes differentially expressed between ABA- and mock-treated leaves, there were 36 genes without a known Arabidopsis homolog (Supplementary Dataset S3) but only five of them also showed an altered temporal pattern of transcript abundance (Table 2).

Discussion

In this study, we applied ABA to induce CAM in a controlled manner in our facultative CAM model *T. triangulare*. To the best of our knowledge, this is the first study to describe the earliest changes in both the transcriptome and the metabolome during the induction phase. By following transcript abundances in ABA-treated and mock-treated leaves in parallel, we aimed to reduce the effect of circadian regulation when identifying DEGs. However, it should be noted that most pronounced changes (Fig. 1A) were observed after the dark period, suggesting a combined effect of ABA and darkness on transcript abundances. At the same time, it seems unlikely that these changes were caused by alterations to circadian clock genes. Out of 34 genes involved in circadian regulation (according to MapMan), nine showed a significantly altered temporal pattern of transcript abundances, but only *LATE ELONGATED HYPOCOTYL* (*LHY*) and *ELF4-LIKE 4* (*ELF4-L4*) were assigned to ABA-specific clusters 5 and 7, respectively (Supplementary Table S3), suggesting that (i) the central circadian clock was not affected by the ABA treatment and (ii) the detected and further discussed changes in transcript abundances are a response to ABA or ABA and darkness, but not the diel cycle alone.

ABA induced changes in transcript abundance of CAM core and related genes as well as nocturnal acidification

Amounts of transcripts encoding central CAM enzymes were influenced by exogenous ABA. These transcripts included *PPC1* (Supplementary Fig. S6) and *PPC3* (Fig. 2), in agreement with ABA-induced transcript accumulation (Taybi and Cushman, 1999) and increased measurable PPC activity in *M. crystallinum* (Dai *et al.*, 1994). While *PPC1* was the most abundant isoform (Supplementary Dataset S1), only the temporal pattern of *PPC3* was significantly altered by ABA (Supplemental Dataset S2). In contrast, exogenous ABA induced the depletion of *PPC3* transcripts in Arabidopsis (Song *et al.* 2016), indicating different transcriptional responses to ABA between a C_3 species and a facultative CAM species. It might therefore be of interest to further investigate the contribution of different PPC isoforms to CAM induction. Transcript abundance of *PPCK1*, which post-translationally regulates PPC activity, has been reported to be under circadian control (Nimmo *et al.*, 1999; Taybi *et al.*, 2000; Boxall *et al.*, 2017); here, the amount of *PPCK1* transcript in ABA-treated leaves declined after 1280 min (i.e. shortly after the onset of light).

Assuming a requirement for active CA together with PPC, co-expression of their genes could be expected. This was observed for chloroplastic *BCA5* only, transcripts of which started to accumulate after 160 min (Fig. 2). In a comparison of four CAM species, CA activity was detected in chloroplasts of ME species, while activity in the cytosol was detected in phosphoenolpyruvate carboxykinase species (Tsuzuki *et al.*, 1982). As *T. triangulare* relies on ME (Brilhaus *et al.*, 2016), we hypothesize that *BCA5* is the CAM-specific isoform. In this context, it will be of interest to investigate the role of CAs in the CAM pathway, their localization, and their regulatory mechanism(s). Transcript levels of both cytosolic and chloroplastic isoforms of *NADP-ME* were responsive to exogenous ABA. Transcript accumulation of chloroplastic *NADP-ME4* was restricted to 160 and 1280 min, while transcript levels of cytosolic *NADP-ME1* were increased from 320 to 1280 min (Fig. 2), further indicating adjustment of the CAM enzymatic toolkit.

Increased transcript levels of carboxylating enzymes correlated with the nocturnal increase in titratable acidity and organic acid content (Fig. 2). Nocturnal titratable acidity of *T. triangulare* was previously reported to reach 10–30 μmol H^+ g^{-1} FW with progressing drought (Brilhaus *et al.*, 2016) and 40–60 μmol H^+ g^{-1} FW in salt-induced recycling CAM (Montero *et al.*, 2018). Here, during the first dark period after

Table 1. *Genes identified as possible CAM regulators in* Talinum triangulare

	Transcript ID	AGI	Name	TF family	log_2-FC pool (ABA/mock)	Cluster (NMF)
Transcriptional activators	KMT03072	AT5G19650	*OFP8*		−1.30	2
	KMT14500	AT3G20910	*NF-YA9*	NF-YA	1.38	5
	KMT02260	AT4G14540	*NF-YB3*	NF-YB	0.71	6
	KMT18169	AT3G11090	*LBD21*	LBD	−1.88	2
	KMT07141	AT3G01470	*HB-1*	HD-ZIP	1.02	6
	KMS96841	AT2G26150	*HSFA2*	HSF	2.58	5
	KMS95225	AT1G25440	*BBX15*	CO-like	−1.80	2
	KMS99638	AT1G19850	*MP*	ARF	1.23	6
	KMT11833	AT2G44730		Trihelix	0.41	6
	KMT10399	AT4G00990			0.62	7
	KMT03067	AT3G24520	*HSFC1*	HSF	1.63	7
Amino acid metabolism	KMT00186	AT1G29900	*CARB*		1.06	5
	KMT12470	AT1G64660	*MGL*		2.67	5
	KMT14300	AT1G55510	*BCDH BETA1*		1.50	6
	KMT10213	AT1G03090	*MCCA*		1.57	6
	KMT12470	AT1G64660	*MGL*		2.67	5
	KMS98028	AT1G08630	*THA1*		1.49	5
Solute transport	KMT02985	AT2G26900	*BASS2*		1.05	6
	KMT13799	AT1G08960	*CAX11*		1.02	6
	KMT03912	AT1G30360	*ERD4*		1.60	5
	KMT03911	AT1G30360	*ERD4*		1.50	5
	KMS95179	AT1G19910	*AVA-P2*		0.45	6
Carbohydrate metabolism	KMT07128	AT1G69830	*AMY3*		1.08	6
Chromatin organisation	KMT10399	AT4G00990			0.62	7
Lipid metabolism	KMT19551	AT1G43620	*UGT80B1*		0.48	6
Nucleotide metabolism	KMT00186	AT1G29900	*CARB*		1.06	5
Protein degradation	KMT12139	AT3G61180			1.28	5
Protein modification	KMT01924	AT1G28480	*GRX480*		2.88	5
Secondary metabolism	KMS96359	AT2G20340	*AAS*		4.15	5
	KMT17647	AT2G07050	*CAS1*		1.03	6
Vesicle trafficking	KMT15610	AT2G44140			0.46	6

These genes were identified based on their differential expression between mock treatment and ABA treatment, time-dependent pattern of transcript abundance (NMF analysis), altered temporal pattern of transcript abundance (maSigPro analysis), differential transcript accumulation in drought-induced CAM in *T. triangulare* (Brilhaus *et al.*, 2016), and lack of ABA responsiveness in C_3 Arabidopsis (Song *et al.*, 2016). Log_2-FC pool (ABA/mock) is based on DESeq analysis of all ABA-treated samples against all mock-treated samples, i.e. across all time points. All shown log_2-FC are significant at the 0.01 level. AGI, Arabidopsis Genome Initiative; TF, transcription factor.

Table 2. *ABA-responsive genes in* Talinum triangulare *without a known Arabidopsis homolog*

Transcript ID	log_2-FC pool (ABA/mock)	*q*-value	NMF cluster	Domain	Description
KMT11995	−1.36	4.30E-04	2	IPR005079	Peptidase C45
KMT13506	1.32	3.24E-03	6	n.d.	n.d.
KMT00528	1.51	2.85E-04	6	n.d.	n.d.
KMT09356	2.59	5.72E-03	5	n.d.	n.d.
KMS99845	3.31	2.46E-05	6	n.d.	n.d.

Analysis of differential expression between mock and ABA treatments, considering only genes with altered temporal patterns of transcript abundance (maSigPro analysis). Log_2-FC pool (ABA/mock) is based on DESeq analysis of all ABA-treated samples against all mock-treated samples, i.e. across all time points. Protein domains were determined using InterProScan. n.d., Not determined.

the ABA treatment, end-of-night titratable acidity reached 52.7±4.1 μmol H^+ g^{-1} FW, and the difference in titratable acidity between the early hours of the dark period and the subsequent morning was comparable to that observed by Herrera *et al.* (1991) in sun-exposed leaves after 4 days of water withdrawal. It has been shown in both obligate and facultative CAM species that the major storage acid is malate (Chen and Nose, 2004; Abraham *et al.*, 2016; Rainha *et al.*, 2016). However, diel citrate fluctuations during early phases of CAM (e.g. in *M. crystallinum* and *Clusia minor*) (Herppich *et al.*, 1995; Borland *et al.*, 1998) and a contribution of succinate to nocturnal acidification [e.g. in *Ananas comosus* (pineapple) and *Agave americana*] have also been reported (Medina *et al.*, 1993; Abraham *et al.*, 2016; Rainha *et al.*, 2016). In ABA-induced CAM in *T. triangulare*, the increase in succinate levels was also accompanied by α-ketoglutarate accumulation (Fig. 2).

Transcript levels of several transport proteins were affected by exogenous ABA, as observed in CAM-performing *M. crystallinum* (Cushman *et al.*, 2008). The reduced transcript levels of the chloroplast malate transporter *DiT1* in ABA-treated leaves after 320 and 1280 min (Fig. 2) could indicate reduced flow of malate into the chloroplast towards the end of the light period and in the dark period, suggesting increased flow into the vacuole. However, transcripts of *ALUMINIUM-ACTIVATED MALATE TRANSPORTER*s did not accumulate significantly upon ABA treatment and were of low abundance (Supplementary Dataset S1), an observation also made in drought-induced CAM (Brilhaus *et al.*, 2016). In contrast, *BASS2* transcripts accumulated promptly (starting at 320 min, Fig. 2; cluster 6, Supplementary Dataset S1), suggesting adjustment for increased transport of pyruvate originating from malate decarboxylation to the chloroplast.

Storage of malate in the vacuole requires a proton gradient to fuel its transport (Lüttge, 1987). Indeed, *VACUOLAR-TYPE H^+ ATPASE C2* (*AVA-P2*) (cluster 6) was slightly but significantly up-regulated in response to ABA. In addition, transcripts of a putative ATPase activator *AT5G58110* (cluster 5) accumulated (Supplementary Dataset S1). Although it is not clear whether V-ATPase activity is controlled at the level of transcript abundance, ABA-induced activity of V-ATPase was observed in *M. crystallinum* (Barkla *et al.*, 1999).

Utilization of carbohydrate reserves

Reprogramming of carbohydrate metabolism ensures the flow of carbon skeletons between storage carbohydrates and the PEP pool. Genes involved in starch metabolism were down-regulated in ABA-treated leaves, especially at the final two time points (Supplementary Fig. S5B). Transcript levels of *ISA1* (after 160 min), *DBE1* (320 min), and *GBBS1* (1280 min) were depleted (Fig. 2). Glycolytic genes were also down-regulated (Supplementary Fig. S5B), including early-responsive *PGI1* (after 160 min) and *FBA2*, which showed the strongest transcript depletion after 1280 min (Fig. 2). While the time-course experiment presented here covered a period of time too short to follow full diel cycling of transcript abundances, these early changes suggest ABA- and time-dependent adjustment of glycolysis, as observed in *M. crystallinum* for *FRUCTOSE-BISPHOSPHATE ALDOLASE 1* and *GLUCOSE-6-PHOSPHATE ISOMERASE 1* (Cushman *et al.*, 2008) and by fluctuating mid-day and mid-night transcript levels of glycolytic genes in drought-induced CAM in *T. triangulare* (Brilhaus *et al.*, 2016).

Starch breakdown products enter the cytosolic glycolytic pathway and, while *GLUCOSE-6-PHOSPHATE TRANSPORTER* (*GPT2*) transcripts accumulated (after 640 and 1280 min; Supplementary Dataset S1), the overall pattern of abundance did not differ significantly between the treatments. The nocturnal accumulation of *GPT2* transcripts agrees with observations in pineapple (Borland *et al.*, 2016) and *M. crystallinum* (Neuhaus and Schulte, 1996; Häusler *et al.*, 2000). *PPDK* transcripts accumulated rapidly (after 320 and 640 min; Fig. 2), suggesting production of the PEP pool, which correlates with the trend towards declining amounts of starch, also starting after 320 min (Supplementary Fig. S8), and accumulating transcripts of starch-degrading enzymes such as *BAM1* and *AMY3* (Fig. 2). Besides starch, sucrose levels also showed a declining trend with progressing time (Supplementary Fig. S8), suggesting the utilization of sucrose as an additional source of carbon backbones. Alternatively, carbohydrate reserves could be used in respiration to provide energy for synthesis and necessary transcriptional and translational adjustment during the induction phase of CAM.

Transcripts of the vacuolar hexose transporters *VGT2* and *ERD4* accumulated in ABA-treated leaves (Fig. 2), indicating increased sugar flow between the cytosol and vacuole. Depending on current demands, glucose stored in the vacuole either supports PEP regeneration or provides energy for maintenance processes and growth (Borland *et al.*, 2016). In addition, exogenous ABA induced accumulation of transcripts of plasma membrane-localized *STP7* and *SWEET12* (Fig. 2). While diel cycling between synthesis and degradation of storage carbohydrates is well described (Borland and Taybi, 2004; Taybi *et al.*, 2017), transcriptional responsiveness of numerous transporters suggests the requirement for proper partitioning of sugar resources between compartments and possibly cells or tissues. In addition, sugars could also play a signalling role during CAM induction, as supported by the enrichment of sugar signalling in cluster 5 (Fig. 4B).

Transcript accumulation of components of ABA biosynthetic and signalling pathways

Exogenous ABA affected the abundance patterns of several transcripts encoding components of ABA signalling, ABA-synthesizing enzymes, and ABA transporters. Transcript levels of the ABA receptors *RCAR1/PYL9* and *RCAR3/PYL8* declined in response to ABA after 640 and 1280 min, respectively. *PYL4, PYR1*, and *PYR4* were also down-regulated at least at one time point (Supplementary Dataset S1), but the overall expression pattern did not differ significantly between the treatments. While depletion of ABA receptor transcripts is a common response to both ABA and abiotic stress, occurring in a wide range of species such as Arabidopsis (Chan, 2012; Gonzalez-Guzman *et al.*, 2012; Song *et al.*, 2016), cotton (Zhang *et al.*, 2017), and maize (Fan *et al.*, 2016; Li *et al.*, 2017), diverse responses have been observed. For example, reversed expression patterns between maize roots and shoots (Fan *et al.*, 2016) and time-dependent expression patterns in cotton (Zhang *et al.*, 2017) were described. Given the spatiotemporal expression of ABA receptors observed in other species, it is intriguing to hypothesize that differentiation between the stress response and CAM induction could occur at the level of ABA interaction partners.

In contrast to the transcript levels of genes encoding ABA receptors, transcript levels of downstream components changed more promptly and were generally up-regulated; among them were the PP2CA phosphatases *ABI1*, *HAI3*, and *PP2CA*, as well as their target kinases *SnRK2.6/OST1 SnRK2.8*. In contrast, *SnRK2.5* transcripts diminished upon ABA treatment (Fig. 3). This finding is not necessarily contradictory, as only selected *SnRK2*s were shown to act in an ABA-dependent manner

under hyperosmotic conditions in Arabidopsis (Boudsocq *et al*., 2004) and differentiation in *SnRK2* transcriptional responsiveness depending on the type of stress was observed in *Zea mays* (Huai *et al*., 2008).

Transcripts of several genes encoding ABA-synthesizing enzymes accumulated in response to ABA (Supplementary Dataset S1), but only *ABA1* and *AAO1* showed altered temporal patterns (Fig. 3). These changes were accompanied by transcript accumulation of *ABCG25*, which encodes an ABA exporter from the site of synthesis (Kuromori *et al*., 2010). Induction of *AAO3* and *NCED3* expression by exogenous ABA has been observed in Arabidopsis, even though the latter was detected only in the Landsberg *erecta* background (Xiong and Zhu, 2003). Exogenous ABA led to increased endogenous ABA levels in *M. crystallinum* (Taybi and Cushman, 2002), suggesting that ABA-induced ABA biosynthesis is conserved between species.

Protein degradation during CAM induction

Previous studies have revealed the importance of *de novo* protein synthesis in the adjustment of enzymatic machinery and in signalling during CAM induction in *M. crystallinum* (Höfner *et al*., 1987; Thomas *et al*., 1992; Taybi and Cushman, 2002). In the present study, the total protein amount declined after 320 min in response to ABA treatment, which correlated with the observed transcriptional changes. Transcripts encoding protein-degrading enzymes accumulated in ABA-treated leaves, starting at 40 min and peaking after 1280 min (Supplementary Fig. S7). Besides up-regulation of protein degradation, the enrichment analysis revealed down-regulation of protein-synthesis enzymes (Supplementary Fig. S5). These findings are consistent with NMF, which identified under-representation of protein biosynthesis in all three clusters composed of ABA-specific factors (clusters 5–7; Fig. 4B).

Protein degradation could serve multiple purposes. First, protein breakdown would release amino acids for the synthesis of new proteins, including CAM-specific enzymes. Rapid utilization of released amino acids is supported by the lack of accumulation of free amino acids (Supplementary Dataset S3). Second, protein degradation may contribute to adjustment of the enzymatic machinery by degrading competing C_3 enzymes or proteins with diurnally fluctuating amounts, such as PPCK (Nimmo, 2003). For completeness, turnover of components of the ABA signalling pathway should be mentioned (Wu *et al*., 2016; Li *et al*., 2017). Finally, degraded proteins might provide energy to fuel synthesis and transport processes. Nocturnal accumulation of TCA intermediates in ABA-treated leaves suggest increased flow through this pathway, as proposed by flux-balance analysis by Cheung *et al*. (2014). Under conditions of rapid CAM induction, this might enable preferential use of carbohydrate reserves for PEP synthesis.

Short-term ABA treatment did not induce a general stress response

Given the altered transcript levels of ABA signalling components, the next question was whether the metabolome of ABA-treated leaves showed a general stress response, manifested through the accumulation of sugars, sugar alcohols, or amino acids (Papageorgiou and Murata, 1995; Bohnert and Jensen, 1996; Hare *et al*., 1998). In Arabidopsis, increased salinity as well as exogenous ABA reduced sucrose and starch content, while maltose content increased (Kempa *et al*., 2008). This result agrees with the response observed in ABA-treated leaves of *T. triangulare*. However, given the nocturnal increase in titratable acidity (Fig. 2), it is likely that starch degradation primarily served PEP generation. Raffinose levels remained largely unchanged in ABA-treated leaves of *T. triangulare* (Supplementary Dataset S4B), indicating that raffinose accumulation after water withdrawal reported in the same species (Brilhaus *et al*., 2016) was (i) independent of ABA and (ii) unrelated to CAM induction. The former agrees with observations made in Arabidopsis, where raffinose accumulation is also independent of ABA (Kempa *et al*., 2008; Urano *et al*., 2009). Amounts of the sugar alcohols glycerol, mannitol, myo-inositol, and sorbitol did not increase after ABA treatment (Supplementary Dataset S4B). This is not surprising at least in the case of myo-inositol, as its biosynthesis was shown to be independent of ABA in Arabidopsis (Urano *et al*., 2009). GABA accumulates in response to various stresses, possibly having a signalling role (Bouché and Fromm, 2004), but no increase in GABA levels was observed in ABA-treated *T. triangulare* leaves. In fact, GABA levels even decreased transiently, an observation also made in Arabidopsis (Kempa *et al*., 2008; Urano *et al*., 2010). In Arabidopsis, stress-induced accumulation of amino acids was observed within 15–18 hours (Nambara *et al*., 1998; Urano *et al*., 2009). While the accumulation of branched-chain amino acids, tyrosine, and histidine depends on ABA, dehydration-induced accumulation of proline can also occur independently of ABA (Kempa *et al*., 2008; Urano *et al*., 2009, 2010). In *T. triangulare*, no significant accumulation of amino acids was observed either under drought stress (Brilhaus *et al*., 2016) or in response to exogenous ABA (Supplementary Dataset S4B).

In summary, among the stress-related metabolites, those most affected by ABA application were starch and sucrose, but we attribute their reduced levels to the PEP regeneration typical of CAM and energy provision to fuel the adjustment of metabolism. The lack of stress response may be due to the use of exogenous ABA as the only stimulus. It has been proposed that stress (e.g. drought) is sensed by multiple sensors, which induce a variety of secondary signals (Xiong *et al*., 2002). Besides ABA, stress signalling also relies on Ca^{2+} and reactive oxygen species, and it is possible that the sole application of ABA was not enough to induce a stress response as rapidly as it induced the CAM pathway. Measurements of photosynthetic yield, which is sensitive to environmental and other stresses (Maxwell and Johnson, 2000), did not reveal any damage to photosystem II (Supplementary Protocol S3), in agreement with the metabolome analysis.

Candidate regulators of CAM induction

The currently available transcriptome data with ~25% ABA-responsive genes (this study) and ~39% drought-responsive genes (Brilhaus *et al*., 2016) in *T. triangulare* represent a rich

source for the identification of regulators downstream of the two, partially overlapping, signalling networks. Based on the DESeq analysis and additional filtering criteria, seven candidate transcriptional regulators were identified (Table 1). *HEAT SHOCK TRANSCRIPTION FACTOR A2* (*HSFA2*), the dominant *HSF* involved in thermotolerance of both Arabidopsis and tomato (Charng *et al.*, 2007; Chan-Schaminet *et al.*, 2009), also contributes to salt and osmotic stress tolerance (Ogawa *et al.*, 2007). Since CAM can be induced by both drought and salinity, a TF responsive to various environmental stresses in an ABA-dependent manner would be a meaningful CAM inducer. *HSFC1* was also up-regulated. In Arabidopsis, *HSFC1* has been shown to be heat- and drought-responsive (Rizhsky *et al.*, 2004).

NUCLEAR FACTOR Y (*NF-Y*) *SUBUNITS A9* and *B3* were up-regulated in ABA-treated leaves (Table 1). NF-Ys act as heterodimers and heterotrimers of NF-YA, NF-YB, and NF-YC subunits and form complexes with other proteins, including bZIP-family TFs and components of phytochrome, gibberellic acid, and ABA signalling. NF-Y TFs regulate plant growth, development, and stress responses (Zhao *et al.*, 2017). This is also true for NF-YB3, which is part of the complex binding to the DREB2A promoter motif, thus regulating the heat stress response (Sato *et al.*, 2014). Over-expression of poplar *NF-YA9* in Arabidopsis revealed its role in promoting stomatal closure in an ABA-dependent manner (Lian *et al.*, 2018). While *NF-YB3* up-regulation might suggest a developing stress response, it could also connect CAM induction with the necessary adjustment of stomatal behaviour.

JmjC DOMAIN-CONTAINING PROTEIN 27 (*JMJ27*) acts as a demethylase targeting histone H3 at lysine 9 (H3K9) and plays a role in both pathogen response and flowering time regulation in Arabidopsis. In the defence response, JMJ27 acts as a suppressor of WRKY TF expression (Dutta *et al.*, 2017). Transcript abundances of several WRKYs were affected in ABA-treated leaves, but these were primarily up-regulated, including *WRKY33* (Supplementary Dataset S1). *MONOPTEROS/AUXIN RESPONSE FACTOR 5* (*MP/ARF5*) regulates development in an auxin-dependent manner (Hardtke *et al.*, 1998; Krogan *et al.*, 2012). Since phytohormone signalling pathways frequently overlap (Jaillais and Chory, 2010), the involvement of *MP* in the ABA response, at least under some circumstances, cannot be excluded. TFs identified as putative CAM regulators share their involvement in responses to stressors, including heat, a biological function that was over-represented in the late ABA cluster 5 (Fig. 4B). In addition, numerous identified TFs are related to development and response to other phytohormones.

The abundances of transcripts encoding transport proteins and enzymes of amino acid metabolism were affected by exogenous ABA as well, but it can only be speculated whether transcription of any of these genes is regulated by either of the abovementioned TFs. The altered flow of metabolites could directly contribute to their flow through the CAM pathway, as in the case of BASS2 (Table 1), transcripts of which are highly abundant in the C_4 species *Flaveria*, which employs a type of photosynthesis that shares many similarities with CAM (Furumoto *et al.*, 2011). Alternatively, transporters may contribute to signalling, which might be the role of CATION EXCHANGER 11 (CAX11), even though its role in Ca^{2+} signalling under stress conditions has been described only in Arabidopsis roots so far (Wang *et al.*, 2016).

The number of ABA-responsive transcripts encoding enzymes of amino acid metabolism could contribute to protein synthesis. For example, *METHIONINE GAMMA-LYASE* (*MGL*) has been shown to catalyse conversion of methionine to cysteine (Goyer *et al.*, 2007) and isoleucine synthesis (Joshi and Jander, 2009). Moreover, amino acid metabolism could play a more global role in the adjustment of metabolism. Mitochondrial *MCCA*, encoding methylcrotonyl-CoA carboxylase alpha chain, is involved in the catabolism of leucine and branched-chain amino acids, thus providing a respiratory substrate to satisfy energy demands. Besides *MCCA*, *THA1* was up-regulated in response to both ABA and drought (Table 1). In Arabidopsis, both *MCCA* and *THA1* are targets of bZIP TFs (Doidy *et al.*, 2016), leading to the hypothesis that both genes could be co-regulated in *T. triangulare* as well.

Based on Arabidopsis coexpression network analysis, a hierarchical structure of catabolism genes was suggested in which smaller pathways, such as single amino acid degradation, form a part of larger biochemical modules, meaning that specialized genes also possesses a broader metabolic function (Mentzen *et al.*, 2008). However, whether such regulation could also contribute to CAM induction (e.g. via amino acid catabolism genes) has not been investigated yet.

Conclusions

Based on the observed transcriptional changes and increased nocturnal acidification, foliar application of 200 µM ABA was sufficient to induce CAM in the facultative CAM species *T. triangulare* within 24 hours. CAM induction by exogenous ABA has been reported previously (Ting, 1981; Chu *et al.*, 1990; Dai *et al.*, 1994; Taybi *et al.*, 1995; Minardi *et al.*, 2014) but we are not aware of any other work showing the rapid pace of adjustment of the carbon assimilation strategy. Establishing ABA-inducible CAM models could boost research in the field and expand our knowledge about this pathway, including but not limited to functional confirmation of proposed 'CAM switches', gas exchange studies and labelling experiments to disentangle carbon flow during the induction phase of CAM.

Supplementary data

Supplementary data are available at *JXB* online.

Dataset S1. Quantitative information and annotation for all *Talinum triangulare* reads mapped on to the reference transcriptome of *Beta vulgaris*.

Dataset S2. Genes with significantly altered temporal pattern of transcript abundances upon ABA treatment of *Talinum triangulare*; analysis with maSigPro package for R.

Dataset S3. A complete list of significantly differentially expressed genes based on comparison of the pool of ABA-treated leaves of *Talinum triangulare* with the pool of mock-treated leaves.

Dataset S4. Relative metabolite amounts in *Talinum triangulare* leaves as determined by GC-MS, and amino acid amounts in *Talinum triangulare* leaves as determined by LC-MS.

Fig. S1. Titratable acidity in mature *Talinum triangulare* leaves at the end of the light and dark phases upon daily treatments with 10–600 μM ABA.

Fig. S2. Principal component analysis of *Talinum triangulare* RNA-seq data.

Fig. S3. Principal component analysis of metabolite amounts detected in leaves of *Talinum triangulare* harvested 40, 80, 160, 320, 640, and 1280 min after the first ABA treatment.

Fig. S4. Comparison of performance of Lee and Seung, Brunet, and KL algorithms on the transcriptome dataset of *Talinum triangulare*.

Fig. S5. Over-representation of differentially expressed genes for biological functions in *Talinum triangulare* at individual sampling time points.

Fig. S6. Transcript abundance of three detected isoforms of *PPC* in *Talinum triangulare*.

Fig. S7. ABA-induced changes in abundances of transcripts encoding components of protein degradation/synthesis and amino acid degradation/synthesis pathways in *Talinum triangulare*.

Fig. S8. Diurnal pattern of sucrose and starch abundance in ABA- and mock-treated leaves of *Talinum triangulare*.

Protocol S1. Validation of transcript levels determined by RNA-seq.

Protocol S2. Time series RNA-seq analysis.

Protocol S3. Pulse-amplitude modulated (PAM) fluorescence measurements.

Table S1. Sequencing statistics of the RNA-seq experiment in *Talinum triangulare*.

Table S2. Gene assignment to MapMan and to homemade reduced categories with expanded Photosynthesis.CAM/C4 photosynthesis category.

Table S3. Genes involved in clock regulation and analysis of time-dependent abundances of their transcript levels in *Talinum triangulare*.

Acknowledgements

We thank M. Graf, E. Klemp, and K. Weber for their expert technical assistance with the GC-MS and LC-MS measurements. This work was funded by grants (to APMW) of the Deutsche Forschungsgemeinschaft (IRTG 1525, and under Germany's Excellence Strategy – EXC-2048/1 – Project ID: 390686111), and a graduate fellowship by the German Academic Exchange Service (DAAD) to EM.

References

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

Antoni R, Rodriguez L, Gonzalez-Guzman M, Pizzio GA, Rodriguez PL. 2011. News on ABA transport, protein degradation, and ABFs/WRKYs in ABA signaling. Current Opinion in Plant Biology **14**, 547–553.

Barkla BJ, Vera-Estrella R, Maldonado-Gama M, Pantoja O. 1999. Abscisic acid induction of vacuolar H+-ATPase activity in *Mesembryanthemum crystallinum* is developmentally regulated. Plant Physiology **120**, 811–820.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological) **57**, 289–300.

Bohnert HJ, Jensen RG. 1996. Strategies for engineering water-stress tolerance in plants. Trends in Biotechnology **14**, 89–97.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **30**, 2114–2120.

Borland AM, Guo HB, 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, Weston DJ, Schlauch KA, Tschaplinski TJ, Tuskan GA, Yang X, Cushman JC. 2014. Engineering crassulacean acid metabolism to improve water-use efficiency. Trends in Plant Science **19**, 327–338.

Borland AM, Taybi T. 2004. Synchronization of metabolic processes in plants with Crassulacean acid metabolism. Journal of Experimental Botany **55**, 1255–1265.

Borland AM, Técsi LI, Leegood RC, Walker RP. 1998. Inducibility of crassulacean acid metabolism (CAM) in *Clusia* species; physiological/biochemical characterisation and intercellular localization of carboxylation and decarboxylation processes in three species which exhibit different degrees of CAM. Planta **205**, 342–351.

Bouché N, Fromm H. 2004. GABA in plants: just a metabolite? Trends in Plant Science **9**, 110–115.

Boudsocq M, Barbier-Brygoo H, Laurière C. 2004. Identification of nine sucrose nonfermenting 1-related protein kinases 2 activated by hyperosmotic and saline stresses in *Arabidopsis thaliana*. The Journal of Biological Chemistry **279**, 41758–41766.

Boxall SF, Dever LV, Kneřová J, Gould PD, Hartwell J. 2017. Phosphorylation of phosphoenolpyruvate 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*. The Plant Cell **29**, 2519–2536.

Brilhaus D, Bräutigam A, Mettler-Altmann T, Winter K, Weber AP. 2016. Reversible burst of transcriptional changes during induction of crassulacean acid metabolism in *Talinum triangulare*. Plant Physiology **170**, 102–122.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics **10**, 421.

Chan Z. 2012. Expression profiling of ABA pathway transcripts indicates crosstalk between abiotic and biotic stress responses in Arabidopsis. Genomics **100**, 110–115.

Chan-Schaminet KY, Baniwal SK, Bublak D, Nover L, Scharf KD. 2009. Specific interaction between tomato HsfA1 and HsfA2 creates heteroligomeric superactivator complexes for synergistic activation of heat stress gene expression. The Journal of Biological Chemistry **284**, 20848–20857.

Charng YY, Liu HC, Liu NY, Chi WT, Wang CN, Chang SH, Wang TT. 2007. A heat-inducible transcription factor, HsfA2, is required for extension of acquired thermotolerance in Arabidopsis. Plant Physiology **143**, 251–262.

Chen LS, Lin Q, Nose A. 2002. A comparative study on diurnal changes in metabolite levels in the leaves of three crassulacean acid metabolism (CAM) species, *Ananas comosus*, *Kalanchoë daigremontiana* and *K. pinnata*. Journal of Experimental Botany **53**, 341–350.

Chen LS, Nose A. 2004. Day-night changes of energy-rich compounds in crassulacean acid metabolism (CAM) species utilizing hexose and starch. Annals of Botany **94**, 449–455.

Cheung CY, Poolman MG, Fell DA, Ratcliffe RG, 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.

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

Cushman JC, Tillett RL, Wood JA, Branco JM, Schlauch KA. 2008. Large-scale mRNA expression profiling in the common ice plant, *Mesembryanthemum crystallinum*, performing C_3 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.

Di Martino C, Delfine S, Pizzuto R, Loreto F, Fuggi A. 2003. Free amino acids and glycine betaine in leaf osmoregulation of spinach responding to increasing salt stress. New Phytologist **158**, 455–463.

Dohm JC, Minoche AE, Holtgrwe D, *et al*. 2014. The genome of the recently domesticated crop plant sugar beet (*Beta vulgaris*). Nature **505**, 546–549.

Doidy J, Li Y, Neymotin B, Edwards MB, Varala K, Gresham D, Coruzzi GM. 2016. "Hit-and-Run" transcription: *de novo* transcription initiated by a transient bZIP1 "hit" persists after the "run". BMC Genomics **17**, 92.

Dutta A, Choudhary P, Caruana J, Raina R. 2017. JMJ27, an Arabidopsis H3K9 histone demethylase, modulates defense against *Pseudomonas syringae* and flowering time. The Plant Journal **91**, 1015–1028.

Emanuelsson O, Nielsen H, Brunak S, von Heijne G. 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. Journal of Molecular Biology **300**, 1005–1016.

Fan W, Zhao M, Li S, Bai X, Li J, Meng H, Mu Z. 2016. Contrasting transcriptional responses of PYR1/PYL/RCAR ABA receptors to ABA or dehydration stress between maize seedling leaves and roots. BMC Plant Biology **16**, 99.

Fiehn O, Kopka J, Dörmann P, Altmann T, Trethewey RN, Willmitzer L. 2000. Metabolite profiling for plant functional genomics. Nature Biotechnology **18**, 1157–1161.

Furumoto T, Yamaguchi T, Ohshima-Ichie Y, *et al.* 2011. A plastidial sodium-dependent pyruvate transporter. Nature **476**, 472–475.

Gaujoux R, Seoighe C. 2010. A flexible R package for nonnegative matrix factorization. BMC Bioinformatics **11**, 367.

Gonzalez-Guzman M, Pizzio GA, Antoni R, *et al.* 2012. *Arabidopsis* PYR/PYL/RCAR receptors play a major role in quantitative regulation of stomatal aperture and transcriptional response to abscisic acid. The Plant Cell **24**, 2483–2496.

Goyer A, Collakova E, Shachar-Hill Y, Hanson AD. 2007. Functional characterization of a methionine γ-lyase in *Arabidopsis* and its implication in an alternative to the reverse trans-sulfuration pathway. Plant & Cell Physiology **48**, 232–242.

Hardtke CS, Berleth T. 1998. The *Arabidopsis* gene *MONOPTEROS* encodes a transcription factor mediating embryo axis formation and vascular development. The EMBO Journal **17**, 1405–1411.

Hare PD, Cress WA, van Staden J. 1998. Dissecting the roles of osmolyte accumulation during stress. Plant, Cell and Environment **21**, 535–553.

Hartigan JA, Wong MA. 1979. Algorithm AS 136: A *K*-means clustering algorithm. Journal of the Royal Statistical Society. Series C (Applied Statistics) **28**, 100–108.

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

Häusler RE, Baur B, Scharte J, Teichmann T, Eicks M, Fischer KL, Flügge UI, Schubert S, Weber A, Fischer K. 2000. Plastidic metabolite transporters and their physiological functions in the inducible crassulacean acid metabolism plant *Mesembryanthemum crystallinum*. The Plant Journal **24**, 285–296.

Herppich M, von Willert DJ, Herppich WB. 1995. Diurnal rhythm in citric acid content preceded the onset of nighttime malic acid accumulation during metabolic changes from C_3 to CAM in salt-stressed plants of *Mesembryanthemum crystallinum*. Journal of Plant Physiology **147**, 38–42.

Herrera A, Delgado J, Paraguatey I. 1991. Occurrence of inducible crassulacean acid metabolism in leaves of *Talinum triangulare* (Portulacaceae). Journal of Experimental Botany **42**, 493–499.

Höfner R, Vazquez-Moreno L, Winter K, Bohnert HJ, Schmitt JM. 1987. Induction of crassulacean acid metabolism in *Mesembryanthemum crystallinum* by high salinity: mass increase and *de novo* synthesis of PEP-carboxylase. Plant Physiology **83**, 915–919.

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

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

Huai J, Wang M, He J, Zheng J, Dong Z, Lv H, Zhao J, Wang G. 2008. Cloning and characterization of the SnRK2 gene family from *Zea mays*. Plant Cell Reports **27**, 1861–1868.

Jaillais Y, Chory J. 2010. Unraveling the paradoxes of plant hormone signaling integration. Nature Structural & Molecular Biology **17**, 642–645.

Jin J, Tian F, Yang DC, Meng YQ, Kong L, Luo J, Gao G. 2017. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Research **45**, D1040–D1045.

Joshi V, Jander G. 2009. Arabidopsis methionine γ-lyase is regulated according to isoleucine biosynthesis needs but plays a subordinate role to threonine deaminase. Plant Physiology **151**, 367–378.

Kempa S, Krasensky J, Dal Santo S, Kopka J, Jonak C. 2008. A central role of abscisic acid in stress-regulated carbohydrate metabolism. PLoS ONE **3**, e3935.

Kent WJ. 2002. BLAT—the BLAST-like alignment tool. Genome Research **12**, 656–664.

Krogan NT, Ckurshumova W, Marcos D, Caragea AE, Berleth T. 2012. Deletion of *MP/ARF5* domains III and IV reveals a requirement for *Aux/IAA* regulation in *Arabidopsis* leaf vascular patterning. New Phytologist **194**, 391–401.

Kuromori T, Miyaji T, Yabuuchi H, Shimizu H, Sugimoto E, Kamiya A, Moriyama Y, Shinozaki K. 2010. ABC transporter AtABCG25 is involved in abscisic acid transport and responses. Proceedings of the National Academy of Sciences, USA **107**, 2361–2366.

Li P, Cao W, Fang H, *et al*. 2017. Transcriptomic profiling of the maize (*Zea mays* L.) leaf response to abiotic stresses at the seedling stage. Frontiers in Plant Science **8**, 290.

Lian C, Li Q, Yao K, Zhang Y, Meng S, Yin W, Xia X. 2018. *Populus trichocarpa PtNF-YA9*, a multifunctional transcription factor, regulates seed germination, abiotic stress, plant growth and development in *Arabidopsis*. Frontiers in Plant Science **9**, 954.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology **15**, 550.

Lüttge U. 1987. Carbon dioxide and water demand: crassulacean acid metabolism (CAM), a versatile ecological adaptation exemplifying the need for integration in ecophysiological work. New Phytologist **106**, 593–629.

Lüttge U. 1988. Day-night changes of citric-acid levels in crassulacean acid metabolism: phenomenon and ecophysiological significance. Plant, Cell and Environment **11**, 445–451.

Lüttge U. 2004. Ecophysiology of Crassulacean acid metabolism (CAM). Annals of Botany **93**, 629–652.

Maxwell K, Johnson GN. 2000. Chlorophyll fluorescence—a practical guide. Journal of Experimental Botany **51**, 659–668.

Medina E, Popp M, Olivares E, Janett H-P, Lüttge U. 1993. Daily fluctuations of titratable acidity, content of organic acids (malate and citrate) and soluble sugars of varieties and wild relatives of *Ananas comosus* L. growing under natural tropical conditions. Plant, Cell and Environment **16**, 55–63.

Mentzen WI, Peng J, Ransom N, Nikolau BJ, Wurtele ES. 2008. Articulation of three core metabolic processes in Arabidopsis: fatty acid biosynthesis, leucine catabolism and starch metabolism. BMC Plant Biology **8**, 76.

Minardi BD, Voytena APL, Santos M, Randi ÁM. 2014. Water stress and abscisic acid treatments induce the CAM pathway in the epiphytic fern *Vittaria lineata* (L.) Smith. Photosynthetica **52**, 404–412.

Montero E, Francisco AM, Montes E, Herrera A. 2018. Salinity induction of recycling Crassulacean acid metabolism and salt tolerance in plants of *Talinum triangulare*. Annals of Botany **121**, 1333–1342.

Nambara E, Kawaide H, Kamiya Y, Naito S. 1998. Characterization of an *Arabidopsis thaliana* mutant that has a defect in ABA accumulation: ABA-dependent and ABA-independent accumulation of free amino acids during dehydration. Plant & Cell Physiology **39**, 853–858.

Neuhaus HE, Schulte N. 1996. Starch degradation in chloroplasts isolated from C_3 or CAM (crassulacean acid metabolism)-induced *Mesembryanthemum crystallinum* L. Biochemical Journal **318**, 945–953.

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.

Nimmo HG, Fontaine V, Hartwell J, Jenkins GI, Nimmo GA, Wilkins MB. 1999. PEP carboxylase kinase is a novel protein kinase controlled at the level of expression. New Phytologist **151**, 91–97.

Nueda MJ, Tarazona S, Conesa A. 2014. Next maSigPro: updating maSigPro bioconductor package for RNA-seq time series. Bioinformatics **30**, 2598–2602.

Ogawa D, Yamaguchi K, Nishiuchi T. 2007. High-level overexpression of the *Arabidopsis HsfA2* gene confers not only increased themotolerance but also salt/osmotic stress tolerance and enhanced callus growth. Journal of Experimental Botany **58**, 3373–3383.

Papageorgiou GC, Murata N. 1995. The unusually strong stabilizing effects of glycine betaine on the structure and function of the oxygen-evolving Photosystem II complex. Photosynthesis Research **44**, 243–252.

Park SY, Fung P, Nishimura N, *et al.* 2009. Abscisic acid inhibits type 2C protein phosphatases via the PYR/PYL family of START proteins. Science **324**, 1068–1071.

Rainha N, Medeiros VP, Ferreira C, Raposo A, Leite JP, Cruz C, Pacheco CA, Ponte D, Silva AB. 2016. Leaf malate and succinate accumulation are out of phase throughout the development of the CAM plant *Ananas comosus*. Plant Physiology and Biochemistry **100**, 47–51.

Rizhsky L, Liang H, Shuman J, Shulaev V, Davletova S, Mittler R. 2004. When defense pathways collide. The response of Arabidopsis to a combination of drought and heat stress. Plant Physiology **134**, 1683–1696.

Sato H, Mizoi J, Tanaka H, *et al.* 2014. Arabidopsis DPB3-1, a DREB2A interactor, specifically enhances heat stress-induced gene expression by forming a heat stress-specific transcriptional complex with NF-Y subunits. The Plant Cell **26**, 4954–4973.

Silvera K, Neubig KM, Whitten WM, Williams NH, Winter K, Cushman JC. 2010. Evolution along the crassulacean acid metabolism continuum. Functional Plant Biology **37**, 995–1010.

Song L, Huang SC, Wise A, Castanon R, Nery JR, Chen H, Watanabe M, Thomas J, Bar-Joseph Z, Ecker JR. 2016. A transcription factor hierarchy defines an environmental stress response network. Science **354**, aag1550.

Taybi T, Cushman JC. 1999. Signaling events leading to crassulacean acid metabolism induction in the common ice plant. Plant Physiology **121**, 545–556.

Taybi T, Cushman JC. 2002. Abscisic acid signaling and protein synthesis requirements for phospho*enol*pyruvate carboxylase transcript induction in the common ice plant. Journal of Plant Physiology **159**, 1235–1243.

Taybi T, Cushman JC, Borland AM. 2017. Leaf carbohydrates influence transcriptional and post-transcriptional regulation of nocturnal carboxylation and starch degradation in the facultative CAM plant, *Mesembryanthemum crystallinum*. Journal of Plant Physiology **218**, 144–154.

Taybi T, Patil S, Chollet R, Cushman JC. 2000. A minimal serine/threonine protein kinase circadianly regulates phosphoenolpyruvate carboxylase activity in crassulacean acid metabolism-induced leaves of the common ice plant. Plant Physiology **123**, 1471–1482.

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.

Thimm O, Bläsing O, Gibon Y, Nagel A, Meyer S, Krüger P, Selbig J, Müller LA, Rhee SY, Stitt M. 2004. MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. The Plant Journal **37**, 914–939.

Thomas JC, McElwain EF, Bohnert HJ. 1992. Convergent induction of osmotic stress-responses: abscisic acid, cytokinin, and the effects of NaCl. Plant Physiology **100**, 416–423.

Ting IP. 1981. Effects of abscisic acid on CAM in *Portulacaria afra*. Photosynthesis Research **2**, 39–48.

Tsuzuki M, Miyachi S, Winter K, Edwards GE. 1982. Localization of carbonic anhydrase in Crassulacean acid metabolism plants. Plant Science Letters **24**, 211–218.

Urano K, Kurihara Y, Seki M, Shinozaki K. 2010. 'Omics' analyses of regulatory networks in plant abiotic stress responses. Current Opinion in Plant Biology **13**, 132–138.

Urano K, Maruyama K, Ogata Y, *et al.* 2009. Characterization of the ABA-regulated global responses to dehydration in *Arabidopsis* by metabolomics. The Plant Journal **57**, 1065–1078.

Wang F, Chen ZH, Liu X, Colmer TD, Zhou M, Shabala S. 2016. Tissue-specific root ion profiling reveals essential roles of the CAX and ACA calcium transport systems in response to hypoxia in Arabidopsis. Journal of Experimental Botany **67**, 3747–3762.

Winter K, Garcia M, Holtum JA. 2008. On the nature of facultative and constitutive CAM: environmental and developmental control of CAM expression during early growth of *Clusia*, *Kalanchöe*, and *Opuntia*. Journal of Experimental Botany **59**, 1829–1840.

Winter K, Holtum JA. 2014. Facultative crassulacean acid metabolism (CAM) plants: powerful tools for unravelling the functional elements of CAM photosynthesis. Journal of Experimental Botany **65**, 3425–3441.

Wu Q, Zhang X, Peirats-Llobet M, *et al.* 2016. Ubiquitin ligases RGLG1 and RGLG5 regulate abscisic acid signaling by controlling the turnover of phosphatase PP2CA. The Plant Cell **28**, 2178–2196.

Xiong L, Schumaker KS, Zhu JK. 2002. Cell signaling during cold, drought, and salt stress. The Plant Cell **14** (Suppl 1), S165– S183.

Xiong L, Zhu JK. 2003. Regulation of abscisic acid biosynthesis. Plant Physiology **133**, 29–36.

Yoshida T, Fujita Y, Sayama H, Kidokoro S, Maruyama K, Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K. 2010. AREB1, AREB2, and ABF3 are master transcription factors that cooperatively regulate ABRE-dependent ABA signaling involved in drought stress tolerance and require ABA for full activation. The Plant Journal **61**, 672–685.

Zhang G, Lu T, Miao W, Sun L, Tian M, Wang J, Hao F. 2017. Genome-wide identification of ABA receptor PYL family and expression analysis of *PYLs* in response to ABA and osmotic stress in *Gossypium*. PeerJ **5**, e4126.

Zhao H, Wu D, Kong F, Lin K, Zhang H, Li G. 2017. The *Arabidopsis thaliana* nuclear factor Y transcription factors. Frontiers in Plant Science **7**, 2045.

Supplementary figures

Figure S1. Titratable acidity in mature *Talinum triangulare* leaves at the end of the light and dark phases upon daily treatments with 10 to 600 μM ABA

Figure S2. Principle component analysis of *Talinum triangulare* RNA-seq data

Figure S3. Principle component analysis of metabolite amounts detected in leaves of *Talinum triangulare* harvested 40, 80, 160, 320, 640 and 1280 min after the first ABA treatment

Figure S4. Comparison of performance of Lee and Seung Brunet, and KL algorithms on the transcriptome dataset of *Talinum triangulare*

Figure S5. Over-representation of differentially expressed genes for biological functions in *Talinum triangulare* at individual sampling time points

Figure S6. Transcript abundance of three detected isoforms of *PPC* in *Talinum triangulare*

Figure S7. ABA-induced changes in abundances of transcripts encoding components of protein degradation/synthesis and amino acid degradation/synthesis pathways in *Talinum triangulare*

Figure S8. Diurnal pattern of sucrose and starch abundance in ABA- and mock-treated of *Talinum triangulare*

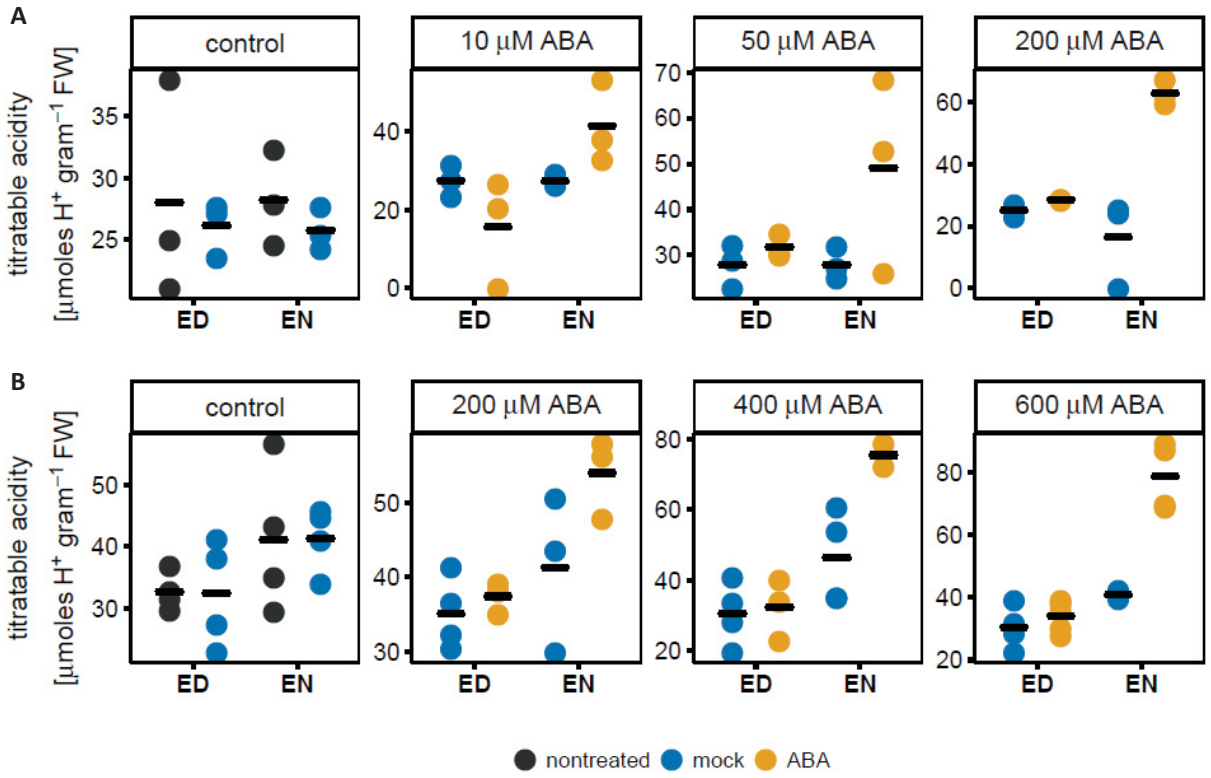


Figure S1. End of day (ED) and end of night (EN) titratable acidity in mature leaves of *Talinum triangulare;* two independent treatments, (A) and (B), with ABA solutions of the given concentration or mock solution (1.86% (v/v) methanol, 0.02% (v/v) Tween-20) applied as foliar spray within the first hour of the light. Plants were grown in 12/12 L/D regime. Paired leaves were treated with mock or ABA solution, nontreated controls were included as well. Leaves were harvested and shock frozen after daily treatments for five consecutive days. Approximately 0.5 g leaf material (fresh weight) was boiled in 40 ml 50% (v/v) methanol until the volume halved. Upon refilling to 40 ml with water, leaves were boiled further until the volume halved again. Finally, extract volume was refilled back to 40 ml one final time with water. After cooling to room temperature, obtained extracts were titrated with 5mM KOH to pH = 7.0, using a pH meter to monitor the change in pH.

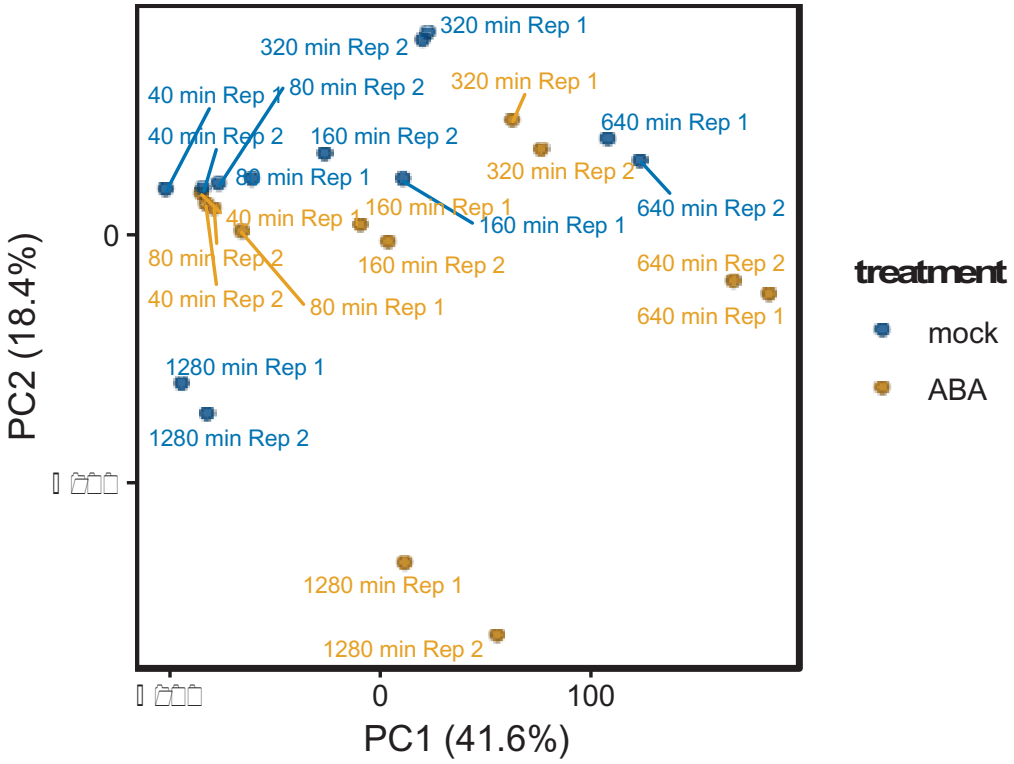


Figure S2. Principle component analysis of RNA-seq data for mock- (blue) and ABA-treated (orange) leaves of *Talinum triangulare* harvested 40, 80, 160, 320, 640 and 1280 min after the first ABA treatment.

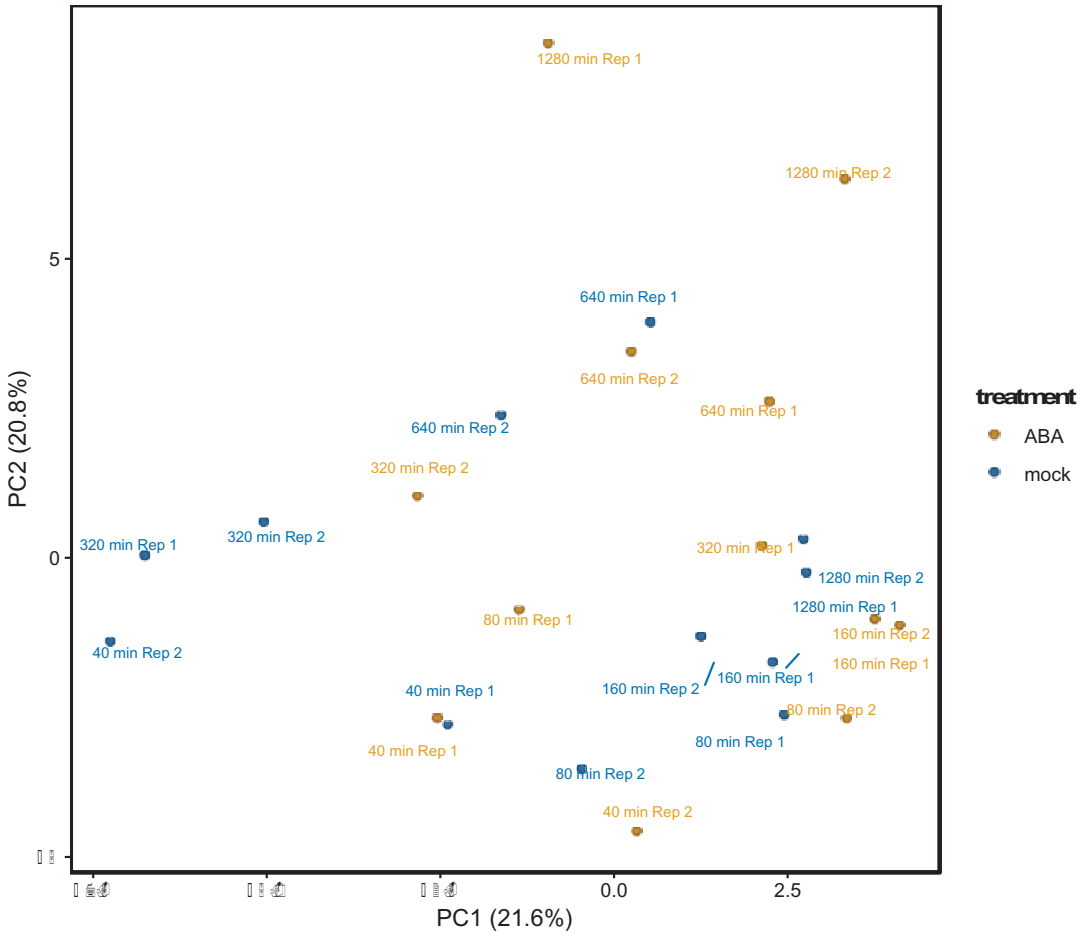


Figure S3. Principle component analysis of metabolite amounts detected in leaves of *Talinum triangulare* harvested 40, 80, 160, 320, 640 and 1280 min after the first ABA treatment (mock in blue and ABA in orange). The shown PCA is based on metabolite amounts measured with GC-MS and LC-MS.

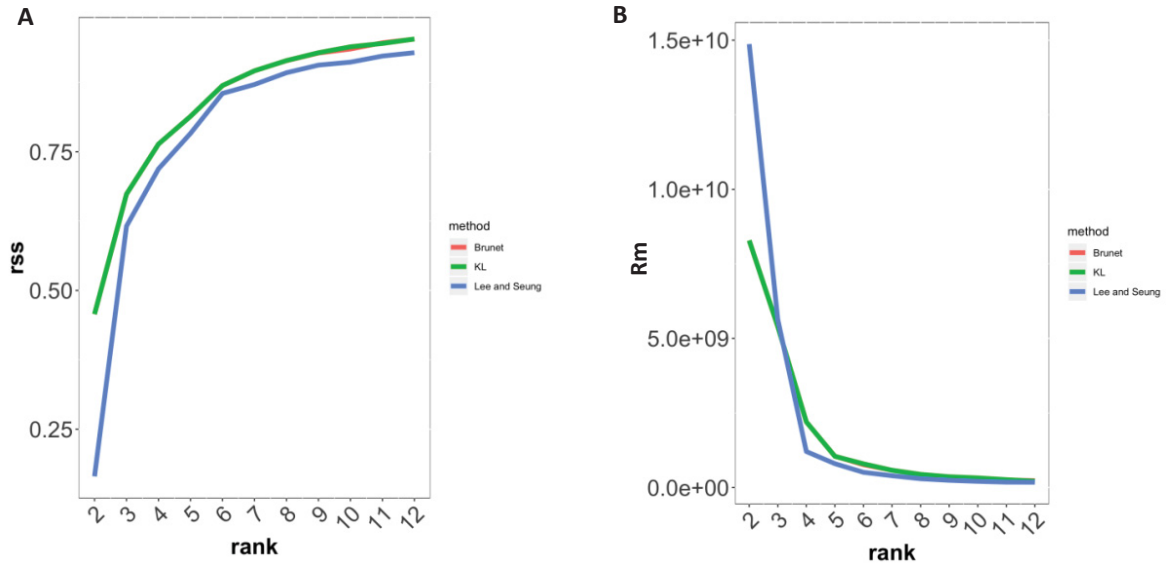


Figure S4. Comparison of performance of Lee and Seung (blue), Brunet (red), and KL (green) algorithms in terms of (A) residual sum of squares and (B) average regression coefficient (R^2). The algorithms were tested prior to running nonnegative matrix factorization (NMF) on the transcriptome data set obtained for *Talinum triangulare*. All three algorithms identified five factors.

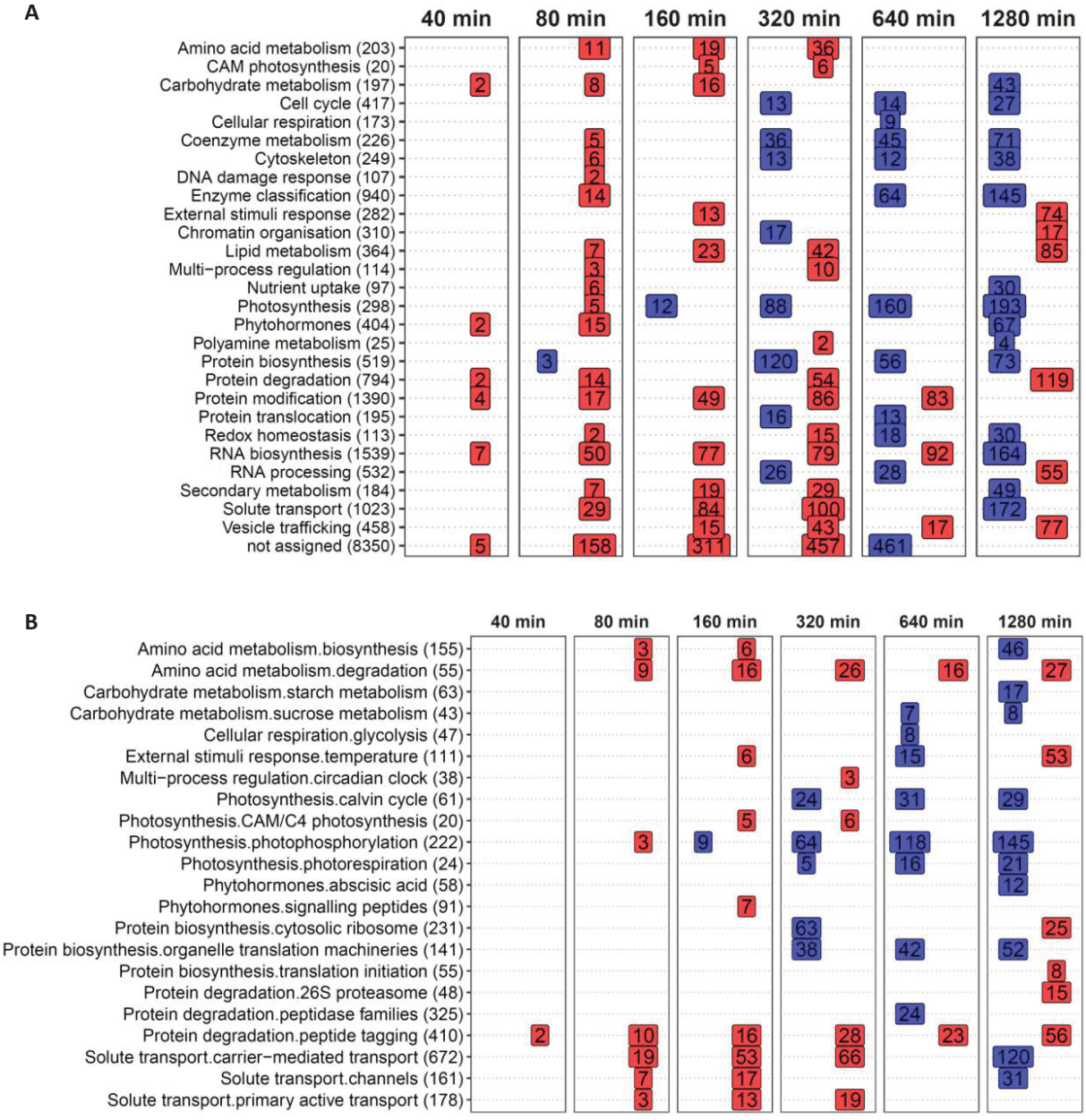


Figure S5. Over-representation of biological functions at individual sampling time points. (A) MapMan categories in the first bin. (B) Selected MapMan categories at the level of second bin. *Talinum triangulare* reads were mapped against *Beta vulgaris* reference and after mapping, only transcripts with a known homolog in Arabidopsis were used in the enrichment analysis. Up- and down-regulation of individual transcripts was based on DESeq analysis (*q*-value 0.01 after false discovery rate correction according to Benjamini & Hochberg). Enrichments are based on Wilcoxon test with false discovery rate correction according to Benjamini & Hochberg and significance level of 0.05. When multiple *Beta vulgaris* targets mapped against a single Arabidopsis gene, all hits were counted. Numbers in brackets stand for total number of mapped genes in each category. In red is number of up-regulated and in blue number of down-regulated genes The complete list of gene assignments to MapMan bins, incuding manually expanded CAM/C_4 photosynthesis, is available as Supplemental Table 2.

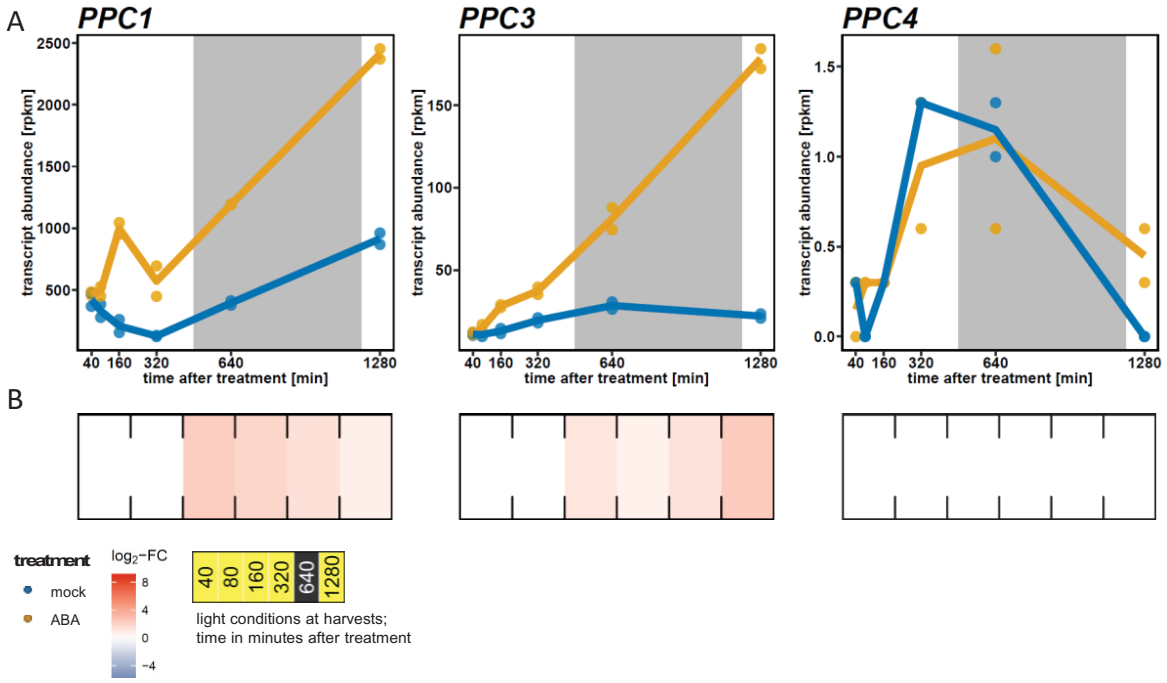


Figure S6. Transcript abundance of three detected isoforms of *PPC* in *Talinum triangulare* leaves. (A) Absolute transcript levels [rpkm]. (B) Differential expression in ABA-treated compared to mock-treated leaves (DESeq analysis). In the DESeq analysis, only significantly different (*q*-value 0.01) changes are shown.

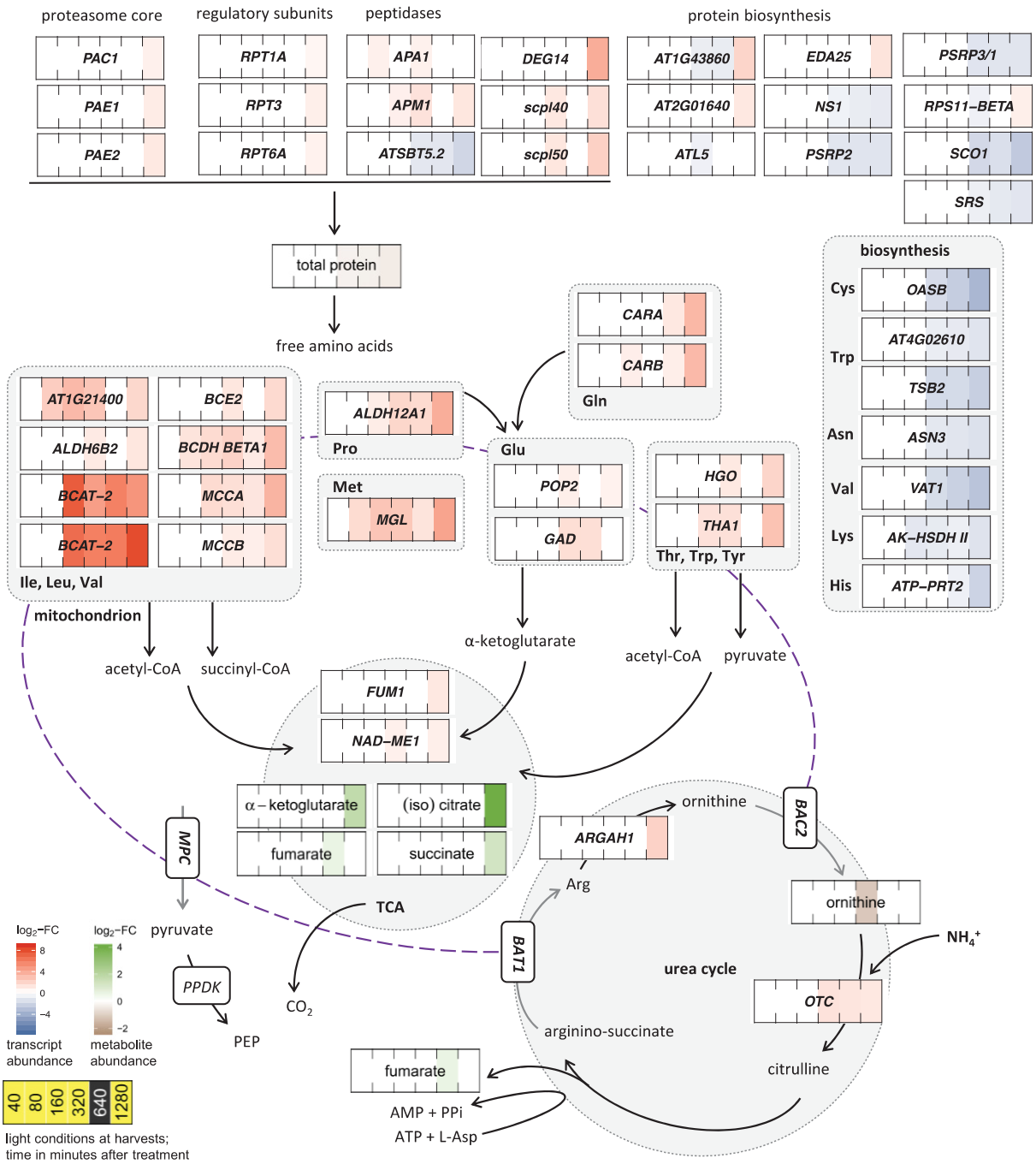


Figure S7. ABA-induced changes in abundances of transcripts encoding components of protein degradation/synthesis and amino acid degradation/synthesis pathways in *Talinum triangulare*. Transcript abundances are expressed as $\log_2$-fold changes of ABA- compared to mock-treated leaves (n = 2). Only genes with significantly different transcript abundance (DESeq with Benjamini & Hochberg correction, *q*-value < 0.01) and altered temporal pattern (maSigPro with Benjamini & Hochberg correction, *q*-value < 0.01 and R^2 > 0.8) are shown. In response to ABA, primarily protein and amino acid degradation pathways are up-regulated, with up-regulation of amino acid-degrading enzymes preceeding that of enzymes of protein degradation. Degraded amino acids provide carbon backbones in the form of pyruvate or TCA intermediates. Pyruvate could be subsequently phosphorylated by PPDK to produce PEP, a substrate for nocturnal carbon assimilation. Generated organic acids can enter the TCA cycle and thus support energy production required *e.g.* to fuel synthesis and transmembrane transport processes. Additionally, CO_2 originating both from respiration and amino acid degradation could be potentially captured by PPC besides ambient CO_2. MPC, mitochondrial pyruvate carrier; OAA, oxaloacetate; PEP, phospho*enol*pyruvate; TCA, tricarboxylic acid cycle.

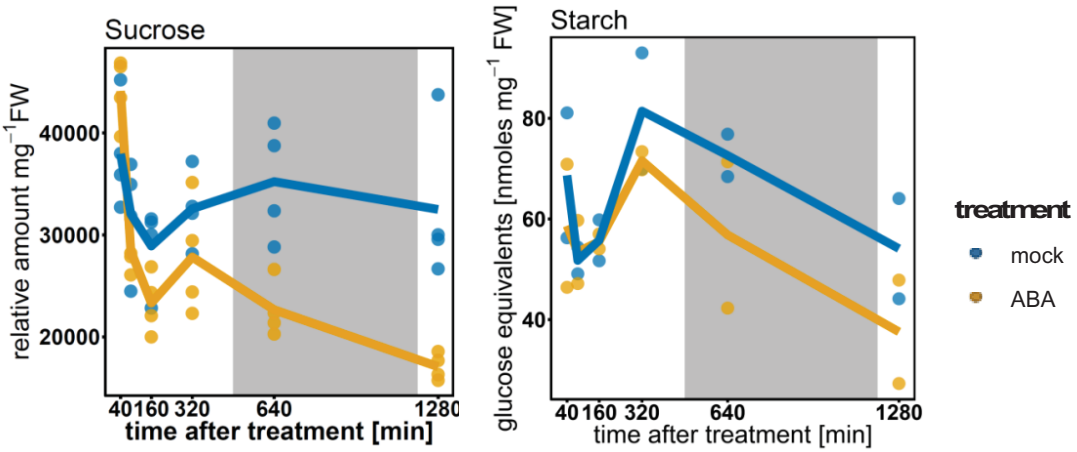


Figure S8. Diurnal pattern of sucrose and starch abundances in mock- (blue) and ABA-treated (orange) leaves of *Talinum triangulare*. While the observed changes are not significant (*q*-value 0.05), there seems to be a trend towards nocturnal depletion of both carbon sources already during the first dark period after the ABA treatment.

Supplementary tables

Table S1. Sequencing statistics of the RNA-seq experiment in *Talinum triangulare*

Table S2. Gene assignment to MapMan and to homemade reduced categories with expanded Photosynthesis.CAM/C4 photosynthesis category *(available as a supplementary material at the publisher's website only)*

https://academic.oup.com/jxb/article-lookup/doi/10.1093/jxb/erz189

Table S3. Genes involved in clock regulation and analysis of time-dependent abundances of their transcript levels in *Talinum triangulare*

Table S1. Sequencing statistics of the RNA-seq experiment in *Talinum* triangulare

[a]Total number of reads, total reads obtained per sample

[b]Reads after trimming, number of reads remaining after trimming

[c]Percent of reads dropped after the trimming

[d]Mapped reads, number of reads mapped against the Arabidopsis reference

[e]Mapping efficiency, proportion of reads sucessfully mapped against *Beta vulgaris* reference in protein space (for more details see Materials and Methods)

[f]BLAT matches, number of reference targets with at least one read mapped

[g]Sum of trimmed reads, total number of reads after trimming per time point and treatment

[h]Sum mapped reads, total number of reads (after trimming) per time point and treatment that could be mapped against the *Beta vulgaris* reference

[i]Average mapping efficiency, proportion of reads sucessfully mapped against Beta vulgaris reference in protein space per time point and treatment

[j]Targets with at least one read in rep group, number of *Beta vulgaris* reference targets with at least one read per time point and condition

[k]Proportion of the *Beta vulgaris* reference represented by the mapped reads per time point and treatment

Treatment	Time after treatment [min]	Biological replica	Sample ID	Total reads[a]	Reads after trimming[b]	Reads dropped [%][c]	Mapped reads[d]	Mapping efficiency [%][e]	BLAT matches[f]	Sum trimmed reads[g]	Sum mapped reads[h]	Averge mapping efficiency [%][i]	Targets with at least one read in rep group[j]	Representation of RefBeet-1.2.2 (28 721 transcripts) [%][k]
mock	40	replica 1	311	34,383,904	32,305,179	6.05	24,222,536	75.0	18,512	65,166,053	48,847,041	75.0	19,051	66.3
mock	40	replica 2	313	33,503,347	32,860,874	1.92	24,624,505	74.9	18,459					
ABA	40	replica 1	312	29,615,024	28,147,275	4.96	20,962,972	74.5	18,394	58,265,045	43,172,894	74.1	18,967	66.0
ABA	40	replica 2	314	31,257,959	30,117,770	3.65	22,209,922	73.7	18,460					
mock	80	replica 1	315	30,104,699	29,684,734	1.4	22,088,629	74.4	18,541	65,749,978	48,935,208	74.4	19,102	66.5
mock	80	replica 2	317	37,039,274	36,065,244	2.63	26,846,579	74.4	18,584					
ABA	80	replica 1	316	38,777,553	32,365,938	16.53	23,722,846	73.3	18,557	61,294,259	44,964,735	73.4	19,040	66.3
ABA	80	replica 2	318	35,883,841	28,928,321	19.38	21,241,889	73.4	18,389					
mock	160	replica 1	319	24,474,812	23,396,038	4.41	16,577,222	70.9	18,594	50,713,371	36,132,981	71.2	19,118	66.6
mock	160	replica 2	321	27,689,677	27,317,333	1.34	19,555,759	71.6	18,600					
ABA	160	replica 1	320	35,103,812	33,736,335	3.9	24,447,415	72.5	18,792	59,060,439	42,437,497	71.8	19,161	66.7
ABA	160	replica 2	322	26,135,742	25,324,104	3.11	17,990,082	71.0	18,523					
mock	320	replica 1	323	32,627,470	32,088,658	1.65	23,101,642	72.0	18,818	62,903,234	45,300,352	72.0	19,222	66.9
mock	320	replica 2	325	31,784,966	30,814,576	3.05	22,198,710	72.0	18,686					
ABA	320	replica 1	324	36,955,093	36,294,440	1.79	24,948,298	68.7	18,933	65,247,592	44,841,150	68.7	19,277	67.1
ABA	320	replica 2	326	29,892,017	28,953,152	3.14	19,892,852	68.7	18,638					
mock	640	replica 1	327	30,164,205	29,388,774	2.57	19,672,660	66.9	18,921	61,153,343	40,403,656	66.1	19,422	67.6
mock	640	replica 2	329	32,216,852	31,764,569	1.4	20,730,996	65.3	19,030					
ABA	640	replica 1	328	32,620,324	31,108,522	4.63	19,487,377	62.6	18,904	63,445,704	40,060,498	63.1	19,408	67.6
ABA	640	replica 2	330	33,811,657	32,337,182	4.36	20,573,121	63.6	19,008					
mock	1280	replica 1	331	31,813,716	29,662,032	6.76	20,542,923	69.3	18,434	60,260,778	41,145,401	68.3	18,989	66.1
mock	1280	replica 2	333	31,647,609	30,598,746	3.31	20,602,478	67.3	18,456					
ABA	1280	replica 1	332	35,091,218	34,587,186	1.44	21,849,517	63.2	18,801	63,161,766	39,918,792	63.2	19,289	67.2
ABA	1280	replica 2	334	29,550,632	28,574,580	3.3	18,069,275	63.2	18,731					
Average				**32,172,725**	**30,684,232**	**4.45**	**21,506,675**	**70.1**	**18,657**	**61,368,464**	**43,013,350**	**70.1**	**19,171**	**66.7**

Table S3. Genes involved in clock regulation and analysis of time-dependent abundances of their transcript levels in *Talinum triangulare*

[a]target_id, transcript identifier of Beta vulgaris reference (RefBeet-1.2.2)
[b]AGI, Arabidopsis gene identifier for the closest hit to the Beta vulgaris transcript (based on BlastX)
[c]Araport11 annotation, gene product annotation based on Arabidopsis
[d]log2-FC pool(ABA/mock), log2-FC of transcript abundances in all ABA-treated leaves (pool of all samplings) compared to transcript abundances in all mock-treated leaves
[e]q-value, Benjamini-Hochberg corrected p-value of DESeq2 analysis of differential expression (column d); only DEGs with $q < 0.01$ are included
[f]changed temporal pattern, result of time-series analysis (maSigPro analysis as described in Supplemental Note 3)
[g]Cluster, Assignment to a cluster based on transcriptional contribution of the five identifed factors to expression pattern of a gene (see Materials and Methods for details)

target Id[a]	AGI[b]	annotation[c]	log_2-FC pool (ABA/mock)[d]	*q*-value[e]	changed temporal pattern[f]	cluster (NMF)[g]
KMT10771	AT5G61380	TOC1	n.d.	n.d.	yes	1
KMT07824	AT1G12910	ATAN11	n.d.	n.d.	yes	1
KMT16167	AT5G52660	AT5G52660	n.d.	n.d.	yes	1
KMT05202	AT5G59570	BOA	n.d.	n.d.	yes	2
KMT18207	AT2G40080	ELF4	n.d.	n.d.	n.s.	2
KMS99319	AT5G02810	PRR7	n.d.	n.d.	n.s.	3
KMT07412	AT3G27010	TCP20	1.51	0.006	n.s.	4
KMT17957	AT2G25930	ELF3	n.d.	n.d.	yes	4
KMS95263	AT3G22380	TIC	n.d.	n.d.	yes	4
KMT01429	AT2G21070	FIO1	n.d.	n.d.	n.s.	4
KMT01453	AT2G21150	XCT	n.d.	n.d.	yes	4
KMT20192	AT1G01060	LHY	n.d.	n.d.	yes	5
KMT01479	AT5G64170	AT5G64170	n.d.	n.d.	n.s.	5
KMS97011	AT3G09600	RVE8	n.d.	n.d.	n.s.	5
KMS94973	AT1G17455	ELF4-L4	n.d.	n.d.	yes	7
KMT16488	AT5G24470	PRR5	n.d.	n.d.	n.s.	8
KMT20206	AT2G46790	PRR9	n.d.	n.d.	n.s.	9
KMT17076	AT3G54500	AT3G54500	n.d.	n.d.	n.s.	9
n.d.	AT2G46830	CCA1	n.d.	n.d.	n.s.	n.d.
n.d.	AT5G60100	PRR3	n.d.	n.d.	n.s.	n.d.
n.d.	AT3G26640	LWD2	n.d.	n.d.	n.s.	n.d.
KMT06287	AT5G08330	TCP11	n.d.	n.d.	n.s.	n.d.
KMT02583	AT5G23280	AT5G23280	n.d.	n.d.	n.s.	n.d.
n.d.	AT5G43630	TZP	n.d.	n.d.	n.s.	n.d.
n.d.	AT1G01520	ASG4	n.d.	n.d.	n.s.	n.d.
n.d.	AT4G01280	AT4G01280	n.d.	n.d.	n.s.	n.d.
n.d.	AT5G02840	LCL1	n.d.	n.d.	n.s.	n.d.
KMT00177	AT3G46640	PCL1	n.d.	n.d.	n.s.	n.d.
KMT01722	AT3G21320	AT3G21320	n.d.	n.d.	n.s.	n.d.
n.d.	AT1G72630	ELF4-L2	n.d.	n.d.	n.s.	n.d.
n.d.	AT2G06255	ELF4-L3	n.d.	n.d.	n.s.	n.d.
n.d.	AT2G29950	ELF4-L1	n.d.	n.d.	n.s.	n.d.
n.d.	AT3G63180	TKL	n.d.	n.d.	n.s.	n.d.

Supplementary protocols

Protocol S1. Validation of transcript levels determined by RNA-seq

Protocol S2. Time series RNA-seq analysis

Protocol S3. Pulse-amplitude modulated (PAM) fluorescence measurement

Supplementary Protocol S1: Validation of transcript levels determined by RNA-seq

Quantitative real-time PCR (qPCR) was used to quantify transcript levels of selected targets, and thus validate the accuracy of RNA-sequencing. In total, four targets, encoding for components of both CAM pathway and ABA signalling, were analysed.

For qPCR confirmation, the same RNA samples of *Talinum triangulare* were used as for the RNA-sequencing. Synthesis of complementary DNA (cDNA) was carried out using SuperScript™ II Reverse Transcriptase (Thermo Fisher Scientific) according to the manufacturer's instructions. All primer pairs were designed using Primer-BLAST (Ye *et al.*, 2012) and targeting at contigs of currently available assembly of *Talinum triangulare* transcriptome (Brilhaus *et al.*, 2016). Contigs of the transcriptome assembly were annotated based on their homology to Arabidopsis transcripts (Tab. 1)

Luna® Universal qPCR Master Mix (New England Biolabs Inc.) chemistry was used for qPCR and the amplification was monitored in StepOne™ Real-Time PCR System (Applied Biosystems™). The reaction composition was scaled down for a total volume of 10 µl and each reaction contained 1.5 µl cDNA, which was previously diluted 1:10. The PCR cycling conditions were: 95 °C for 60 s, 40 cycles of 95 °C for 15 s and 60 °C for 30 s. The melting curve was measured after 40 cycles to verify primer specificity. For each sample and primer pair, analysis was performed in three technical replicates and average C_T values were used to calculate mean normalized expression (MNE) as described by Simon (2003). Transcript levels of *ATPase, F0/V0 complex, subunit C protein* (AGI: AT2G25610, *Talinum triangulare* contig Tt50206_2) were used for normalization.

To evaluate correlation between RNA-seq and qPCR, RNA-seq reads were additionally quantified against the reference transcriptome using Kallisto (Bray *et al.*, 2016) and the degree of correlation between MNE and transcripts per million (tpm) was evaluated by Pearson's correlation.

Table 1. Primer sequences used to amplify selected targets in *Talinum triangulare* by qPCR and annotation of the targets. *ATPase, F0/V0 complex, subunit C* was used as a house-keeping gene.

Arabidopsis gene	***T. triangulare* contig**	**Forward primer (5'-3')** **Reverse primer (5'-3')**
ABF2 (AGI: AT1G45249)	Tt42370	TAGGGATGGAGTCGTTGGGT CATTCCTCCACCTGGCAACA
ATPase, F0/V0 complex, subunit C (AGI: AT2G25610)	Tt50206_2	GTGTCCCATCGTCCCAAGTT CGCAGACAAGGTTGGCAAAA
PPC1 (AGI: AT1G53310)	Tt63271	TTGGCAACTCTACAAGGCCC TTCACCCCGAATTGCTTGGA
PYL9 (AGI: AT1G01360)	Tt22302	CTACCAGCACGGAGAGGCTA AGGGACACGACGGAAGAGTA
SnRK2.2 (AGI: AT3G50500)	Tt31423_2	AGGCGATGAATCGGGAACAG TGTCGTGCATGATCGGCATA

Results

qPCR was performed and mean normalized expression (MNE) was calculated and correlated with quantification of respective contigs based on RNA-sequencing. For *PPC1, PYL9* and *SNRK2.2*, the correlation coefficient was 0.998, 0.705 and 0.91, respectively. In all three cases, the correlation was significant at the 0.01 significance level. For *ABF2*, the correlation coefficient was 0.615 with *p*-value of 0.03 (Fig. 1).

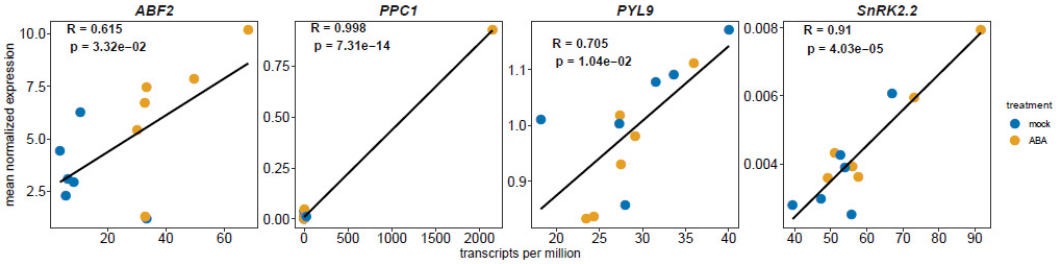


Figure 1. Correlation between mean normalized expression and transcript abundance [tpm] in leave tissue of *Talinum triangulare*. p indicates *p*-value. R stands for correlation coefficient.

References for Supplementary Protocol 1

Brilhaus D, Bräutigam A, Mettler-Altmann T, Winter K, Weber APM. 2016. Reversible Burst of Transcriptional Changes During Induction of Crassulacean Acid Metabolism (CAM) in Talinum triangulare. Plant Physiology **170**, 102–122.

Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. Nature Biotechnology **34**, 525–527.

Simon P. 2003. Q-Gene: Processing quantitative real-time RT-PCR data. Bioinformatics **19**, 1439–1440.

Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden T. 2012. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics **13**, 134-145.

Supplementary Protocol S2: Time series RNA-seq analysis

The maSigPro package (version 1.54; Nueda *et al.*, 2014) for R was used to identify genes with significantly different temporal patterns between mock and ABA treatments. Obtained results were used to further support the analysis of differential genes expression at individual time points.

Talinum triangulare reads mapped against *Beta vulgaris* reference transcriptome RefBeet-1.2.2 (Dohm *et al.*, 2014) were normalized to reads per million (rpm) and low abundant genes were filtered prior to the time-course analysis. Different temporal patterns were detected by generalized linear model (GLM) with the following settings: $\theta = 10$ (default), false discovery rate Q = 0.01 and BH adjustment for multiple testing. Profiling of significant differences between the experimental groups was done using `"two.ways.backward"` step method. Additional parameters of the `t.fit()` function included the minimum of 10 true numerical observation per gene and $\alpha = 0.05$. To select temporally differentially expressed genes, the results of the second regression model were filtered using the `"group"` comparison (`get.siggenes()` function) and setting the R^2 cut-off value to 0.85 for the regression model.

Subsequently, hierarchical clustering (`hclust`) was performed to group temporally differentially expressed genes based on similarity in their transcript abundance profiles. The number of required clusters was set to six, which was determined as an optimum by within-cluster sum of squares. Finally, genes within each cluster were tested for enrichment of biological functions based on MapMan categorisation (Thimm *et al.*, 2004). For the enrichment analysis, only transcripts with known Arabidopsis homologs were considered and Arabidopsis-based functional annotations were used. Genes, whose temporal pattern did not differ significantly between mock and ABA, were considered as a separate cluster in the enrichment analysis. Fisher's exact test with *p*-value adjustment according to Benjamini-Hochberg (Benjamini and Hochberg, 1995) was used to find significant enrichments.

Results

With the above specified parameters, temporal patterns significantly different between mock and ABA treatments were discovered for 4,765 genes. The complete list of all genes with significantly altered temporal pattern is available as Supplemental Dataset 2.

References for Supplementary Protocol 2

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society B **57**, 289–300.

Dohm JC, Minoche AE, Holtgrawe D, Capella-Gutierrez S, Zakrzewski F, Tafer H, Rupp O, Sorensen TR, Stracke R, Reinhardt R, Goesmann A, Kraft T, Schulz B, Stadler PF, Schmidt T, Gabaldon T, Lehrach H, Weisshaar B, Himmelbauer H. 2014. The genome of the recently domesticated crop plant sugar beet *(Beta vulgaris)*. Nature **505**, 546-549.

Kent WJ. 2002. BLAT—The BLAST-Like Alignment Tool. Genome Research **12**, 656–664.

Nueda MJ, Tarazona S, Conesa A. 2014. Next maSigPro: Updating maSigPro bioconductor package for RNA-seq time series. Bioinformatics **30**, 2598–2602.

Thimm O, Bla O, Gibon Y, Nagel A, Meyer S, Kru P, Selbig J, Mu LA, Rhee SY, Stitt M. 2004. MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. The Plant Journal **37**, 914–939.

Supplementary Protocol S3: Pulse-amplitude modulated (PAM) fluorescence measurements

To investigate whether exogenous ABA induced stress response in *Talinum triangulare* at a physiological level, chlorophyll fluorescence was analysed. Chlorophyll fluorescence ratio F_v/F_m is a robust indicator of the maximum quantum yield of photosystem II. Under stress conditions, the capacity to repair photosystem II is reduced, which translates in the irreversible inhibition of photosystem II, detectable as reduced F_v/F_m (Baker 2008; Murchie and Lawson, 2013).

Two independent experiments were performed, one to investigate a long-term effect of exogenous ABA on photosynthetic performance/stress in *T. triangulare*, and another one to accompany the RNA-seq experiment and focusing on the early response to exogenous ABA. In both experiments, plants were grown under 12 h light/ 12 h dark and 25 °C/ 23°C conditions.

In the long-term experiment, mature leaves of *T. triangulare* were treated daily with 10 µM, 50 µM, 200 µM ABA or mock solution in four biological replicates; non-treated controls were included as well. All treatments were done on a daily basis within the first hour of the light period. F_v/F_m was measured within the last 30 minutes of dark *(i.e.,* in dark-adapted state) using a JUNIOR PAM (Heinz Walz GmbH) device. The same leaves were used during the whole experiment.

For the early response experiment, plants independent of the RNA-seq experiment were used but originating from the same seed batch and grown under identical conditions, and experimental design as used in the RNA-seq experiment (Fig. 1 of the main text). For each time point and treatment, two biological replicates were used and four leaf-discs originating from the treated leaves were analysed per plant. Immediately after cutting, leaf discs were placed in water and incubated in dark for 15 minutes prior to the measurement with an IMAGING-PAM *M-Series* (MAXI version) connected to an IMAG-K6 camera (Heinz Walz GmbH).

Results

Daily ABA treatments (concentration range 10-200 µM) of mature leaves of *T. triangulare* for up to six days did not have any adverse effect on the photosynthetic performance of the treated leaves (Fig. 1). Beginning on day 11, statistically significant decline of F_v/F_m occurred but it could also be a result of growth conditions, as a decline was observed also in mock-treated leaves (day 11, 50 µM ABA).

Consistent with the findings of the long-term experiment, exogenous ABA did not induce damage to photosystem II within 1280 min after the treatment (Fig. 2). The drop of Fv/Fm to [illegible] and ABA-treated leaves at 640 min, respectively, is likely a result of diurnal changes to the photosystem and not caused by the treatment (Fig. 2).

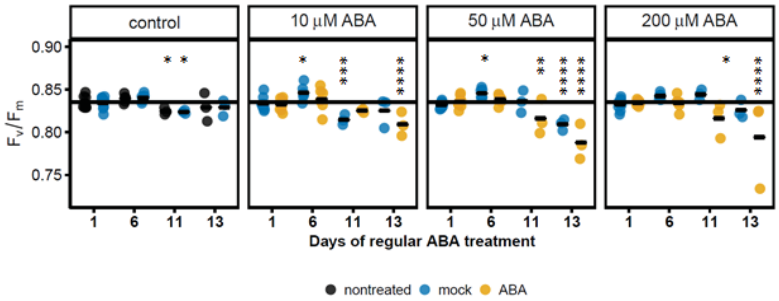


Figure 1. Effect of daily ABA treatments on F_v/F_m of mature leaves of *Talinum triangulare*. Non-treated, mock- and ABA-treated leaves were analysed (n = 4). Treatments were done daily within the first hour of the light. F_v/F_m measurements were done shortly before dawn, while the plants were still dark-adapted. The horizontal line depicts a reference F_v/F_m value (0.835), which is based on measurements of all leaves chosen for this experiment prior to the first treatment. Asterisks indicate difference compared to the reference F_v/F_m value. One sample *t*-test, **** [illegible]

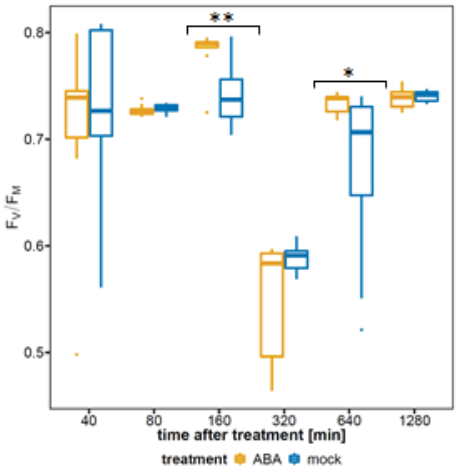


Figure 2. Short-term effect of exogenous ABA on F_v/F_m of mature *Talinum triangulare* leaves. The treatment and harvest schemes were the same as in the RNA-seq experiment (see Materials and Methods). Two biological replicates were used and four leaf-discs per plant were obtained and analysed. After cutting, leaf-discs were kept in water and dark adapted prior to the measurement. Two sample *t*-[illegible]
* [illegible]

References for Supplementary Protocol 3

Baker NR. 2008. Chlorophyll Fluorescence: A Probe of Photosynthesis In Vivo. Annual Review of Plant Biology **59**, 89–113.

Murchie EH, Lawson T. 2013. Chlorophyll fluorescence analysis: A guide to good practice and understanding some new applications. Journal of Experimental Botany **64**, 3983–3998.

Manuscript II

Maleckova, E., Brilhaus, D. & Weber, A. P. M. The genome of *Talinum triangulare* provides insights into regulation of crassulacean acid metabolism *(in preparation)*

Own contribution: Creation of homozygous population, part of gDNA extraction, genome assembly, tissue sampling and RNA extraction, transcriptome assembly, all genome and transcriptome analyses, graphical presentation, preparation of manuscript

The genome of *Talinum triangulare* provides insights into regulation of crassulacean acid metabolism

Eva Maleckova[1], Dominik Brilhaus[1], Andreas P. M. Weber[1]

[1] Institute of Plant Biochemistry, Cluster of Excellence on Plant Sciences (CEPLAS), Heinrich Heine University, Universitätsstraße 1, 40225 Düsseldorf, Germany

Abstract

Crassulacean acid metabolism (CAM) has evolved as the most water-use efficient mode of photosynthesis. CAM engineering was therefore proposed as one strategy to improve drought resilience in crops. This requires a detailed understanding of the genetic blueprints enabling CAM emergence and proper regulation of the underlying protein activities. To identify such CAM-enabling signatures, we generated genome and transcriptome assemblies of the facultative CAM species *Talinum triangulare*. Promoter analysis of genes showing altered expression pattern during abscisic acid-mediated CAM induction revealed enrichment of transcription factor binding sites among genes with similar expression profiles. These included a clock-associated motif, a motif bound by transcription factors responsive to abscisic acid and several predicted motifs of TEOSINTE BRANCHED 1, CYCLOIDEA AND PCF TRANSCRIPTION FACTOR (TCP) family. Majority of enriched motifs were identified in genes down-regulated in response to abscisic acid, suggesting importance of down-regulation during CAM induction. Interspecies comparison of promoter sequences identified motifs enriched in *T. triangulare* promoters of core CAM genes including *PHOSPHOENOLPYRUVATE CARBOXYLASE (PPC), PHOSPHOENOLPYRUVATE CARBOXYLASE KINASE* and *MALATE DEHYDROGENASE* as well as genes encoding several transporters and enzymes of glycolysis and starch metabolism. Protein-coding sequences were analysed as well, focusing on amino acid substitutions associated with C_4 and CAM PHOSPHO*ENOL*PYRUVATE CARBOXYLASE (PEPC) isoforms. A high degree of diversification among *Talinum triangulare* isoforms as well as across plant lineages and photosynthesis modes was revealed.

Keywords: *cis*-elements, *cis* regulation, crassulacean acid metabolism, genome, *Talinum triangulare*, transcriptome

Abbreviations: CAM, crassulacean acid metabolism; CRE, *cis*-regulatory element; DEG, differentially expressed gene; lnRNAs, long non-coding RNAs; ORF, open reading frame; PEPC/*PPC*, phospho*enol*pyruvate carboxylase; TF, transcription factor; TFBS, transcription factor binding site(s)

Introduction

Crassulacean acid metabolism (CAM) has evolved as one of the carbon-concentrating mechanisms in terrestrial plants and also the most water-use efficient photosynthetic strategy – some CAM species require as little as one sixth of the water budget of C_3 species (Yang *et al.*, 2015*a*). The key towards the reduced water demands as compared to C_3 and C_4 species is the rescheduling of stomatal opening to the dark, when transpiration rates are low, and temporal separation of primary and secondary carbon assimilation. In brief, the CAM cycle operates in four phases. In the dark, stomata open and atmospheric CO_2 is fixed in form of HCO_3^- by PHOSPHO*ENOL*PYRUVATE CARBOXYLASE (PEPC) to oxaloacetate, which is immediately reduced to malate by MALATE DEHYDROGENASE (MDH) and stored in the vacuole as malic acid (Phase I). During the light, stomata close and nocturnally accumulated malate is decarboxylated enzymatically to provide CO_2 for the secondary fixation catalysed by RIBULOSE-1,5-BISPHOSPHATE CARBOXYLASE/OXIDASE (RuBisCO) (Phase III). At dawn and dusk, two transitory phases occur. At the dark/light transition, stomata can open widely when switching from PEPC- to RuBisCO-mediated carbon fixation (Phase II). Under favourable conditions, stomata may re-open already towards the end of the light period to enable CO_2 assimilation by RuBisCO as malate reserves get depleted (Phase IV) (Osmond, 1978; Lüttge, 2002; Borland *et al.*, 2009).

CAM is a highly plastic adaptation with only few species relying exclusively on carbon assimilation via the CAM pathway. One example of this plasticity is presented by facultative CAM plants, which mostly rely on C_3 assimilation and only upon exposure to stress switch to the CAM mode. Unlike ontogenetically pre-programmed C_3 to CAM transition (*e.g. Ananas comosus* pineapple*, Phalaenopsis equestris* and *Kalanchoë fedtschenkoi*), facultative CAM is reversible upon stress removal (Winter and Holtum, 2014; Wai *et al.*, 2019; Winter, 2019). *Mesembryanthemum crystallinum* has been used as a model species for a long time, but more cases of facultative CAM have been described with more of the plant diversity being explored. Other facultative CAM species include *Talinum triangulare* and several C_4-CAM *Portulaca* species. Facultative CAM also evolved in monocots (*e.g.* orchid *Dendrobium catenatum*) and in a fern *Vittaria lineata* (Minardi *et al.*, 2014; Zhang *et al.*, 2016).

The flexibility between fast growth and inducible stress resilience, makes facultative CAM optimal for introduction into crop plants. Successful CAM engineering efforts however require detailed understanding of the regulatory mechanisms controlling CAM induction and how re-programming of carbon assimilation is synchronized with other metabolic processes (Borland *et al.*, 2014; Yang *et al.*, 2015*a*; Liu *et al.*, 2018). Besides offering a promising CAM biodesign strategy, facultative CAM species are an excellent system to identify i) novel genes associated with the CAM pathway and ii) CAM-related genetic changes in the case of genes orthologous to C_3 and C_4 isogenes (Yang *et al.*, 2019). The genetic changes leading to CAM are expected to be of two kinds: i) amino acid substitutions affecting kinetic, regulatory and binding properties of target proteins, and ii) altered timing of gene expression, possibly accompanied by changed maximal abundance (Cushman and Bohnert, 1999; Cushman *et al.*, 2008; Silvera *et al.*, 2010).

Due to the limited number of genomic resources, the knowledge about exact CAM-enabling changes has been limited. This is however changing at rapid pace and more available genome sequences

covering the CAM diversity will facilitate the study of convergent CAM evolution. For example, the role of whole genome and tandem gene duplications followed by neofunctionalization in CAM evolution remains to be clarified. In pineapple, the hypothesis of gene duplications contributing to CAM evolution was rejected, while analysis of the *K. fedtschenkoi* genome provided strong evidence for whole genome duplications and identified several CAM genes with high copy numbers (*e.g. PPC* and *MDH*) (Ming *et al.*, 2015; Yang *et al.*, 2017).

Evidence of CAM-enabling adjustment of gene expression was obtained at the level of both *cis*- and *trans*-regulation. Promoters of genes expressed diurnally in green leaf tissue, including core CAM genes, contain clock-associated *cis*-regulatory elements (CREs) at frequencies differing from their random occurrence in the genome (Ming *et al.*, 2015, 2016). Further, the altered frequency of these elements was confirmed in a comparison to C_3 and C_4 orthologs (Ming *et al.*, 2016). Increased expression of several CAM genes, including *PPC* was attributed to both loss of repressors and gain of activators in pineapple (Ming *et al.*, 2015). In *K. fedtschenkoi*, Moseley *et al.* (2019) identified candidate genes extending the transcriptional feedback loops of the circadian oscillator as known in Arabidopsis. Similar approaches need to be applied on facultative CAM species to identify whether *trans*-acting "CAM switches" evolved that drive the metabolic re-programming.

Long non-coding RNAs (lncRNAs) and short regulatory RNAs such as microRNAs (miRNAs) provide an additional layer of transcriptional control both in plants and animals (Nelson *et al.*, 2003; Lee, 2012; Heo *et al.*, 2013). In pineapple, both lncRNAs and miRNAs with strong diurnal expression pattern were identified as well as their putative targets, including CAM genes (Wai *et al.*, 2017; Bai *et al.*, 2019). *M. crystallinum* is currently the only facultative CAM species, for which a collection of miRNAs was generated but studied in the context of resistance to high salinity only (Chiang *et al.*, 2016). It remains to be researched, whether RNA-mediated gene expression is crucial for effective CAM expression or whether non-coding RNAs provide an additional layer for fine-tuning the gene expression.

T. triangulare, together with *M. crystallinum* and *Portulaca* spp., belongs to the order of *Caryophyllales*, which is a true hotspot of multiple independent CAM origins, including obligate CAM *Cactaceae* and *Crassulaceae*. Besides that, more than 20 independent origins of C_4 photosynthesis were identified within *Caryophyllales* (Silvera *et al.*, 2010; Sage *et al.*, 2011; Christin *et al.*, 2014). *T. triangulare* is the experimental model of our choice because of fast and reversible CAM induction in response to stress (Brilhaus *et al.*, 2016; Montero *et al.*, 2018). In addition, exogenous abscisic acid (ABA) can be used to induce the CAM pathway in a controlled and very rapid way (Maleckova *et al.*, 2019). The aim of our work was to expand the genomic resources available for *T. triangulare* and to identify CAM-enabling signatures in this species. Besides the first *Talinum* genome assembly, we generated an improved transcriptome assembly, including putative lncRNAs. Analysis of promoter sequences of genes differentially expressed during ABA-mediated CAM induction revealed transcription factor binding sites (TFBS) associated with certain expression profiles.

Materials and Methods

Plant material and growth conditions

Seeds of *Talinum triangulare* obtained after five subsequent, controlled self-pollination events were germinated in D 400 soil with Cocopor (Stender) and grown in a controlled-environment plant chamber (MobyLux GroBanks, CLF Plant Climatics) under the following conditions: 12 h light/12 h dark at 25 °C/23 °C. The light intensity at the leaf level was 150–200 µmol s^{-1} m^{-2}. In parallel, a batch of the same seed stock was sterilised by washing twice in 70% ethanol for 15 min followed by two washes with sterile water (15 min each) and germinated on ½ Murashige and Skoog (MS) medium with 3% sucrose.

For PacBio transcriptome sequencing, the following samples were harvested by snap-freezing respective tissues in liquid nitrogen: seedlings grown on soil harvested 6 to 12 after sowing, roots of plants grown in soil in which case roots were washed in water and rinsed with 70 % ethanol prior to freezing, roots of plants grown *in vitro* on ½ MS medium (25 days after sowing), young leaves grown under control conditions (21 days after sowing), mature leaves harvested 160 min and 640 min after foliar application of 200 µM abscisic acid (ABA) for presence of ABA-inducible transcripts (Maleckova *et al.*, 2019) (40 days after sowing), whole leaves harvested at the end of day and at the end of night after water withdrawal for seven days for presence of drought-responsive and/or CAM-related transcripts (Brilhaus *et al.*, 2016) (40 days after sowing), and whole flowers harvested continuously (53 to 70 days after sowing).

For genomic DNA extraction, seeds from the same batch as for transcriptome sequencing were germinated on soil and grown under the same conditions as described above. At the age of 25 days, plants were kept at constant darkness for 36 hours to remove starch. Afterwards, whole shoots were harvested, snap-frozen in liquid nitrogen immediately and stored at - 80 °C.

Genome size estimation

The genome size was measured by flow cytometry. Leaf tissue of approximately two-month old *T. triangulare* was chopped together with young leaves of *Solanum lycopersicum* L. 'Stupické polní rané' (obtained from Institute of Experimental Botany, Centre of the Region Haná for Biotechnological and Agricultural Research, Olomouc, Czech Republic) which was used as an internal standard. *T. triangulare* was used in excess (approximately 3:1) as determined before measurements. Galbraith's buffer (Galbraith *et al.*, 1983; Doležel *et al.*, 2007) with addition of 15 mM β-mercaptoethanol and 1 % (w/v) PVP-40 was used for nuclei extraction. Nuclei were released free by chopping both leaf tissues in the isolation buffer, obtained nuclei suspension was filtered via miracloth and RNase A treatment followed for 30 min at 37 °C using 5 µl [10 mg/ml] RNase A. Shortly before the measurement, 50 µl propidium iodide [1 mg/ml] were added per 1 ml of sample. Nuclei suspensions prepared in triplicates were measured on two different days using CyAn™ ADP Analyzer (Beckman Coulter) up to a total of 5,000 particles at each measurement. Summit v 4.3.04 software was used for selection of measured events and peak evaluation. The amount of nuclear DNA in *T. triangulare* tissue was calculated based on G_1 peaks using the following formula: Sample 2C DNA content = [(sample G_1 peak mean) / (standard G_1 peak mean)] × standard 2C DNA content [pg DNA] (Doležel and Bartoš, 2005). For genome size in nucleotides, the formula published by Doležel *et al.* (2003) was used.

Genomic DNA extraction and Nanopore sequencing

Approximately 10 g frozen shoots were ground in liquid nitrogen to fine powder which was transferred to 200 ml homogenization buffer (HB) (10 mM Tris-Cl pH 9.5, 80 mM KCl, 100 mM EDTA, 1 mM spermidine, 1 mM spermine, 17.1 % (w/v) sucrose). β-mercaptoethanol was added to a final concentration of 0.15 % (v/v) and the sample was stirred for 10 min on ice. Subsequently, it was filtered through two layers of cheesecloth and one layer of miracloth, 5 ml 20 % Triton X-100 (in HB) were added to the filtrate and stirred gently for additional 20 min on ice. Obtained lysate was centrifuged at 2,500 *g* for 20 min at 4 °C, the pellet was resuspended in 30 ml HB and filtered through two layers of miracloth. Afterwards, two more washing steps without any further filtering were performed.

Purified pelleted nuclei were resuspended in 2 ml G2 buffer (800 mM guanidine hydrochloride, 30 mM EDTA, 30 mM Tris base, 5 % (v/v) Tween 20, 0.5 % (v/v) Triton X-100), 4 µl RNase A [10 mg/ml] were added and incubated at 37 °C for 30 min. Afterwards, 45 µl proteinase K [20 mg/ml] were added and incubated at 50 °C for two hours, while occasionally inverting the tube. After spinning down at 5,000 *g* for 10 min at 4 °C, the supernatant was loaded on equilibrated Genomic-tip 20/G column (QIAGEN) and washed and eluted according to manufacturer's instructions. DNA was precipitated with 0.7 volume of isopropanol and spinned down at 5,000 *g* for 15 min at 4 °C. The pellet was washed two times in 500 µl cold 70 % ethanol, each time followed by centrifugation at 5,000 *g* for 10 min at 4 °C. Upon discarding ethanol from the last washing step, the DNA pellet was air dried, 50 µl 10 mM Tris pH 8.5 were added and DNA was allowed to resuspend by shaking overnight at 300 rpm at 4 °C.

gDNA size and integrity was examined with Fragment Analyzer™ (Advanced Analytical Technologies GmbH) and quantified with dsDNA HS Qubit assay, in both cases following the manufacturer's protocols. Libraries were prepared from 1.5 µg gDNA with SQK-LSK109 1D Ligation Sequencing Kit (Oxford Nanopore Technologies) according to the manufacturer's instructions and sequenced on the GridION™ platform. On-instrument base-calling was performed with Guppy v3.0.6 (Oxford Nanopore Technologies).

Total RNA extraction, library preparation and isoform sequencing (Iso-Seq, PacBio)

Total RNA was extracted from various tissues of *T. triangulare* using GeneMATRIX Universal RNA Purification Kit v3.0 (EURx Ltd.) according to the manufacturer's instruction with the following modifications: for soil-grown roots, leaf and flower samples, volumes of buffers added to the ground material were increased to 400 µl LG and 200 µl RL buffer in step 1 and homogenization column activation was omitted for leaf and flower samples.

To eliminate any contaminating gDNA, 1 µg total RNA was treated with 1 U DNase I (New England Biolabs Inc.) at room temperature for 5 minutes, scaling reaction volumes as needed. Subsequently, individual samples were pooled at equal amounts and to satisfy the minimal concentration requirements for SMRTbell™ library preparation, the pooled sample was concentrated up by lyophilisation at -23 °C until the sample volume reduced down to a fifteenth of its initial volume (approximately 1.5 hrs with Alpha 1-2 LDplus (Christ) freeze-dryer). RNA integrity was analysed with a 2100 Bioanalyzer (Agilent) throughout the sample preparation.

Prior to library preparation, input RNA was purified with 1.3 X volume AMPure XP beads (Beckman Coulter) and 900 ng purified total RNA was used as input. First cDNA strand was synthesized with Clontech SMARTer® cDNA kit (TaKaRa) and amplified for 11 cycles with PrimeSTAR® GXL DNA polymerase (TaKaRa). Subsequently, size selection was performed with BluePippin (Sage Science) to enrich for fragments 5-10 kbp in size and libraries for individual fractions were prepared using SMRTbell™ Template Prep Kit 1.0 (Pacific Biosciences) according to the manufacturer's instructions. Prepared libraries were loaded on four SMRT cells (Sequel™ SMRT® Cells LR 3.0, Pacific Biosciences). Total yield as well as zero mode waveguide (ZMW) productivity are summarized in Tab. S4.

Draft genome assembly and quality assessment

In total, 6,168,032 Nanopore raw reads were (Tab. S1) used for *de novo* genome assembly with the Flye v2.7 (Kolmogorov *et al.*, 2019). Draft genome assembly was improved with a subset of RNA-seq reads of *T. triangulare* obtained earlier (Maleckova *et al.*, 2019) using Pilon v1.23 (Fig. 1e; Walker *et al.*, 2014). Afterwards, the assembly was scanned for the presence of contaminating sequences with BLAST+ v2.9.0 (Kent, 2002) against non-redundant nucleotide NCBI database (downloaded on 19.8.2019).

Basic parameters of the genome assembly were obtained with QUAST v5.0.2 (Tab. S2). Assembly completeness was verified by mapping RNA-seq reads (Maleckova *et al.*, 2019) with HISAT2 v2.1.0 (Kim *et al.*, 2015; Tab. S3). Gene-set completeness was determined using Benchmarking Universal Single-Copy Orthologs (BUSCO) v3.0.1 (Simão *et al.*, 2015*a*) with embryophyte_odb10 database of single-copy orthologs.

Reconstruction of full-length unique transcripts

First, raw PacBio reads were used to generate circular consensus sequences (CCS) using ccs v3.4.1 (available as bioconda package pbccs) with the parameters: `--noPolish --minLength=50 --maxLength=15000 --minPasses=1 --minPredictedAccuracy=0.8.` Obtained CCS were trimmed with lima v1.9.0 (available as bioconda package lima) and full-length non-chimeric reads (FLNCs) were obtained by trimming poly(A) regions from the CCS sequences. FLNCs from individual flow cells were merged at this point using dataset create command (dataset v0.2.4; available as package pbcoretools). Running isoseq3 v3.1.2 clustering and polish algorithms, FLNCs were clustered based on sequence similarity and subsequently, polished using all the initial subreads to obtain high quality (HQ) transcripts with predicted accuracy of ≥ 99 % (Fig. 1a-d). HQ transcripts were subsequently examined for presence of contaminating sequences by running BLAST+ v2.9.0 (Kent, 2002) against non-redundant NCBI nucleotide database (downloaded on 19.8.2019) and the results were visualized in MEGAN v6.18.5 (Huson *et al.*, 2016). Significant sequence similarities with other than *Viridiplantae* hits were examined manually to exclude significant hits with potential contaminants.

Subsequently, minimap2 v2.12 was used to map HQ transcripts against the draft genome of *T. triangulare* and the resulting sorted SAM file served as an input for the `collapse_isoforms_by_sam.py` script (included with cDNA Cupcake; https://github.com/Magdoll/cDNA_Cupcake) collapsing redundant HQ transcripts to unique isoforms (Fig. 1f). Finally, `filter_by_count.py` (Cupcake) was applied to keep isoforms with at least two full-

length counts and potentially 5'-degraded isoforms were removed with `filter_away_subset.py` (cDNA Cupcake).

Functional annotation of coding sequences

Protein coding sequences were predicted using ANGEL (available from https://github.com/PacificBiosciences/ANGEL), following the provided tutorial. ORF prediction with the trained ANGEL classifier was done alongside with dumb prediction using the following arguments: and returning only the longest predicted ORF as specified by use of: `--min_angel_aa_length 100 --max_angel_secondORF_distance 10 --min_dumb_aa_length 100 –output_mode=best`.

Obtained ORFs were subjected to nucleotide BLAST search (BLAST+ v2.9.0) against nonredundant NCBI nucleotide database and coding sequences (CDS) of the Ensembl Plants database release 46 (Yates *et al.*, 2020) and protein BLAST search against the Swiss-Prot database downloaded on 27.2.2020 (Consortium, 2019). Completeness of predicted ORFs was characterized using BUSCO v3.0.1 (Simão *et al.*, 2015*a*) in the proteome mode with embryophyte_odb10 database of single-copy orthologs.

OrthoFinder (Emms and Kelly, 2019) was used to compare all predicted ORFs were subsequently to protein sequences of 80 species selected to represent the plant diversity and with focus on sequenced *Caryophyllales* and CAM species (see Tab. S1 for a complete species list and sequence sources). For Phytozome (v12) assemblies, only sequences based on primary transcripts were used and for shotgun transcriptome sequencing projects, entire sequence sets were used. The majority of those was obtained from the 1 KP transcriptomes initiative (Carpenter *et al.*, 2019). *Agave deserti* and *A. tequilana* (Gross *et al.*, 2013), *Mesembryanthemum crystallinum* (Tsukagoshi *et al.*, 2015) and *Erycina pusilla* (Chou *et al.*, 2013) publicly available transcriptome assemblies were first processed with TransDecoder v5.4.0 to predict coding sequences, while incorporating blastp search against the UniProt database as described in the user's instructions. The same procedure was also applied to the *M. crystallinum* sequences obtained from the RNA-seq experiment performed by Tsukagoshi *et al.* (2015). OrthoFinder v2.3.3 (Emms and Kelly, 2019) with default tools at the individual analysis steps (*i.e.* gene tree inference with dendroblast, sequence search with diamond, mafft for multiple sequence alignment and fasttree as tree inference method) was used for coding sequence comparison. Biological functions of predicted proteins were inferred based on the presence of Arabidopsis orthologs in the same orthogroups, using Araport11 annotation (https://www.araport.org). In the case of proteins without any Arabidopsis orthologs, searches against InterPro signature and Pfam domain databases were performed, using InterProScan v5.36-75.0 (Jones *et al.*, 2014) and hmmscan with HMMER v3.1b2 (http://hmmer.org/), respectively.

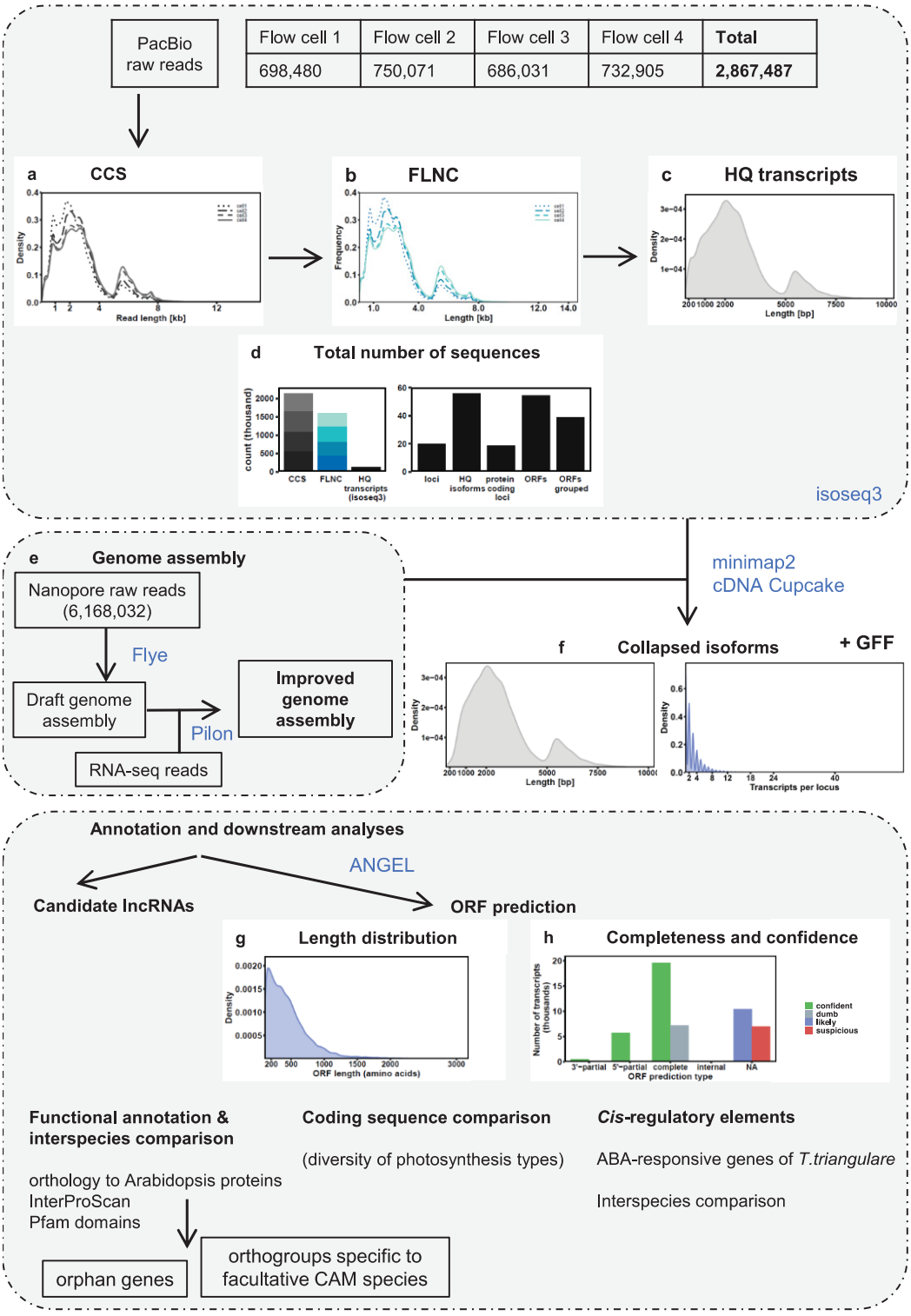


Figure 1. Pipeline leading to transcriptome assembly (Iso-Seq, Pacbio) with the supported of draft genome assembly of *Talinum triangulare.* Nanopore sequencing was employed in whole-genome sequencing and RNA-seq reads were used to correct problematic homopolymeric regions (e). The process of transcriptome assembly included initial read processing with isoseq3 tools, transforming the

Figure 1. (*Continued.*)
raw reads into high-quality (HQ) transcripts (a-d). Upon their alignment to the draft genome assembly, these were collapsed to unique isoforms (f), which were a basis for identification of open reading frames with ANGEL (g-h). Obtained protein-coding sequences were annotated based on their Arabidopsis orthologs and presence of known protein domains (Pfam, InterProScan). The downstream analyses included identification of orthologs specific to CAM species (OrthoFinder) and amino sequence comparison of selected CAM genes. Besides protein-coding sequences, candidate long non-coding RNAs were identified. Enrichment of cis-regulatory elements in promoters of ABA-responsive genes and interspecies comparison of promoter sequences were performed to identify cis-regulatory elements associated with CAM induction. CCS, circular consensus sequence; FLNC, full-length non-chimeric; HQ transcripts, high-quality transcripts; lncRNA, long non-coding RNA; ORF, open reading frame.

Orphan genes and their annotation

Predicted protein sequences of *T. triangulare* with no orthologs in any species included in the OrthoFinder analysis and with no similar sequence in public databases were considered orphan genes, possibly unique to the *Talinum* lineage. HMMER v2.3.2 (available from http://hmmer.org/) search in the Pfam database and InterProScan v5.36-75.0 (Jones *et al.*, 2014) were used to identify known domains in these proteins.

Identification of long non-coding RNAs (lncRNAs)

Isoforms obtained after collapsing HQ transcripts with cDNA Cupcake which were not predicted to contain an ORF according to ANGEL were assessed for their coding potential using Coding-Non-Coding Index (CNCI; Sun *et al.*, 2013), Coding Potential Calculator 2 (CPC2; Kang *et al.*, 2017) and LGC (Wang *et al.*, 2019*c*). Transcripts predicted as non-coding in all three independent analyses were considered candidate lncRNAs and their sequences were screened for homology to CDS and UTR sequences of *T. triangulare*. Both sequence types were obtained with ANGEL.

Transcript abundance and abundance pattern analysis

The effect of exogenous ABA on transcript abundance was considered as an additional criterion during coding sequence comparisons as well as while identifying regulatory motifs in upstream sequences. For this purpose, RNA-seq reads generated previously (Maleckova *et al.*, 2019) were mapped with Kallisto v45.1 against the CDS and lncRNA sequences of the reference transcriptome generated in this work using Kallisto v45.1. The R package sleuth (Pimentel *et al.*, 2017) was used with R v3.6.3 to collect transcript abundances. Temporal patterns of transcript abundance were analysed using the R package lmms (Straube *et al.*, 2015) to identify differentially expressed genes (DEGs). Transcript abundances (transcripts per million) were *z*-scored and clustered based on pattern similarity using *k*-means clustering. The optimal number of clusters was seven as determined with the clValid package (Brock *et al.*, 2008) using the Dunn index.

Time-ordered gene co-expression network

Co-expression was analysed via TO-GCN (Chang *et al.*, 2019) according to the authors' instructions. As an input, DEGs (both TFs and structural genes) as identified with lmms were used. In cases, when multiple transcripts were different splice variants encoded by the same gene, only the most abundant (determined as the sum of abundances at individual time points) transcript was retained.

Transcription factors (TFs) for the analysis were identified based on orthology to Arabidopsis TFs collated in the PlantTFDB (Jin *et al.*, 2017).

Motif analysis in promoter sequences

For genes with multiple splice variants present in the transcriptome assembly, the earliest starting ORF was considered to identify the start codon. Sequences 500 bp upstream of the predicted start codon were extracted for motif analyses. Enriched motifs were identified using Discriminative Regular Expression Motif Elicitation (DREME; Bailey, 2011). Subsequently, Tomtom (Gupta *et al.*, 2007) was employed to compare the enriched motifs to known transcription factor binding sites (TFBS) collated in the Plant Cistrome Database (O'Malley *et al.*, 2016). Both DREME and Tomtom were used as a part of the MEME Suite v5.1.1 (https://www.meme-suite.org; Timothy L. Bailey and William Noble, Copyright 1994-2019 The Regents of the University of California).

When analysing promoters of genes clustered based on similar expression patterns, random 1,590 *T. triangulare* promoter sequences (500 bp upstream of the predicted start codon) were used to estimate motif background frequencies in a set of size comparable to the size of individual clusters. In interspecies comparison, orthologs were identified with OrthoFinder and sequences 500 bp upstream of predicted stop codons were analysed in *T. triangulare* and the following species: Arabidopsis, *Beta vulgaris, Spinacia oleracea*, *Hordeum vulgare*, *Zea mays*, *Ananas comosus*, *Phalaenopsis equestris*, *Kalanchoë fedtschenkoi*, *K. laxiflora*, and *Dendrobium catenatum*..

Amino acid sequence comparison

Orthologous sequences of species employing diverse carbon assimilation strategies (Tab. S1) were aligned via CLUSTAL W implemented in the msa package for R (Thompson *et al.*, 1994; Bodenhofer *et al.*, 2015). In PEPC and PPDK protein sequence comparison, sites homologous to sites subjected to post-translational regulation and associated with C_4 enzymatic properties were examined in predicted ORFs of *T. triangulare*. Translated PEP1 (CAA33317) and PPDK1 (CAA33054) sequences of *Zea mays* were used to localize such sites.

In the above-described way, multiple alignments of all predicted components of the CAM pathway were created and used for metric multidimensional scaling (MDS) analysis available in the bios2mds package for R (Pelé *et al.*, 2012). The resulting primary component matrices (first and second components) served as a basis to visualize sequence similarity of orthologous proteins.

Results

We previously demonstrated CAM inducibility in *T. triangulare* by foliar application of ABA. The reproducibility and fast responsiveness to the ABA stimulus are major advantages for the study of facultative CAM. To identify the changes in transcript abundance relevant for CAM induction, we decided to generate genome and transcriptome assemblies for this species. With these resources, it will be possible not only to gain a deeper understanding in genetic blueprints of *T. triangulare* and how their expression is regulated but it will also be possible to perform comparative studies. Besides generating assemblies and transcriptome annotation, we therefore performed first analyses addressing transcriptional regulation of CAM induction as well as comparative analyses.

Genome and transcriptome assemblies and transcriptome annotation

Genome assembly

Working with a non-model species, its genome size was estimated by flow cytometry prior to proceeding with whole-genome sequencing. Using *Solanum lycopersicum* L. 'Stupické polní rané' as an internal standard, 2C DNA content of *T. triangulare* was estimated to be 3.97 ± 0.02 pg (Fig. S1A). This corresponds to 3.89 Gbp as calculated after Doležel and Bartoš (2005). Approximately 2.3 µg HMW gDNA were used for library preparation. The libraries were sequenced on five GridION^{TM} flow cells yielding 6,168,032 reads passing the quality criteria (Tab. S2). In total, these reads represented 64,783,646,828 bases and their average N_{50} reached 30,317 bp (Tab. S2).

Raw reads were assembled with Flye assembler (Kolmogorov *et al.*, 2019). In combination with a subset of previously generated RNA-seq reads of the same species (Maleckova *et al.*, 2019), the assembled contigs were subsequently improved with Pilon (Walker et al., 2014; Fig. 1e) to better resolve homopolymeric regions in the coding regions. The final assembly of 608.5 Mbp consisted of 234 contigs, with the largest one being over 47.7 kbp (Fig. 2A) and with N_{50} of 21.6 Mbp (Fig. 2B). The average coverage depth reached 111×, with 0.03 N bases per 100 kbp as assessed with QUAST (Gurevich et al., 2013; Tab. S3).

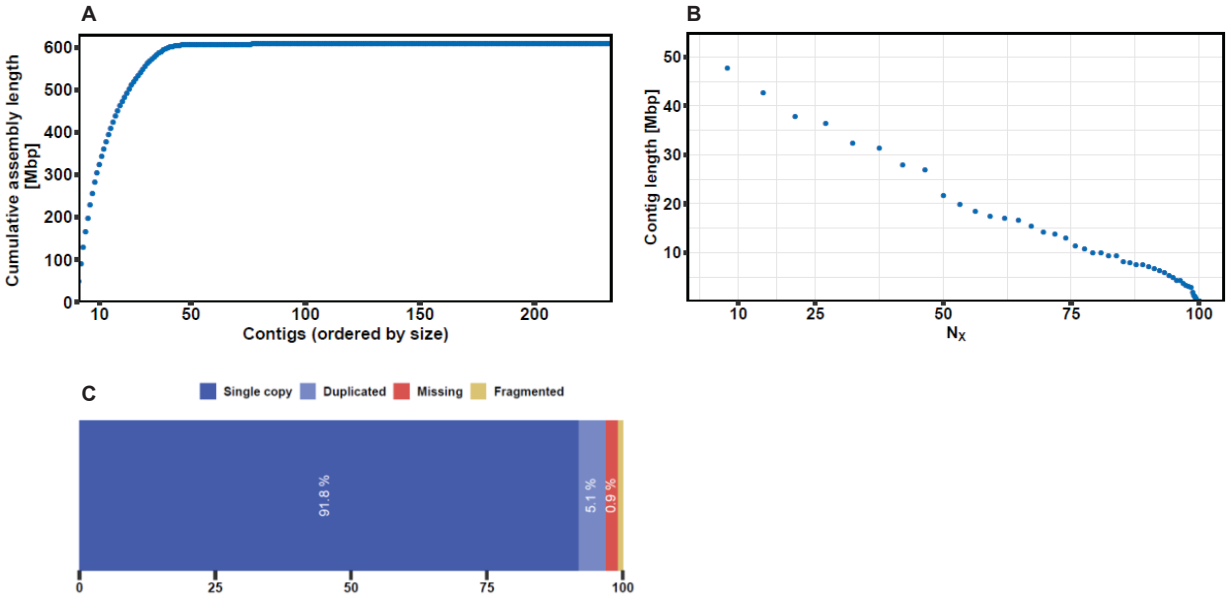


Figure 2. Quality and completeness assessment of the draft genome assembly. (A) The total size of a haploid genome assembly was 608 Mbp and could be reached with the longest 50 contigs. The assembly comprised 234 contigs in total. (B) The assembly was characterized by N_{50} of 21.6 Mbp. (C) Completeness assessment with BUSCO analysis revealed presence of 96.9% of *Embryophyta* single-copy orthologs, with vast majority of them being present in a single copy (91.8% of all identified orthologs).

The assembly's completeness was evaluated by two independent measures. Firstly, RNA-seq reads from our previous work were aligned to the genome assembly with HISAT (Kim *et al.*, 2015), revealing a mean alignment rate of 98.38% (Tab. S4). Secondly, BUSCO analysis (Simão *et al.*, 2015*b*) with 1,614 single-copy orthologs of *Embryophyta* was performed. It indicated presence of 96.9% of the conserved

genes in the assembly, with the vast majority of 91.8% being present as single copies (Fig. 2C). The proportion of fragmented genes remained below 1% and 2.2% of *Embryophyta* orthologs were missing in the genome assembly (Fig. 2C).

Transcriptome assembly

To generate a representative collection of *T. triangulare* transcripts, total RNA was extracted from roots, young leaf tissue, ABA-treated leaves, leaf samples upon water withdrawal and from flower tissue (see Materials and methods for more details). cDNA libraries prepared for PacBio sequencing had an average insert size of 4,144 bp and were sequenced on four flow cells. On average, 716,872 reads were obtained per flow cell and the mean N_{50} polymerase read length reached 80.68 kbp (Tab. S5).

The initial processing steps included trimming of adapter and 3'-poly(A) sequences, merging of FLNC sequences from individual flow cells, clustering of similar transcripts, polishing and examination for contaminating sequences. Following the isoseq3 pipeline, the raw reads were first processed into 2,142,660 circular consensus sequences (CCS) (Tab. S4), ranging from 50 to 14,999 bp in length (Fig. 1a). After adapter trimming, 1,594,167 full-length non-chimeric reads (FLNC) were obtained (Tab. S5), 50 to 40,602 bp long (Fig. 1b). At this point, FLNCs from individual flow cells were merged and clustered, resulting in 127,350 high quality (HQ) transcripts (Fig. 1d), 82 to 10,178 bp in length (Fig. 1c). Next, HQ transcripts were examined for presence of contaminating sequences by running BLAST search against non-redundant NCBI nucleotide database. The vast majority of HQ transcripts (99.999%) showed significant similarity to *Viridiplantae* sequences, while there were no bacterial hits and 17 HQ transcripts aligned to eukaryote entries, primarily *Opisthokonta*. In those cases, the aligning regions did not exceed 30 nucleotides and were therefore not considered as contaminants (Tab. S6).

Obtaining non-redundant transcripts

Subsequently, the HQ transcripts could be aligned to the draft genome with the aim to make the resulting transcriptome non-redundant as well as to identify distinct splice variants present among the HQ transcripts and genes they originated from. After the mapping, cDNA Cupcake was used to collapse HQ transcripts, yielding 66,728 unique isoforms. All of them were supported by at least two FLNC sequences and the default filtering by count support did not result in any losses. However, 10,985 (16.5%) isoforms were potentially 5'-degraded transcripts and were removed to avoid artifacts. The remaining 55,743 isoforms retained in the transcriptome assembly were predicted to result from 19,997 distinct loci (Fig. 1d). Thus, each locus coded for 2.8 transcripts on average (Fig. 1f).

Protein coding transcripts in the transcriptome assembly

As a next step, coding loci and open reading frames they encode for were identified with ANGEL classifier. After training on a non-redundant subset of HQ isoforms, it was run on the entire set of HQ isoforms. Of the 19,997 identified loci, 18,757 (93.8%) were classified as protein-coding. The protein-coding loci were predicted to encode 50,015 distinct transcripts and 54,107 ORFs were predicted for them (Fig. 1d). In other words, each predicted gene gave rise to 2.88 ORFs on average, with a range of amino acid length from 100 to 3,184 (Fig. 2g).

The first measure of the assembly's quality was provided directly by ANGEL, which classifies ORFs based on their confidence and completeness. Of all predicted ORFs, 52.9% were predicted to be complete and originated either from prediction with the trained classifier (tagged as confident; 73% of all complete ORFs) or were a result of "dumb prediction" (*i.e.* the longest ORF in the cases when ANGEL classifier failed to find one) (Fig. 2h). Approximately 20% of predicted ORFs were classified as likely (*i.e.* cases where multiple ORFs were predicted but only the likely one exceeded the minimal length of 100 amino acids) and the remaining 13.8% were suspicious (*i.e.* multiple ORFs exceeding the minimal length were identified within a single transcript) ORFs (Fig. 2h). Since removal of suspicious ORFs resulted in a loss of detectable BUSCO orthologs (Fig. S2), these were retained in the transcriptome assembly as well. ORFs were classified as complete, 5' partial, 3' partial and internal. With nearly 53%, complete ORFs comprised the largest proportion but for over a third of ORFs, the completeness could not be estimated (tagged as NA) (Fig. 2h). Clearly degraded ORFs made up for approximately 1.2% and 11.3% for 3' and 5' degradation, respectively (Fig. 2h).

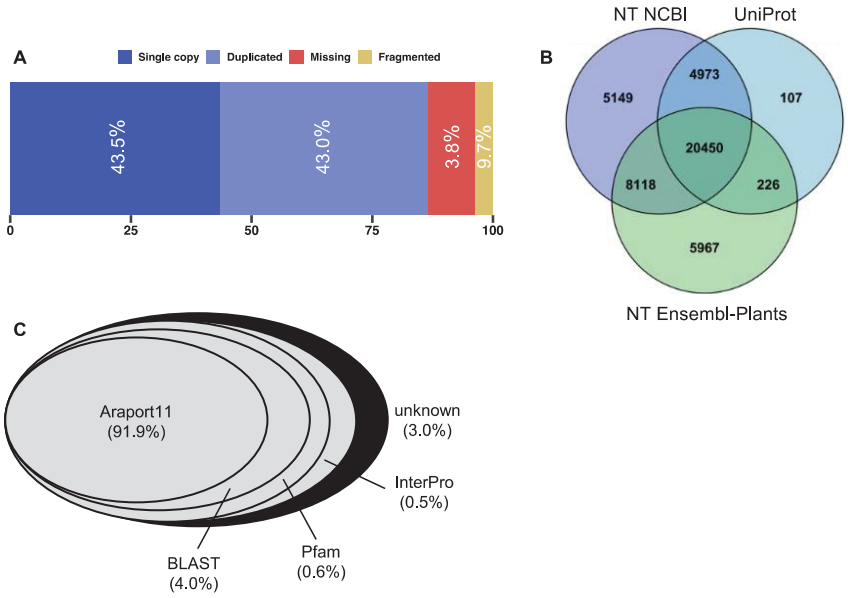


Figure 3. Completeness assessment of the transcriptome assembly. (A) BUSCO analysis identified 86.5% of the expected *Embryophyta* single-copy orthologs, with half of them being duplicated. The proportion of missing and fragmented orthologs was 3.8% and 9.7%, respectively. (B) Predicted open reading frames were used in BLAST search against the protein Swiss-Prot database and nucleotide NCBI and Ensembl-Plants databases. A sequence homologous to 83.2% of predicted open reading frames of *Talinum triangulare* was present in at least one database.

Quality and completeness of the transcriptome assembly were further assessed with i) BUSCO analysis, ii) mapping of previously generated RNA-seq reads, and iii) search in public databases. BUSCO analysis with the *Embryophyta* set of single-copy orthologs revealed presence of 86.5% of conserved genes in the transcriptome assembly of *T. triangulare*, with 43% of all detected genes being present in a duplicated form. Less than 4% of examined genes were classified as fragmented and 9.7% were missing in the transcriptome assembly (Fig. 3A). RNA-seq reads aligned to the predicted CDS sequences with an

average efficiency of 69.1%, but only about 17% mapped uniquely (Tab. S7). Nucleotide BLAST search was performed against nonredundant NCBI nucleotide database and CDS sequences of the Ensembl Plants database with a cut-off E-value ≤ $1e^{-10}$. Protein BLAST search was performed against the Swiss-Prot database with a cut-off E-value ≤ $1e^{-10}$. Most of the coding sequences (44,990; 83.2%) showed significant homology to an entry in at least one database, but only 20,450 CDS sequences (37.8%) returned a significant hit in all three databases (Fig. 3B).

Functional annotation of coding sequences of the transcriptome assembly

To assign the likely function to each identified ORF of *T. triangulare*, translated CDS sequences were compared to proteomes of species representing the diversity of photosynthesizing organisms (OrthoFinder analysis; Tab. S1) and screened for presence of known protein signatures in the Pfam InterPro databases. OrthoFinder accepts proteome sequences and identifies orthogroups based on sequences similarity, thus enabling not only to draw conclusions about likely function of proteins derived from *de novo* assemblies, but also enables evolutionary investigations (Emms and Kelly, 2019). In our case, the primary purpose of OrthoFinder analysis was i) sequence annotation and ii) identification of orthologs for comparison of amino acid sequences. *T. triangulare* ORF sequences were compared to proteomes of 80 species, which included sequences derived from reference genome assemblies, such as of Arabidopsis, *Zea mays* or *Kalanchoe fedtschenkoi* (in which case only the main isoforms were included), as well as translated sequences based on publicly available *de novo* shotgun transcriptome assemblies, which were the source for the majority of analysed CAM and *Caryophyllales* species (Tab. S1).

Reliable inference about orthology relationships depends on adequate species sampling. In the case of *T. triangulare*, 37,132 (95.2%) of the analysed sequences were assigned to orthogroups. Of these, 1,883 translated sequences (3.5% of all predicted ORFs of *T. triangulare*) belonged to 1,282 orthogroups specific to *T. triangulare* among the species analysed (Fig. S3A). The remaining 1,884 sequences (4.8%) were not assigned to any orthogroup (Fig. S3A). This was comparable to results for other proteomes, based both on genome assemblies (Fig. S23-C) and shotgun transcriptome sequencing (Fig. S2A-C). Another quality measure of the OrthoFinder analysis is the resulting species tree. By default, sequence similarity for individual loci (STAG algorithm) is used to infer a species tree and with this approach, *T. triangulare* was placed as a sister species to *Portulaca* spp. within the so called "ACPT" clade of *Caryophyllales* (Fig. 4A).

Based on protein sequence similarity, 26,010 distinct orthogroups were identified, 9,710 (34%) of them with at least one *T. triangulare* ortholog. This was comparable to orthogroup occupancy of majority of other species included in the analysis (Fig. S3F). Closely related C_3 species *B. vulgaris* and *S. oleracea* shared 7,990 (82.3% of all orthogroups with *T. triangulare* orthologs) and 7,876 (81.1%) orthogroups with *T. triangulare*, respectively (Fig. 4B). In the case of well-annotated but less related Arabidopsis, 7,735 orthogroups (79.7%) were shared by the two species (Fig. 4B). The likely function of 49,746 predicted *T. triangulare* proteins orthologous to Arabidopsis could thus be concluded based on Araport11 annotation (91.9% of all predicted ORFs; Fig. 3C). The remaining 4,361 *T. triangulare* CDS without an Arabidopsis ortholog, were subjected to BLAST search in NCBI nucleotide and UniProt databases and searched for known protein domains collated in Pfam and InterPro databases. Combination of approaches led to annotation of 97% of the transcriptome (Fig. 3C).

The

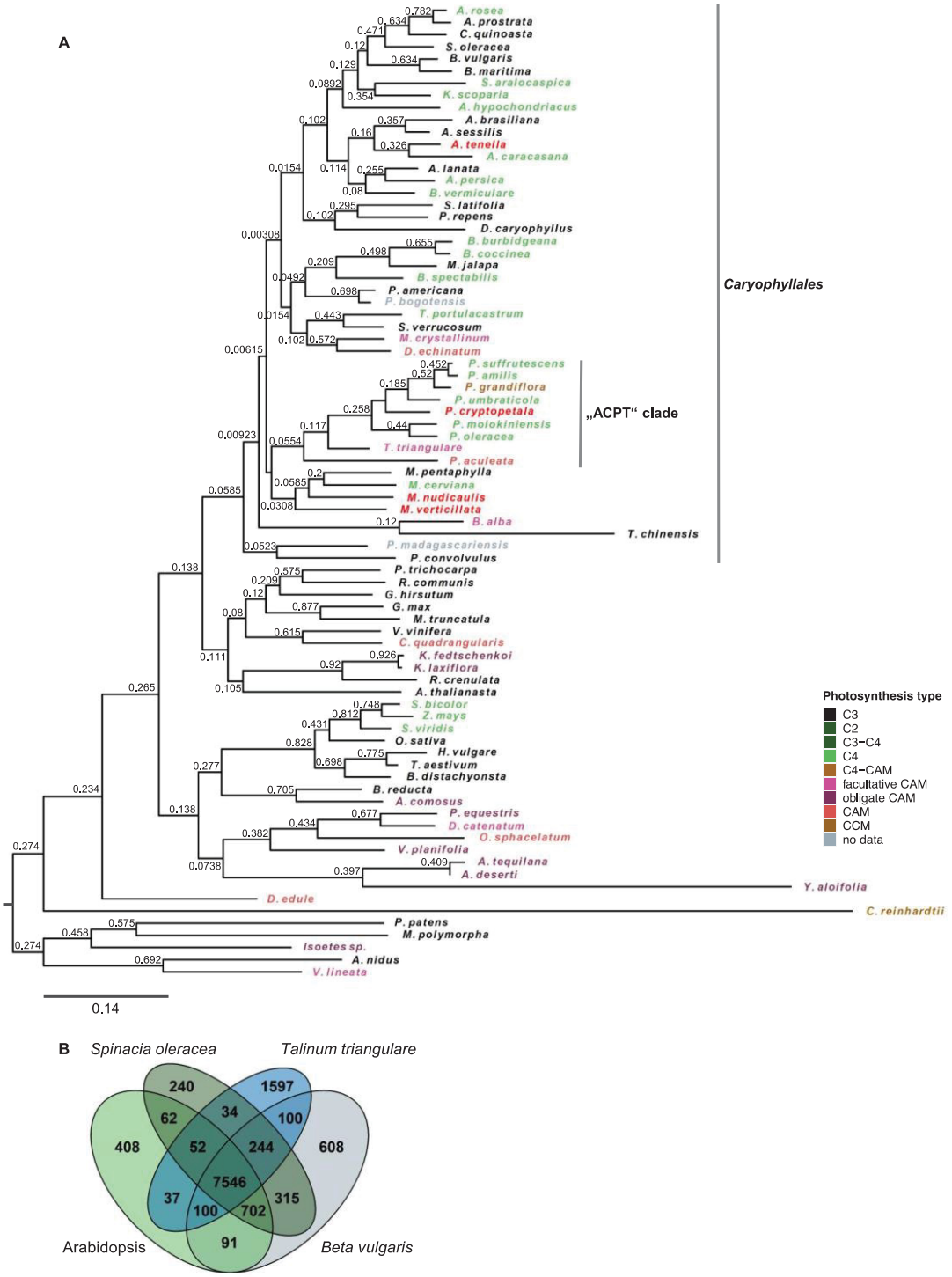


Figure 4. Phylogenetic tree based on proteome-wide sequence comparison, and the proportion of orthogroups shared between *Talinum triangulare* and selected C_3 species (OrthoFinder analysis). (A) The resulting tree is based on single-locus gene trees, with STAG bootstrap values shown. *T. triangulare* was placed as a sister group to *Portulaca* spp. within the “ACPT” clade of *Caryophyllales.*The tree also shows distribution of CAM species among numerous plant families of both dicots and monocots.

Figure 4. (*Continued.*)
(B) *T. triangulare* shared 79.7% of its orthogroups with the C_3 model species Arabidopsis, and 82.3% and 81.1% of orthogroups with more closely related C_3 species *Beta vulgaris* and *Spinacia oleracea*, respectively.

Identification of long non-coding RNAs in the transcriptome assembly

Besides coding sequences, the transcriptome assembly is also a source of non-coding transcripts with possibly regulatory functions. Here, we identified long non-coding RNAs (lncRNAs) based on coding potential analysis and number of exons. First, all protein-coding transcripts as determined by ANGEL were filtered from the assembly, leaving 3,473 HQ isoforms (6.4% of all HQ isoforms) for coding potential assessment. This was achieved by three independent tools: Coding-Non-Coding Index (CNCI) (Sun *et al.*, 2013), Coding Potential Calculator 2 (CPC2) (Kang *et al.*, 2017) and LGC (Wang *et al.*, 2019*c*). CNCI, relying on profiling of adjoin nucleotide triplets to distinguish protein-coding and non-coding sequences, detected 3,356 non-coding transcripts. CPC2 relies on evaluation of sequence intrinsic features (*e.g.* length or pI of putative ORFs) and identified 4,050 non-coding transcripts in the analysed set. Finally, LGC assesses the length of putative ORFs in combination with their GC content and it identified 3,875 non-coding transcripts. Across all the tools, there were 3,184 transcripts classified as non-coding. For a transcript to be considered a lncRNA, additional filtering criteria were employed: minimal length of 200 nucleotides and presence of at least two exons (Tian *et al.*, 2019). In the end, 1,401 putative lncRNAs were retained, which were further characterized in terms of their homology to identified CDS and UTR sequences (both obtained with ANGEL). The vast majority of putative lncRNAs (638; 45.5%) showed significant homology to both CDS and UTR (either 5' or 3') (Fig. 5B). Besides lncRNA homologous to multiple coding regions, there were 101 (7.2%) and 198 (14.1%) lncRNAs specific only to CDS or UTRs, respectively (Fig. 5B).

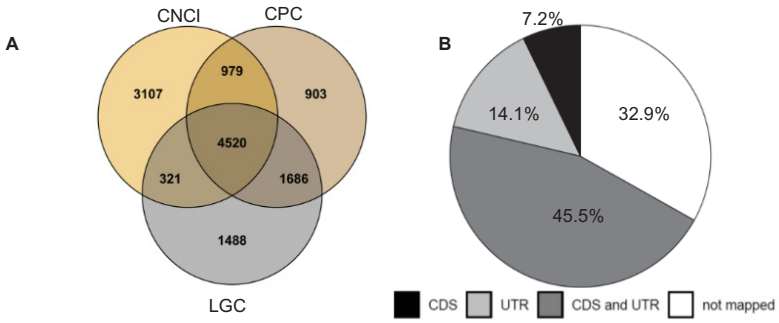


Figure 5. Putative long non-coding RNAs (lncRNA) in the transcriptome of *Talinum triangulare*. (A) Isoforms for which ANGEL did not predict any open reading frames were assessed for their coding potential with Coding-Non-Coding Index (CNCI), Coding Potential Calculator (CPC) and LGC. Of 4,102 transcripts analysed, 3,184 (77.6 %) were predicted as non-coding by all three tools and were used for subsequent filtering based on sequence length and number of exons. (B) Candidate lncRNAs were those non-coding transcripts, which were at least 200 bases long and contained at least two exons. Their sequences were aligned to CDS and UTR sequences of *T. triangulare*, revealing significant homology of the majority (45.5%) of the candidate lncRNAs to both CDS and UTR (either 5' or 3') sequences. lncRNA sequences which could not be aligned (32.9%) were likely derived from non-coding regions of the genome.

Transcriptional regulation of CAM induction

Our previous work revealed the rapid pace of CAM induction in response to exogenous ABA in *T. triangulare*, which was accompanied by extensive changes in transcript abundance (Maleckova *et al.*, 2019). We exploited the additional genomic resources of *T. triangulare* generated here to investigate the regulation governing the observed transcriptional changes. i) With a collection of mainly full-length, annotated transcript sequences of *T. triangulare*, we could re-quantify RNA-seq reads from the ABA time-course experiment (Maleckova *et al.*, 2019) against the reference of the same species (see Materials and methods for more detail), thus avoiding cross-species mapping. Besides increased mapping efficiency, we were able to quantify candidate lncRNAs. ii) The knowledge of genomic context of ABA-responsive loci enabled us to explore their promoter regions for presence of TFBS and compare them to motifs in promoters of other species relying on various photosynthesis modes.

Motif enrichment in promoters of differentially expressed genes

Motifs enriched in promoters of ABA-responsive genes, we relied on RNA-seq reads obtained previously in a time-course experiment (Fig. 1A in Maleckova *et al.*, 2019). In brief, RNA-sequencing was performed on individual mature leaves of *T. triangulare* after spraying with 200 μM ABA or a mock solution. The treated leaves were sampled after 40, 80, 160, 320, 640 and 1280 min. Sampling at 640 min after the treatment took place 160 min into the dark phase and the last sampling took place in 80 minutes into the subsequent light phase. After re-mapping, the lmms package (Straube *et al.*, 2015) was used to identify transcript with significantly ($q \leq 0.01$) altered temporal profiles. There were 11,107 such transcripts, which were grouped into seven distinct clusters based on similar patterns of transcript abundance (Fig. 6).

The question arises, whether genes, the transcripts of which share similar expression profiles, also share common TFBS in their promoter regions. To address it, genes of all transcripts with altered abundance patterns were identified in the draft genome assembly and sequences 500 bp upstream of the predicted transcript's beginning were analysed for TFBS enrichment using Discriminative Regular Expression Motif Elicitation (DREME; Bailey, 2011) implemented in the MEME Suite. TFBS frequency in individual clusters was compared to a set of randomly sampled promoter sequences, size of which corresponded to average size of DEG clusters. With this approach, 23 motifs frequent in promoters of ABA-responsive genes were identified (Fig. 6) and compared to TFBS in the Plant Cistrome Database using Tomtom (Gupta *et al.*, 2007), part of the MEME Suite.

The majority of identified motifs was present in clusters 1 and 4 showing transcript depletion in response to ABA. TFBS enriched in cluster 1 were associated especially with WRKY, TEOSINTE BRANCHED 1, CYCLOIDEA AND PCF TRANSCRIPTION FACTOR (TCP) and AP2/EREB TFs (Fig. 6A). WRKY and TCP binding sites were frequent in cluster 4 as well, alongside with TFBS of G2-LIKE TFs (Fig. 6D). Candidate TFs recognizing WRKY-associated binding sites included WRKY18 and WRKY40 (Tab. S8), both involved in ABA signalling. Promoters of up-regulated and early ABA-responsive genes (cluster 3) were enriched for MYB-RELATED TFs of the circadian oscillator (Fig. 6C, Tab. S8). Genes in clusters 5 and 7 also showed transcript accumulation in response to ABA and their promoters frequently contained motifs resembling TFBS of MYB TFs and GeBP/MYB-REALTED TFs, respectively (Fig. 6E, G). Both clusters also contained a motif which does not resemble any known TFBS.

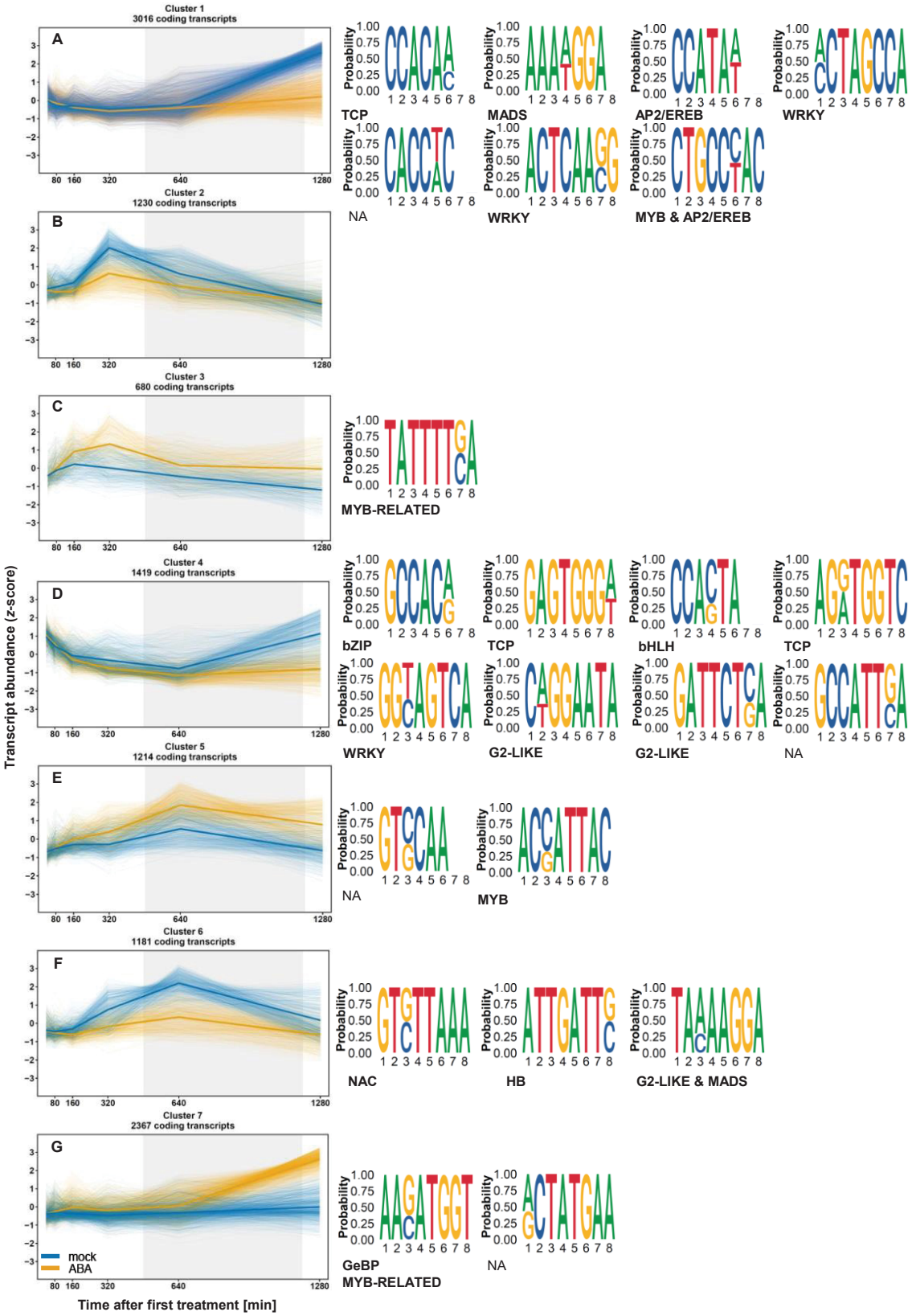


Figure 6. Promoter sequences of genes showing similar abundance profiles share common transcription factor binding sites. RNA-seq reads were aligned to the transcriptome assembly of *Talinum triangulare* and transcripts with altered temporal pattern were identified with lmms package ($q \leq 0.01$) followed by

Figure 6. (*Continued.*)
k-means clustering. Normalized transcript abundance is shown to the left, with duration of the dark phase marked with the grey background. Sequences 500 bp upstream of predicted start codons of differentially expressed genes were compared to a random set of *Talinum triangulare* promoters to identify motifs associated with each cluster (DREME and Tomtom, with and Plant Cistrome Database, with default parameters). Families of transcription factors predicted to bind the identified motifs are shown, complete results are available in Tab. S8.

Genes of the CAM pathway and carbohydrate metabolism are co-expressed with selected transcription factors of ABA signalling

Nearly 21% of the mapped transcripts showed significantly altered temporal pattern of transcript abundance in response to exogenous ABA (Fig. 6). To analyse the connection between ABA signalling and CAM induction from another point of view, time-ordered gene co-expression network (TO-GCN) analysis (Chang *et al.*, 2019), enabling identification of co-expressed gene pairs, was performed. TFs of *T. triangulare* were identified based on orthology to Arabidopsis TFs collated in the PlantTFDB (Jin *et al.*, 2017). However, it was not always possible to identify orthology between *T. triangulare* and Arabidopsis sequences, which were moreover rarely in 1:1 relationship. To reduce the size of the resulting network, only transcripts (both TFs and target genes) with altered temporal pattern as identified by the lmms package (more details in Materials and methods) were considered in the analysis. When a single gene encoded multiple transcripts (*i.e.* splice variants), only the most abundant transcript was retained, leaving 342 transcripts encoding TFs and 6,164 transcripts of structural genes to be analysed.

The resulting network consisted of 26,971 nodes. In the exploratory phase, seven TFs known to play a role in ABA signalling and a TF homologous to MYB-RELATED TFs of the circadian oscillator were selected for inspection of their most immediate interactions (*i.e.* "first neighbour"). The analysed portion of the network included several components of ABA signalling, such as a *SUCROSE NONFERMENTING 1-RELATED PROTEIN KINASE 2 (SnRK2)*, a *PYRABACTIN RESISTANCE 1-LIKE (PYL)*, and an ortholog of *HIGHLY ABA-INDUCED PP2C GENE (HAI)* and *ABSCISIC ACID INSENSITIVE (ABI)* (Fig. 7).

Moreover, several transcripts encoding CAM enzymes, enzymes of carbohydrate turnover and transport proteins were co-expressed with the considered TFs. The core CAM enzymes included *PHOSPHOENOLPYRUVATE CARBOXYLASE (PPC)*, *NADP-MALIC ENZYME (NADP-ME)*, *MDH* and *PYRUVATE ORTHOPHOSPHATE DIKINASE* (Fig. 7). Co-expressed transcripts associated with enzymes of carbohydrate metabolism included *STARCH SYNTHASE (SS)*, *BETA-AMYLASE (BAM1* or *CT-BMY)* and *BMY3* (aka *BAM9*) encoding for enzymes of starch metabolism, and a glycolytic *PHOSPHOFRUKTOKINASE (PFK)* (Fig. 7). Sugar transports included in the network were a *SUGARS WILL EVENTUALLY BE EXPORTED TRANSPORTER (SWEET)*, a *SUGAR TRASNPORTER (STP)*, a *GLUCOSE 6-PHOSPHATE/PHOSPHATE TRANSLOCATOR (GPT)* and a *VACUOLAR GLUCOSE TRANSPORTER 1 (VGT1)* (Fig. 7). Ion transporters co-expressed with ABA-responsive TFs included a *NA+/H+ EXCHANGER (NHX)* and a *POTASSIUM CHANNEL IN ARABIDOPSIS THALIANA (KAT)* orthologs (Fig. 7).

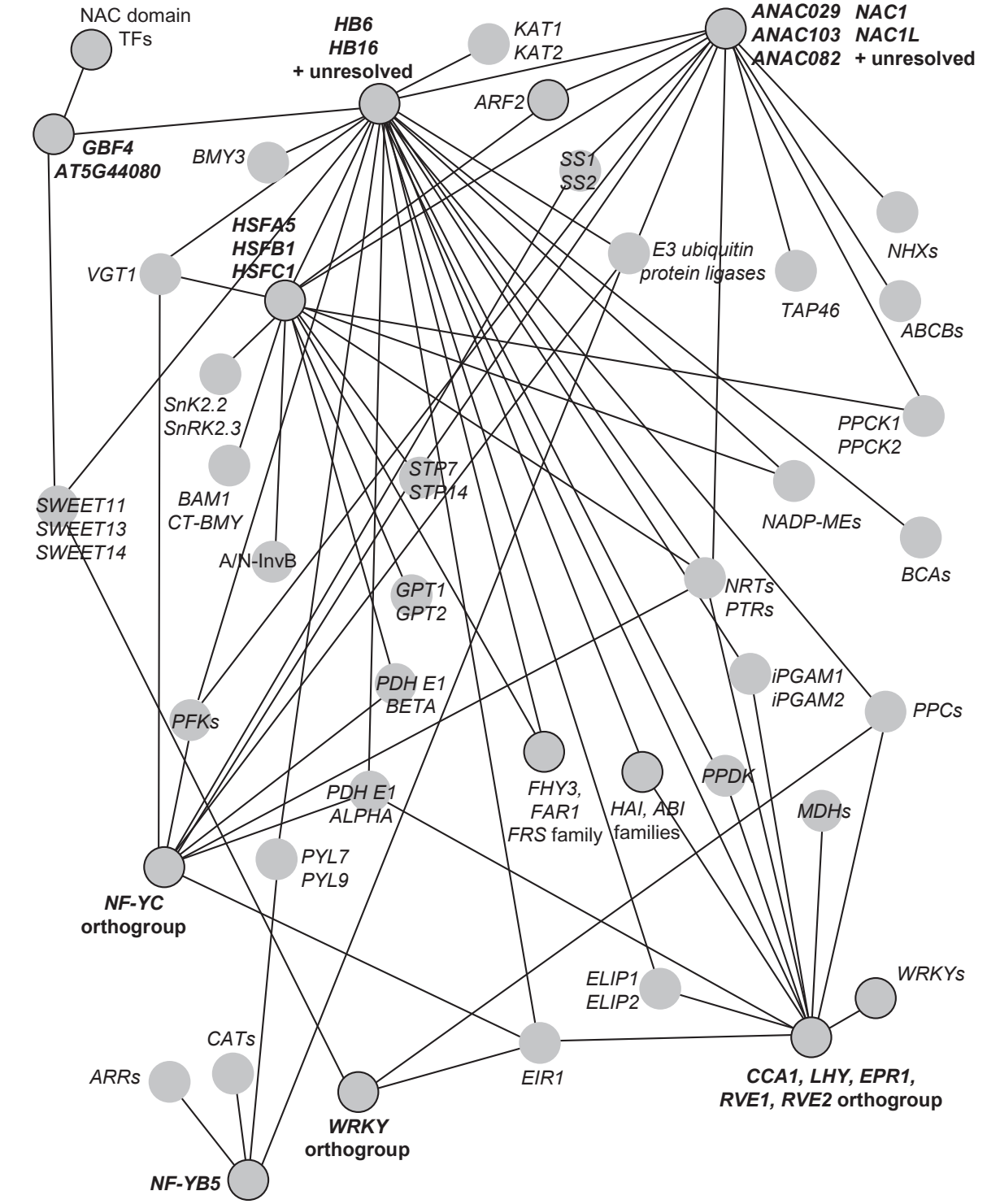


Figure 7. A subset of co-expressed gene pairs identified using the TO-GCN algorithm to obtain time-ordered gene co-expression network. Only differentially expressed transcripts and only the most abundant splice variants (if multiple ones were differentially expressed) were included in the analysis. The depicted nodes are limited to interactions at the "first neighbour" level and to transcripts coding for orthologs of ABA/stress-responsive transcription factors, a transcription factor of the circadian oscillator (bottom right), CAM enzymes, enzymes of carbohydrate metabolism and transporters. When (an) Arabidopsis ortholog(s) could be identified unambiguously for a given *Talinum triangulare* transcript, the respective gene name is given, otherwise a description of the entire orthogroups is used. Transcription factors are marked with a

Figure 7. (*Continued.*)
ABCB, ATP binding cassette B; ABI, ABA-insensitive; ANAC, NAC domain containing protein; ARF, auxin responsive factor; ARR, Arabidopsis response regulator; A/N-InvB, plant neutral invertase family protein; BAM1, ß-amylase 1; BCA, ß-carbonic anhydrase; BMY3, ß-amylase 3; CAT, cationic amino acid transporter; CCA1, circadian clock associated 1; CT-BMY, chloroplast ß-amylase; EIR1, ethylene insensitive root 1; ELIP, early light-induced protein; EPR1, early phytochrome responsive 1; FAR1, far-red impaired response; FHY3, far-red elongated hypocotyl 3; FRS, FAR1 related sequences; GBF4, G-box binding factor 4; GPT, glucose-6-phosphate/phosphate translocator; HAI, highly ABA-induced PP2C gene; HSF, heat shock factor; iPGAM, 2,3-biphosphoglycerate-independent phosphoglycerate mutase; KAT, potassium channel in Arabidopsis thaliana; LHY, late elongated hypocotyl; MDH, malate dehydrogenase; NADP-ME, NADP-dependent malic enzyme; NF-YB5, nuclear factor Y, subunit B5; NF-YC, nuclear factor Y, subunit C; NHX, Na^+/H^+ exchanger; NRT, nitrate transporter; PDH E1 ALPHA, pyruvate dehydrogenase E1 alpha subunit; PFK, phosphofructokinase; PPC, phosphoenolpyruvate carboxylase; PPDK, pyruvate orthophosphate dikinase; PTR, nitrate transporter; PYL, PYR-like; RVE, reveille; SnRK2, SNF1-related protein kinase; SS, starch synthase; STP, sugar transporter protein; SWEET, sugars will eventually be exported transporter; TAP46, 2A phosphatase associated protein of 46 kD; VGT1, vacuolar glucose transporter 1; WRKY, WRKY DNA-binding protein.

Interspecies comparison revealed enrichment of motifs associated with MYB-RELATED, HB and other TFs in promoters of Talinum triangulare

Motif analysis (DREME) of promoters of genes differentially expressed in response to ABA as well as TO-GCN analysis point at the possibility that CAM induction is at least in part regulated at the transcriptional level via the action of TFs of ABA signalling and or the circadian oscillator. To investigate whether genes of *T. triangulare* acquired new TFBS, promoters of genes contained in the analysed portion of TO-GCN (Fig. 7) were compared to promoters of orthologous genes in species employing a variety of carbon assimilation strategies. These included C_3 species Arabidopsis, *B. vulgaris* and *S. oleracea* and *Hordeum vulgare*, C_4 species *Z. mays*, obligate CAM species *A. comosus* and *Phalaenopsis equestris*, *K. fedtschenkoi* and *K. laxiflora*, and facultative CAM species *D. catenatum*. DREME analysis followed by Tomtom was used to identify and annotate enriched motifs.

Among the core CAM genes, enriched motifs were identified in promoters of *MDH, PPC* and *PPCK* orthogroups of *T. triangulare*, in other words enzymes contributing to nocturnal CO_2 assimilation into malic acid. Especially promoters within the *PPC* orthogroups, with four distinct motifs identified, showed a high level of enrichment. The predicted TFs binding to these sites included C2H2 TFs and MYB-RELATED TFs of the circadian oscillator (Fig. 8A).

The CAM cycle is tightly connected to carbohydrate metabolism and enriched motifs were identified in several orthogroups coding for enzymes of glycolysis and starch turnover. These included promoters of *ENOLASE (ENO), FRUCTOSE-BISPHOSPHATE ALDOLASE (FBA)* and *PFK* genes. There was no overlap among TFs that were predicted to bind to these motifs. For example, putative TFBS of FAR1-RELATED SEQUENCE 9 (FRS9) were associated with *FBA* promoters and putative TFBS of MYB TFs with *PFK* promoters (Fig. 8B). *T. triangulare* promoters of *ADP-GLUCOSE PYROPHOSPHORYLASE* (*APL)* were enriched for a motif that was predicted to bind HBs and a motif without similarity to TFBS of Arabidopsis was present in *BAM* promoters of *T. triangulare* (Fig. 8C)

Several motifs were frequently associated with multiple orthogroups of *T. triangulare* encoding transport proteins. In *ABCB* promoters, four distinct motifs were enriched, including TATCTT(T/C)A

resembling TFBS of MYB-RELATED TFs of the circadian oscillator. Another candidate binding to *ABCB* promoters was NUCLEAR FACTOR Y, SUBUNIT B4 (NF-YB4) (Fig. 8D). *SWEET* promoters of *T. triangulare* were predicted to contain TFBS of TFs of the circadian oscillator more frequently than other species included in the analysis (Fig. 8D).

TO-GCN (Fig. 7) revealed many connections to an *AUXIN RESPONSIVE FACTOR (ARF)* transcript and seven motifs were frequently identified in *T. triangulare* promoters of *ARF* genes (Tab. S9). These included putative TFBS of HEAT SHOCK FACTORS (HSF), FRS9, HB TFs and MYB-RELATED TFs of the circadian oscillator (Fig. 8E).

Figure 8. Motif enrichment in promoters of genes in selected orthogroups. Based on TO-GCN analysis and identified genes with altered temporal patterns, targets for interspecies comparison of promoter sequences were selected. Due to a limited resolution of individual orthologs, the comparison was performed at the level of entire orthogroups. Promoters (500 bp upstream of predicted start codon) of *T. triangulare* were compared to promoter sequences of the following species: Arabidopsis, *Beta vulgaris*, *Spinacia oleracea*, *Hordeum vulgare*, *Zea mays*, *Ananas comosus*, *Phalaenopsis equestris*, *Kalanchoë fedtschenkoi*, *K. laxiflora*, and *Dendrobium catenatum*. Motif enrichment and subsequent annotation were performed with DREME and Tomtom (and Plant Cistrome Database) with default parameters. The number of promoters with a hit for a given motif out of the total number of sequences analysed is shown next to each motif both for promoters of *Talinum triangulare* and in the control group containing all other species analysed. Full summary of transcription factors predicted to bind the enriched motifs in Table S9. Genes of (A) the CAM pathway, (B) glycolysis, (C) starch metabolism, (D) several transporters and an orthogroup of (E) an AUXIN RESPONSIVE FACTORs. *ABCB, ATP binding cassette B; APL, ADP-glucose pyrophosphorylase; ARF, auxin responsive factor; BAM, ß-amylase; ENO, enolase; FBA, fructose-bisphosphate aldolase; MDH, malate dehydrogenase; NHX, Na^+/H^+ exchanger; PFK, phosphofructokinase; PPC, phosphoenolpyruvate carboxylase; PPCK, phosphoenolpyruvate carboxylase kinase; SWEET, sugars will eventually be exported transporter.*

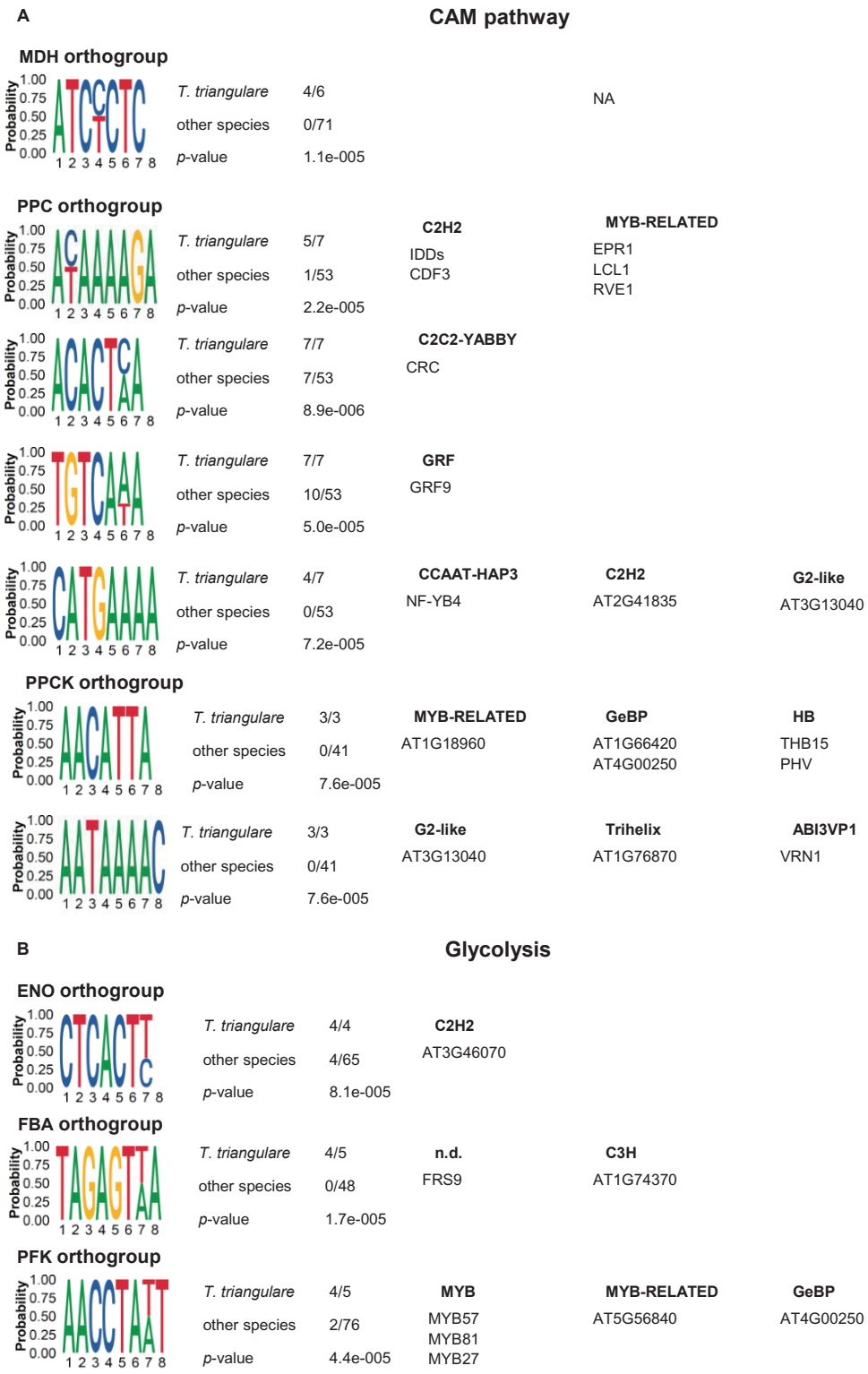


Figure 8. Motif enrichment in promoters of genes in selected orthogroups.

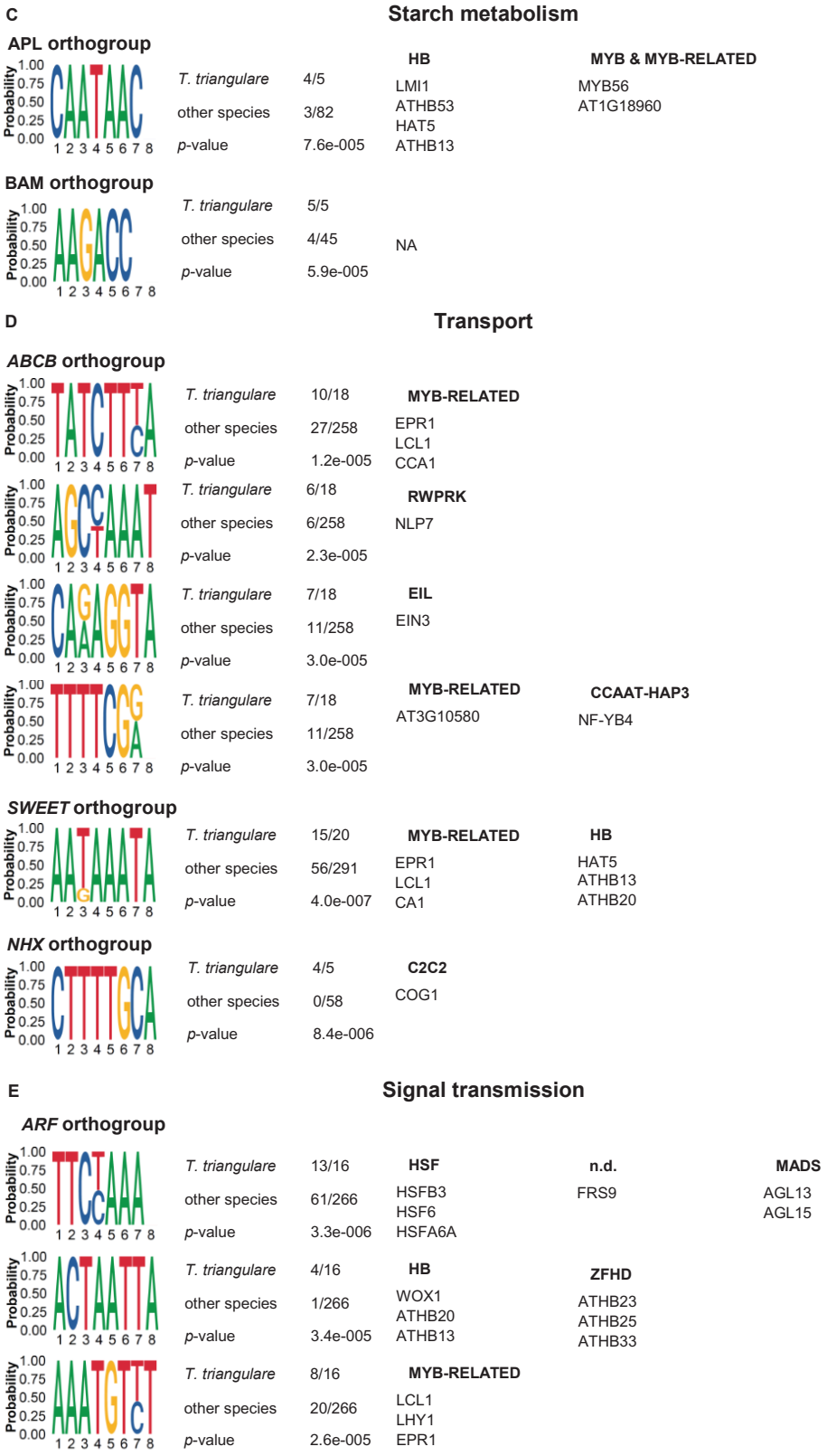
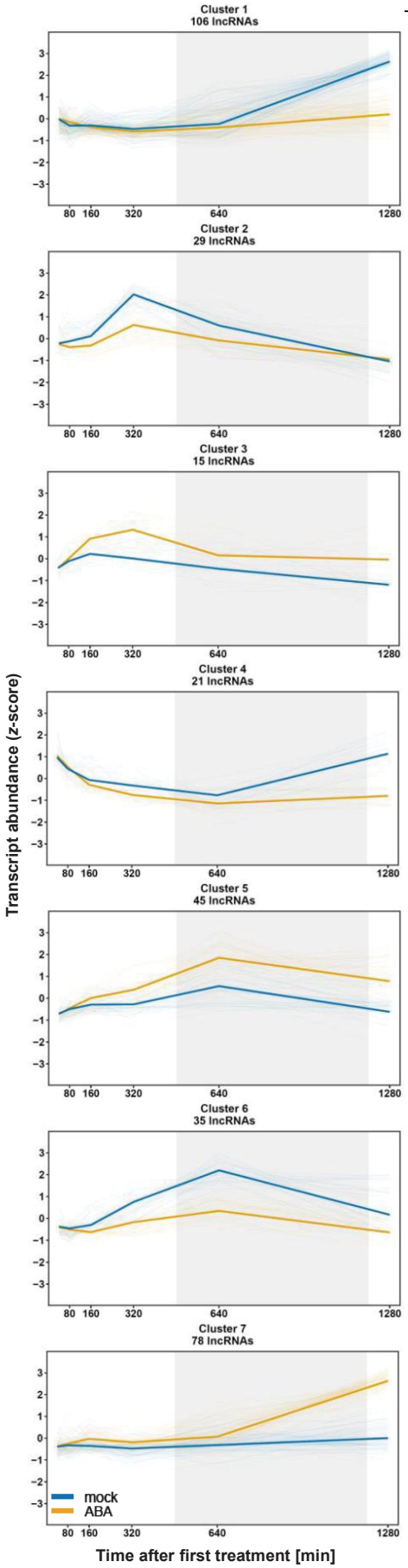


Figure 8 (*continued).* Motif enrichment in promoters of genes in selected orthogroups.



lncRNAs are co-expressed with protein-coding transcripts

Besides TFs, non-coding RNA also act as regulators of gene expression and choosing a long-read sequencing technology for transcriptome sequencing enabled us to obtain their sequences for the first time in a facultative CAM species. Transcripts of lncRNAs were quantified and analysed together with protein-coding transcripts, including lmms analysis. It identified 329 lncRNAs (23.5% of all mapped candidate lncRNAs) with altered temporal patterns. This is comparable to the proportion of protein-coding transcripts (21%) responsive to exogenous ABA. Subsequent *k*-means clustering yielded seven clusters with expression profiles of lncRNAs (Fig. 9) highly similar to those of coding transcripts (Fig. 6). The largest clusters were cluster 1 (106 lncRNAs) and cluster 7 (78 lncRNAs), down- and up-regulated in response to ABA, respectively. Cluster 3, showing very early responsiveness to ABA, contained 15 lncRNAs (Fig. 9).

Figure 9. Transcript abundance patterns of candidate long non-coding RNAs (lncRNAs). Candidate lncRNAs were identified based on lack of confident open reading frames and minimum of two exons. Predicted lncRNAs were retained in the transcriptome assembly of *T. triangulare* together with protein-coding sequences for mapping of RNA-seq reads. Differentially expressed lncRNAs were identified with the lmms package ($q \leq 0.01$), which was followed by *k*-means clustering to group lncRNAs with similar abundance profiles.

Changes in protein-coding sequences associated with facultative CAM

Besides investigations of mechanisms of transcriptional regulation of CAM induction in *T. triangulare*, new genomic resources provide an opportunity to investigate changes in coding sequences with regard to evolution of (facultative) CAM. To limit sequences of *T. triangulare* to unique ORFs only, translated CDS sequences were grouped with cDNA Cupcake to form ORF groups based on identical amino acid sequences (*i.e.* originating from highly similar CDS), which resulted in 39,016 unique "representative" ORFs (Fig. 1d) used for an OrthoFinder analysis. Of the 17,896 identified orthogroups, 8,782 (49%) were present in at least one obligate CAM species, at least one facultative CAM species and at least one species with any other type of photosynthesis (*i.e.* C_3, C_4 and C_3-C_4 intermediate species) (Fig. 10A).

Firstly, we compared translated transcriptome sequences of several species representing the diversity of carbon assimilation strategies (Tab. S1) to identify proteins possibly associated with ability of CAM induction. Secondly, amino acid conservation at sites important for altered kinetic properties or regulation of protein activity was examined in *T. triangulare* sequences of PEPC and PPDK, both well studied enzymes involved both in CAM and C_4 pathways.

Orthogroups shared by facultative CAM species and transcriptional responsiveness in Talinum triangulare to exogenous ABA

We examined whether facultative CAM is associated with emergence of new genes, which could act as "CAM switches" connecting stress-related ABA signalling with metabolic re-programming specific of the CAM. Firstly, we compared entire proteomes at the protein level across a number of species to identify such candidates. Secondly, we identified *T. triangulare* orthologs showing altered temporal patterns of transcript abundance during ABA-mediated CAM induction. Of the 2,100 orthogroups specific to facultative CAM species included in the analysis, 1,538 (15.8% of all orthogroups of *T. triangulare*) were shared by *T. triangulare* and at least one additional facultative CAM species (Fig. 10B). In these orthogroups, there were 2,439 predicted ORFs of *T. triangulare* and their sequences were screened for conserved protein domains (InterProScan). The five most frequently occurring protein domains (based on gene count) are listed in Tab. 1. Besides leucine-rich domains, protein kinases and TF domains, several predicted proteins carry an EF-hand and two genes code for a protein product containing pyruvate-kinase-like domain. One instance of ABA/WDS (water deficit stress) induced protein and three transport proteins were detected as well (Tab. 1).

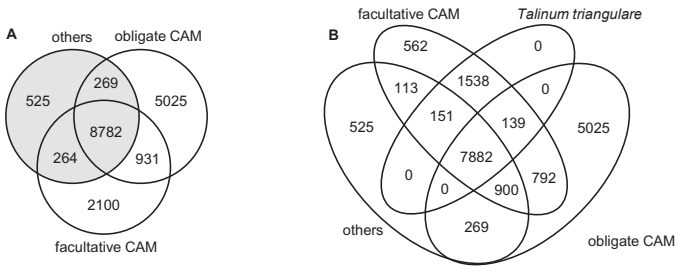


Figure 10. Proportions of shared orthogroups in a proteome-wide comparison (OrthoFinder) of species employing various carbon assimilation strategies. For this purpose, species included in the analysis (Tab. S1) were classified as facultative CAM, obligate CAM and others.

Figure 10. (*Continued.*)
(A) Of 26,010 orthogroups identified, 8,782 (33.8 %) were present in at least one species of each category, while 931 orthogroups contained proteins specific to facultative and obligate CAM species. (B) For a selection of ABA-responsive candidates specific to facultative CAM species, orthogroups shared exclusively between *T. triangulare* and at least one more facultative CAM species were identified. There were 1,538 such orthogroups.

Of the 2,439 predicted ORFs of *T. triangulare* with an ortholog in at least one additional facultative CAM species, 377 (encoded by 249 distinct genes) showed altered temporal pattern in response to exogenous ABA. These were particularly abundant in nocturnally down-regulated cluster 1 and nocturnally up-regulated cluster 7 (lmms analysis; Fig. 6A), comprising 120 and 147 transcripts (encoded by 72 and 97 genes), respectively (Tab. S10). Cluster 3, showing rapid transcript accumulation in response to ABA, included only six DEGs possibly connected to CAM induction (Tab. S10), but the ABA/WDS induced protein from Tab. 1 as well as one of the protein kinases were among them.

Table 1. InterProScan results of *T. triangulare* proteins belonging to orthogroups specific to facultative CAM species. The shown categories were selected to represent the most frequently occurring domains and transport processes.

InterPro annotation	Gene count	ORF count
Leucine-rich repeat (domain)	12	19
Zinc finger (all types)	8	10
Protein kinase (domain)	6	7
Ubiquitin domain and ubiquitin-conjugating enzyme	4	5
Transcription factor (any domain)	4	17
EF-hand domain	3	4
Pyruvate kinase-like domain superfamily	2	3
ABA/WDS induced protein	1	1
ABC transporter type 1, transmembrane domain superfamily	1	2
Amino acid transporter, transmembrane domain	1	1
Ion transport domain	1	1

Variability of protein coding sequences of core CAM enzymes

PEPC and PPDK are two examples of C_4/CAM enzymes with detailed understanding of their activity, amino acid substitutions differentiating C_3 and C_4/CAM isoforms and sites of post-translation modifications (Westhoff and Gowik, 2004; Yang *et al.*, 2017). Predicted PEPC and PPDK protein sequences of *T. triangulare* were therefore compared to a C_4 ortholog of *Z. mays* and a CAM ortholog *K. fedtchenkoi* to identify *T. triangulare* with potentially altered post-translational regulation and/or kinetic properties.

In the case of PEPC, six distinct genes were identified with the pipeline currently in use: two orthologous to Arabidopsis PEPC2 (PB.6486 and PB.6488) and four orthologous to at least two Arabidopsis PEPCs (Fig. 11A). A multiple sequence alignment of PEPC protein sequences revealed the diversity among the proteins encoded by individual PEPC orthologs as well as their numerous splice variants (Fig. 11B). There was no protein variant containing all signatures of a typical C_4 PEPC as for example PEP1 of *Z. mays*, but four full-length PEPC proteins of *T. triangulare* possessed some of them in various combinations. Ser15, Ser136 and Thr138 (numbering based on translation of *Z. mays* CAA33317)

A

Arabidopsis gene	Talinum gene	Talinum splice variants
PPC1 (AT1G53310)	PB.5633	ORFgroup_PB.5633_3, ORFgroup_PB.5633_2
	PB.6485	ORFgroup_PB.6485_4, ORFgroup_PB.6485_9, ORFgroup_PB.6485_8
	PB.6487	ORFgroup_PB.6487_6, ORFgroup_PB.6487_8
	PB.7729	ORFgroup_PB.7729_5, ORFgroup_PB.7729_11, ORFgroup_PB.7729_3, ORFgroup_PB.7729_7
PPC2 (AT2G42600)	PB.6485	ORFgroup_PB.6485_10, ORFgroup_PB.6485_14, ORFgroup_PB.6485_2, ORFgroup_PB.6485_13
	PB.6486	ORFgroup_PB.6486_1
	PB.6487	ORFgroup_PB.6487_7, ORFgroup_PB.6487_5, ORFgroup_PB.6487_1, ORFgroup_PB.6487_2
	PB.6488	ORFgroup_PB.6488_1
PPC3 (AT3G14940)	PB.5633	ORFgroup_PB.5633_3, ORFgroup_PB.5633_2
	PB.6485	ORFgroup_PB.6485_4, ORFgroup_PB.6485_9, ORFgroup_PB.6485_8
	PB.6487	ORFgroup_PB.6487_6, ORFgroup_PB.6487_8
	PB.7729	ORFgroup_PB.7729_5, ORFgroup_PB.7729_11, ORFgroup_PB.7729_3, ORFgroup_PB.7729_7

B

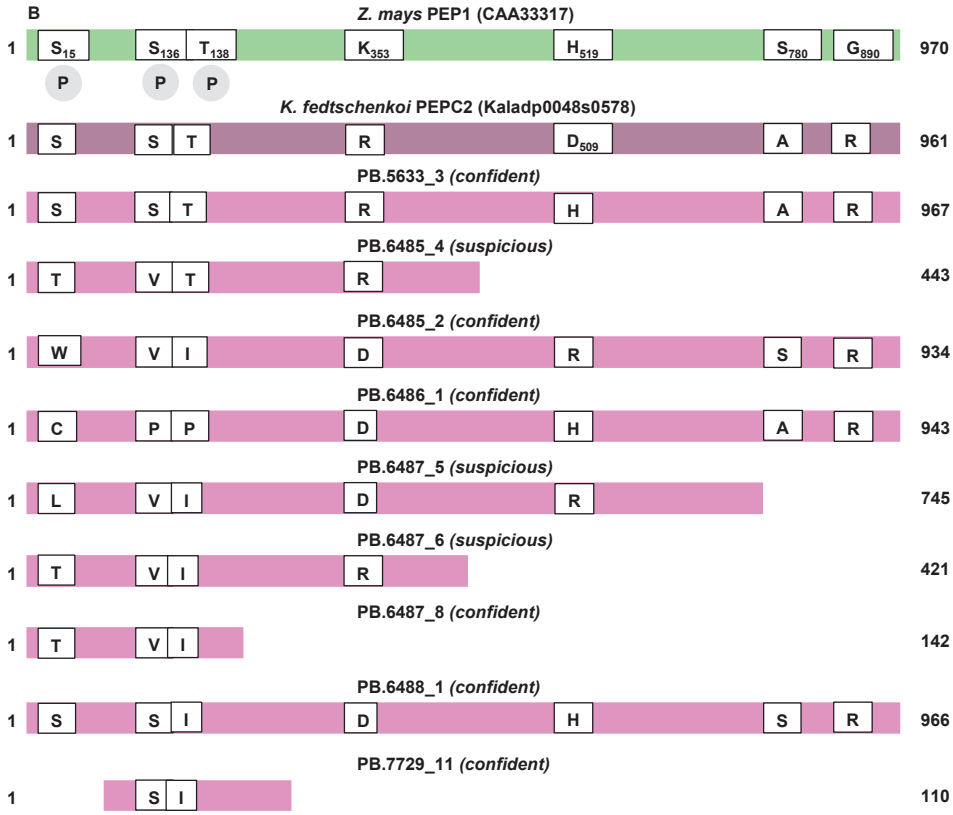


Figure 11. PHOSPHO*ENOL*PYRUVATE CARBOXYLASE-coding sequences of *Talinum triangulare*. (A) Orthologs of Arabidopsis PEPC1, PEPC2 and PEPC3 could be identified unambiguously among translated CDS sequences od *T. triangulare*, yet with a high degree of similarity. (B) A schematic depiction of amino acid variability at sites associated with a C_4 PEPC of *Zea mays* and a predicted CAM isoform of *Kalanchoë fedtschenkoi*. Amino acid numbering is based on translation of CAA33317 of *Z. mays*, except for D_{509} in the sequence of *K. fedtschenkoi* which is a true position in the *K. fedtschenkoi* protein. Total lengths of predicted protein sequences are shown at the right. For *T. triangulare* sequences, confidence of the open reading frame prediction is shown in brackets.

are targets of phosphorylation (Lu *et al.*, 2011) and homologous sites with phosphorylation potential were preserved in several orthologs of *T. triangulare* but frequently with Ser/Thr exchange at position 15 (Fig. 11B). Gly890 is associated with reduced inhibition by malate (Wedding *et al.*, 1990), but surprisingly, it was not preserved in any PEPC of *T. triangulare*. Numerous predicted proteins of *T. triangulare* were truncated prior to this site and all full-length proteins contained a conserved Arg at the homologous site (Fig. 11B). Up to date, there was only a single site proposed that possibly underwent convergent amino acid substitution in CAM species – Asp at position 509 of *K. fedtschenkoi* PEPC2 (corresponds to His519 of *Z. mays* PEP1) (Yang *et al.*, 2017). All full-length proteins of *T. triangulare* contained the ancestral variants and so did for example pineapple, *P. oleracea* and *M. crystallinum* (Fig. S4).

Two PPDK orthologs were identified in *T. triangulare* (Fig. 12A). PPDK is regulated post-translationally by phosphorylation at Thr309, Ser506, Thr527 and Ser528 (numbering based on translation of *Z. mays* CAA33054), with Gly525 contributing to phosphorylation of Thr527 and Ser528, which is mediated by PPDK REGULATORY PROTEINs RP1 and RP2 (Astley *et al.*, 2011). All known phosphorylated sites were preserved in full-length PPDK proteins of *T. triangulare*, but due to alternative splicing, both genes also gave rise to variants lacking the target sites of RP (Fig. 12B).

A

Arabidopsis gene	Talinum gene	Talinum splice variants
PPDK (AT4G15530)	PB.1546	ORFgroup_PB.1546_3, ORFgroup_PB.1546_2, ORFgroup_PB.1546_11, ORFgroup_PB.1546_4, group_PB.1546_7
	PB.15274	ORFgroup_PB.15274_7, ORFgroup_PB.15274_6, ORFgroup_PB.15274_9, ORFgroup_PB.15274_8

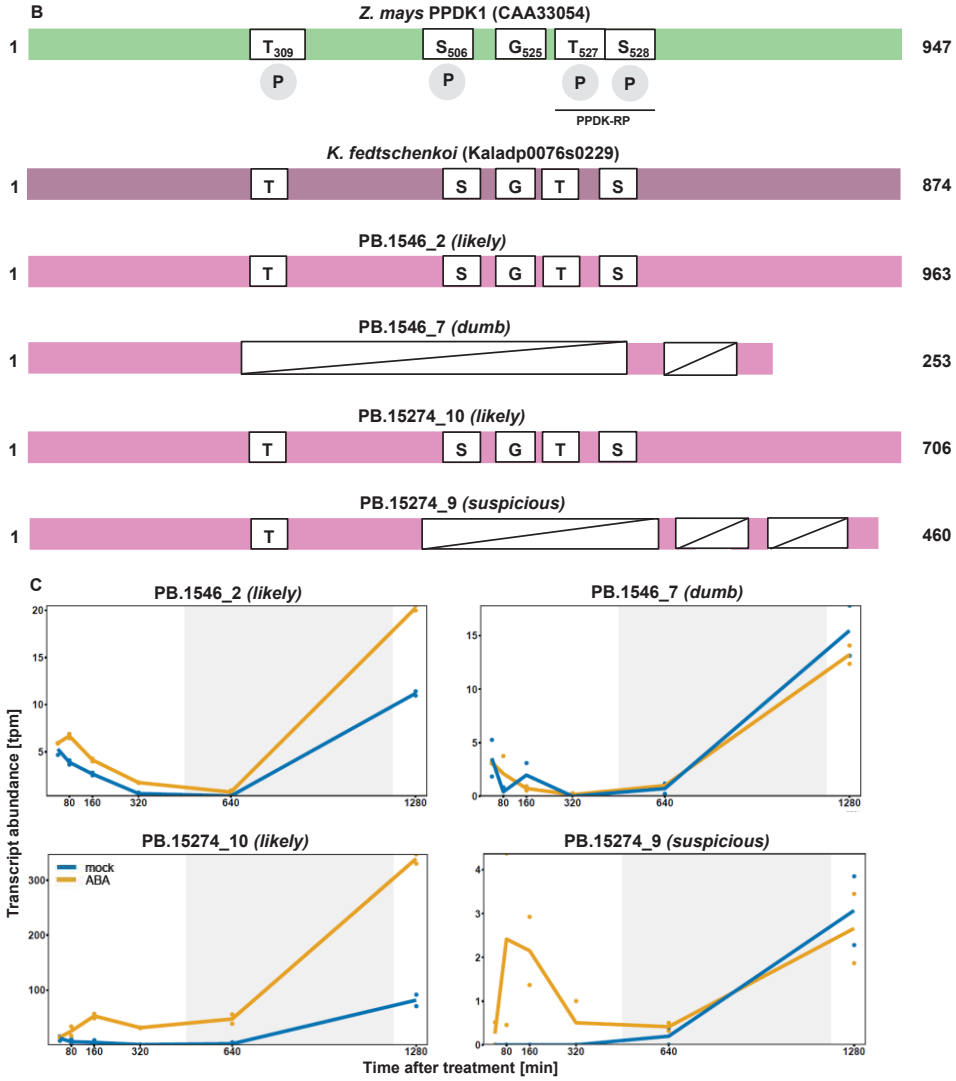


Figure 12. PYRUVATE, PHOSPHATE DIKINASE-coding sequences of *Talinum triangulare*. (A) Two orthologs of a single-copy Arabidopsis gene were identified in the transcriptome of *T. triangulare*. (B) Both gene copies gave rise to proteins predicted to carry all known signatures of a C_4 isoforms, such as that of *Zea mays*, as well as alternatively spliced variants lacking the sites associated with C_4 isoforms. Amino acid numbering is based on translation of CAA33054 of *Z. mays*, total lengths of predicted protein sequences are shown at the right. For *T. triangulare* sequences, confidence of the open reading frame prediction is shown in brackets. (C) Transcript abundance of several isoforms of *T. triangulare* in response to ABA.

Discussion

To expand the available genomic resources of our facultative CAM model species *T. triangulare*, we assembled a (draft) genome and transcriptome. Using long-read sequencing technologies, we identified transcribed loci of *T. triangulare* and estimated the diversity of splice variants. Moreover, we captured non-coding transcripts, making identification of putative lncRNAs and their quantification during CAM induction possible for the first time in this species. Secondly, we identified genes co-expressed with transcription factors involved in ABA/stress signalling. Finally, we used both assemblies for interspecies comparison of CREs and protein-coding sequences to identify changes associated with facultative CAM.

The Talinum triangulare genome and transcriptome

Genome assembly and indications of whole-genome duplication

Nanopore reads were assembled with Flye and polished with a subset of RNA-seq reads into 234 contigs with a total size of 608.5 Mbp (Tab. S3). This is in strong contrast to the expected genome size of 3.89 Gbp (*i.e.* approximately 1.9 Gb for a haploid assembly) as estimated by flow cytometry (Fig.S2A). This contradiction might occur through i) an erroneous assembly (*e.g.* due to a choice of parameter and/or assembler), ii) a recent whole-genome duplication event in the evolutionary history of *T. triangulare* or iii) endoreduplication. Based on flow cytometry measurements, endoreduplication seems the most likely explanation, but experimental confirmation is required for a finite conclusion.

Large plant genomes are known to contain a substantial proportion of repetitive sequences (Flavell, 1980; Schnable *et al.*, 2009; Ming *et al.*, 2015), which are problematic to resolve during the assembly process. However, raw reads with an N_{50} of approximately 30 kbp (Tab. S2) should be long enough to span even long repeats. Additionally, another assembly with highly comparable statistics was obtained in an entirely independent approach using the SMARTdenovo assembler (https://github.com/ruanjue/smartdenovo; Tab. S11), the results of BUSCO analysis with 91.8% of complete *Embryophyta* orthologs identified in the genome assembly (Fig. 2C) as well as a mean alignment rate of 98.38% when mapping RNA-seq reads to the genome assembly with HISAT (Tab. S4) strongly refute the possibility of losses in protein-coding regions during the assembly process.

Both polyploidisation and endoreduplication lead to increased content of nuclear DNA. Polyploidisation events occurred frequently and at different times throughout the evolution of *Caryophyllales*, including the *Portulacineae* clade, which *Talinum* spp. belong to (Yang *et al.*, 2015*b*, 2018). Not only are tetraploid populations of *T. calycinum* and *T. calcaricum* known (Harris *et al.*, 1993), but chromosome counts of both $2n = 24$ and $2n = 48$ were reported for *T. triangulare* (Baquar, 1986; Nyananyo and Olowokudejo, 1986), leaving the possibility that our population of *T. triangulare* is of mixed ploidy as reported for example for *Acacia senegal* (Odee *et al.*, 2015).

During endoreduplication, chromosomal DNA content replicates without an intervening cell division, increasing the ploidy level. It occurs frequently at different stages of development both in plants and animals (Sugimoto-Shirasu and Roberts, 2003; Lee *et al.*, 2009). Endoreduplication can have the same impact on morphology of affected cells as stable and heritable polyploidisation, but unlike during polyploidisation, the proportion of cells with different ploidy levels would vary between plant organs and within a single organ throughout its development. Endoreduplication was confirmed among several

succulent species, including the facultative CAM species *M. crystallinum* (De Rocher *et al.*, 1990). Occurrence of endoreduplication in (leaf) tissue of *T. triangulare* appears a likely explanation of the discrepancy between the estimated genome size and the assembly size. Firstly, flow cytometry revealed the presence of an additional population of *T. triangulare* nuclei with DNA content smaller than both G1- and G2-phase nuclei of both *T. triangulare* and *S. lycopersicum*. Secondly, its proportion seemed to change with plant age, being most highly abundant in leaf tissue of five-week-old plants but absent in leaves of plants more than two months old (Fig. S1B, C). Since measurements with older plants were used for genome size estimation but material of one-month-old plants was used for whole-genome sequencing, it is likely that predominantly diploid nuclei were used for gDNA extraction. Even when endoreduplication was already ongoing at this age, it would not be detectable in the assembly since endoreduplication is a whole-genome duplication.

Increased ploidy – even through developmentally regulated endoreduplication – might be advantageous for CAM species. It was shown for C_4 species that increased gene dosage can lead to increased gene expression (Bianconi *et al.*, 2018). From a CAM perspective, the effect of increased nuclear DNA content on cell morphology is an intriguing possibility. Higher content of nuclear DNA was shown to correlate with increased cell size, largely mediated by expansion of vacuole size (Sugimoto-Shirasu and Roberts, 2003), which is critical for sufficient storage capacity of nocturnally accumulating organic acids. Harris *et al.* (1993) observed higher day/night fluctuations of malic acid in tetraploid *Talinum* spp. compared to diploid ones. Beside the effect on cell size, higher DNA content was connected with increased guard cell length but lower stomatal density (Sugimoto-Shirasu and Roberts, 2003; Beaulieu *et al.*, 2008) and even increased leaf thickness and drought tolerance in the CAM species *Sedum pulchellum* (Harriet, 1946). It should be further explored whether large genome (*e.g.* some *Aloe* spp.; Zonneveld, 2002) or endoreduplication of small genomes are adaptation enabling CAM evolution and whether increased nuclear DNA content contributes to higher gene expression and/or storage capacity of malic acid.

Transcriptome assembly, annotation, and identification of orphan genes

Assembling transcriptomes *de novo* is widely used when annotated reference genomes are not available (*e.g.* Simon *et al.*, 2009; Martin and Wang, 2011). At the same time, *de novo* assembled transcriptome are a valuable source for genome annotation (Venturini *et al.*, 2018). We assembled the first genome of *T. triangulare* with current gene model predictions based solely on identified unique isoforms (HQ transcripts obtained with the isoseq3 pipeline were aligned to the draft genome and collapsed to unique isoforms with cDNA Cupcake) and prediction of ORFs with ANGEL (Fig. 1). Iso-Seq reads were processed using the isoseq3 pipeline and collapsed with cDNA Cupcake to 19,997 loci encoding for 50,015 transcripts (Fig. 1d). ANGEL classified 18,757 (93.8%) loci as protein-coding and predicted 54,107 distinct ORFs (Fig. 1d). BUSCO analysis determined 86.5% completeness of the transcriptome assembly but discovered over 40% duplicated BUSCOs (Fig. 3A). Duplicated BUSCOs probably resulted from multiple transcripts encoded by individual loci. On average, each gene gave rise to 2.66 isoforms and 2.88 ORFs, which is comparable to 2.4 transcripts per gene in Arabidopsis (Zhang *et al.*, 2017). The higher number of ORFs over the detected transcripts suggests that occasionally multiple ORFs per transcript were predicted. Alternative promoter usage has just started to be explored in plants (Hernandez-Garcia and Finer, 2014; Gregory, 2018). Given this phenomenon is more frequent at least in some species

this would increase the number of predicted ORFs. The fact that (at least some) ORFs tagged as suspicious are not artifacts is supported by increased proportion of missing BUSCO orthologs upon their removal from the transcriptome assembly (Fig. S2). Therefore, suspicious ORFs were retained in the transcriptome assembly and further confirmation of predicted ORFs will be done in the future. Combining several approaches, 97% of protein sequences were annotated based on their homology to other proteins or recognized protein domains (Fig. 3C). It must be confirmed whether the unannotated proteins are specific to *Talinum* (and possibly closely related) species or whether these also include erroneously predicted ORFs (*e.g.* by alternative assessment of their coding potential or *de novo* gene prediction in the polished genome assembly).

Besides protein-coding transcripts, we analysed non-coding transcripts in the present assembly. We particularly focused on lncRNAs as they were shown to play regulatory roles in a variety of processes, including development, stress response and plant immunity (Mach, 2017; Wang *et al.*, 2017; Rai *et al.*, 2019). Based on the coding potential estimation as well as minimum transcript length and number of exons, 3,184 putative lncRNAs were identified (Fig. 5). Their sequences were aligned to CDS as well as to UTR sequences. The majority (45.5%) showed homology to regions contained both CDS and UTR sequences, while 32.9% of them did not align. These likely originate from intergenic or intronic regions (Ma *et al.*, 2013; Rai *et al.*, 2019).

Evidence of transcriptional control of CAM induction via CREs and lncRNAs

We previously reported the pace and extent at which transcript levels change in *T. triangulare* during CAM induction but were limited by cross-species quantification (Maleckova *et al.*, 2019). The long-read sequencing-based transcriptome assembly now facilitates differentiation among individual splice variants and quantification of protein-coding sequences not present in Arabidopsis or *B. vulgaris* as well as quantification of lncRNAs. Transcriptional analyses conducted up to date indicate that CAM activities must be co-ordinately regulated (Cushman *et al.*, 2008; Abraham *et al.*, 2016; Borland *et al.*, 2016; Brilhaus *et al.*, 2016; Yang *et al.*, 2017; Heyduk *et al.*, 2019). To test the hypothesis that co-ordinated regulation is achieved at the transcriptional level via shared regulatory motifs in promoter sequences of the target genes, we performed TO-GCN and motif enrichment analyses.

Gene expression profiles correlate with common promoter motifs

For motif enrichment, DREME was used to compare promoter regions of genes in each temporal cluster to their expected frequencies based on a random sample of *T. triangulare* promoters. Common motifs were identified for each promoter set except for genes in cluster 2 (Fig. 6B). As expected, TFBS recognized by ABA signalling TFs occurred with an increased frequency in promoters of DEGs: GCCAC(A/G) and GG(T/C)AGTCA motifs in cluster 4 and ATTGATT(G/C) in cluster 6, transcripts of both clusters showing depletion in ABA-treated leaves (Fig. 6D, F; Tab. S8), illustrating the importance of down-regulation during CAM induction alongside with accumulation of specific transcripts.

The GCCAC(A/G) motif found in promoters of cluster 4 genes is similar to the TFBS of GBF3 and ABI5, ABA-RESPONSIVE ELEMENT BINDING PROTEIN 3 (AREB3) and ABSCISIC ACID RESPONSIVE ELEMENTS-BINDING FACTOR 2 (ABF2/AREB1) (Tab. S8). Even though closely related and involved in ABA-mediated stress responses, functional differentiation exists between these

TFs. In Arabidopsis, *GBF3* transcripts were shown to accumulate in response to ABA and its impact on increased tolerance to multiple abiotic stresses was demonstrated (Lu *et al.*, 1996; Ramegowda *et al.*, 2017). AREB3 contributes to initiation of long-term ABA-induced changes in gene expression (Uno *et al.*, 2000). AREB3 might thus be an interesting candidate in the context of CAM regulation once the pathway was induced. A connection between CAM and ABA signalling was proposed for obligate CAM species, for example in stomatal regulation (Abraham *et al.*, 2016; Heyduk *et al.*, 2019). Overexpression of *ABF2* increased Arabidopsis tolerance to several abiotic stresses including drought and heat, but at the same time led to downregulation of *RIBULOSE BISPHOSPHATE CARBOXYLASE SMALL CHAIN (RBCS)*, *HXK1* and *HXK2*, thus connecting ABA signalling to both photosynthesis and sugar metabolism (Kim *et al.*, 2004). This signalling crosstalk could be further adjusted during CAM evolution. However TFs of the same family share binding sites (Song *et al.*, 2016; Kurihara *et al.*, 2020). Thus, to unambiguously confirm binding of the identified TFs to their putative targets, experimental confirmation is needed. Additional bZIP candidates, not necessarily connected to ABA signalling but potentially recognizing the enriched motif, include bZIP16, bZIP42 or bZIP28 (Tab. S8).

Motifs recognized by the WRKY – another ABA-associated TF family – were abundant in promoters of genes in clusters 4 and 6 (Fig. 6D, F). Two motifs were identified as possible binding sites of WRKY18 and WRKY40, which through a joint action with WRKY60 regulate expression of other components of ABA signalling, including *ABF2, ABF4, ABI5* and *ABI4* TFs (Liu *et al.*, 2012). Additional TF candidates connected to ABA response/CAM induction include ATHB5, ATHB6 and HAT22, all potentially recognizing ATTGATT(G/C) enriched in down-regulated genes of cluster 6 (Fig. 6F, Tab. S8). Here, it was not possible to unambiguously identify ABA-responsive TFs binding the enriched motifs, but we could provide evidence that coordinated down-regulation of gene expression during CAM induction is at least in part due to direct action of ABA-responsive TFs. This is consistent with the role of WRKY TFs and HB6 as negative regulators of ABA signalling in Arabidopsis (Himmelbach *et al.*, 2002; Shang *et al.*, 2010; Liu *et al.*, 2012; Song *et al.*, 2016). Further investigations should focus on differentiation of ABA-induced down-regulation between C_3 and facultative CAM species.

TCP TFs may also act as negative regulators during CAM induction, as the respective TFBS were predicted in genes belonging to clusters 1 and 4 (Fig. 6A, D). In Arabidopsis, the role of TCP TFs has been studied in connection to regulation of growth and early development (rewieved by Li, 2015). In *Vitis vinifera*, expression of many TCPs is inhibited by stimuli as different as drought and waterlogging (Jiu *et al.*, 2019). Candidate TCPs recognising motifs in promoter sequences of genes in cluster 1 and 4 included TCP15, TCP20 and TCP22 (Tab. S8). In Arabidopsis, TCP20 and TCP22 contribute to regulation of expression of the clock gene *CIRCADIAN CLOCK ASSOCIATED 1 (CCA1)* (Wu *et al.*, 2016*a*), and TCP14 and TCP15 regulate cytokinin responses (Steiner *et al.*, 2012). Besides that, *TCP15* expression fluctuates over the diel cycle and protein-protein interaction assays revealed its physical interaction with two clock components, LATE ELONGATED HYPOCOTYL1 (LHY1) and PSEUDO RESPONSE REGULATOR5 (PRR5) (Giraud *et al.*, 2010). Furthermore, TCP15 is a regulator of diurnally regulated protein biosynthesis in organelles, including enzymes of the TCA cycle (Giraud *et al.*, 2010), which might also be relevant for CAM regulation. Several CAM enzymes localize not only to the cytosol but also to mitochondria (*e.g.* NADP-ME) and we observed accumulation of transcripts of several enzymes of the TCA cycle and amino acid degradation during CAM induction (Maleckova *et al.*, 2019). Additionally, a

recent analysis of TCP TFs in grapevine genome revealed that their promoters are rich for *cis*-acting elements (Jiu *et al.*, 2019). This makes them good candidates to integrate various signals and translate external stimuli into a concerted transcriptional regulation of the circadian oscillator together with other genes containing TCP binding sites during CAM induction.

Besides the motif in cluster 3 associated with MYB-RELATED TFs of the circadian oscillator, other motifs associated with genes up-regulated in response to ABA included two motifs in cluster 5 as well as two motifs frequent among genes of cluster 7. The AC(C/G)ATTAC motif abundant in cluster 5 resembles TFBS of MYB TFs, among them MYB77, MYB81, MYB118 and MYB119 (Fig. 6E, Tab. S8). MYB TFs have been studied in connection to auxin signalling, but based on findings in Arabidopsis, MYB77 seems like a good candidate for regulation of gene expression in response to environmental changes. MYB77 interacts with ABA receptor PYL8, which in turn enhances the binding of MYB77 to its target motif (Zhao *et al.*, 2014). AUXIN RESPONSE FACTORs (ARF) are other interaction partners of MYB77 at the protein level (Shin *et al.*, 2007) and *ARF2* ortholog was co-expressed with *ANAC* and *HSF* orthologs in ABA-treated leaves (Fig. 7). Such interaction can have an effect both on binding of MYB77 to promoter sequences of the target genes and on binding of ARFs to other interaction partners. The AA(G/C)ATGGT motif frequent in cluster 7 was predicted to bind several GeBP and MYB-RELATED TFs, including MYB67, was enriched in cluster 7 (Fig. 6G, Tab. S8). MYB67 could also play a role in mediating the coordination between different signalling pathways both at transcriptional and post-translation level. It was shown to bind G-BOX BINDING FACTOR 1 (GBF1), a TF of cryptochrome signalling and to act as a negative regulator of one of the *RIBULOSE BISPHOSPHATE CARBOXYLASE SMALL CHAIN (RBCS)* subunits (Mallappa *et al.*, 2006; Nozawa *et al.*, 2009).

Surprisingly, several motifs associated with ABA-responsive genes did not show significant similarity to any TFBS of Arabidopsis (Fig. 6A, D, E, G). Enriched but unannotated motifs could be present in Arabidopsis but were not identified yet, for example due to low occupancy or TF binding under very specific conditions. Alternatively, they could have evolved as TFBS in certain plants lineages only and in that case, it will be of interest to identify the TFs binding to them. Finally, these sites may act as binding sites for other types of transcriptional regulators, such as non-coding RNAs. Contribution of miRNAs to post-transcriptional regulation of the CAM pathway was confirmed in the obligate CAM species pineapple (Wai *et al.*, 2017). Numerous diurnally fluctuating lncRNAs have been identified in the same species (Bai *et al.*, 2019) and several ABA-responsive lncRNAs have been identified in the transcriptome of *T. triangulare* (Fig. 9).

With an increased number of CAM species sequenced, more thorough comparisons will be possible, but already a simple comparison between the obligate CAM species pineapple (Wai *et al.*, 2017) and the facultative CAM *T. triangulare* identified shared over-represented TFBS. In both cases, motif enrichment analysis was preceded by clustering based on common expression patterns – green *vs* non-photosynthesizing tissue in pineapple and mock- *vs* ABA-treated leaves in *T. triangulare*. One of the clusters of genes showing circadian fluctuation in transcript abundance specifically in the green pineapple tissue (cluster 1, Wai *et al.,* 2017) was enriched for CCA1 and TCP15 TFBS. In the case of *T. triangulare*, a clock-associated motif was enriched in ABA up-regulated cluster 3 (Fig. 6C, Tab. S8) and motifs resembling TCP TFBS were frequent in clusters 1 and 4, both showing transcript depletion in response to ABA (Fig. 6D, Tab. S8). These are promising CAM-enabling *cis*-regulatory sequences for future

investigations. TFs binding to these sites and their target genes should be identified unambiguously in both species. The effect of loss and increased frequency of these TFBS on strength of the CAM phenotype should be investigated.

Co-expression of ABA-responsive TFs, CAM genes and genes of carbohydrate metabolism

To further decipher the transcriptional regulation of CAM, TO-GCN analysis was performed. TO-GCN was developed specifically for time-series data to identify co-expressed pairs of a TF and target genes (encoding either another TF or a structural gene). The resulting network for ABA treatment in *T. triangulare* comprised 26,971 nodes. We explored a subset of the network at the "first neighbour" level focusing on ABA/stress response TFs and an unresolved (*i.e.* without a 1:1 relationship between Arabidopsis and *T. triangulare*) ortholog of *MYB-RELATED* TFs regulating the circadian oscillator. Components of ABA signalling known from Arabidopsis were part of the network as expected for ABA response. These included a *SnRK2*, a PP2C transcript (orthologous to *HAI* and *ABI* proteins) and a *PYL* transcript (Fig. 7; Leung *et al.*, 1997; Fujita *et al.*, 2009; Antoni *et al.*, 2011). Several genes encoding components of the CAM pathway and carbohydrate metabolism as well as transporters were co-expressed with these TFs (Fig. 7).

In Arabidopsis, *HB6* transcript level increases in response to both exogenous ABA and water limitation and acts as a negative regulator of ABA signalling (Söderman *et al.*, 1999; Himmelbach *et al.*, 2002). HB6 was shown to interact physically with ABI1, a PP2C in Arabidopsis (Himmelbach *et al.*, 2002), transcripts of accumulate in response to ABA in Arabidopsis (Goda *et al.*, 2008). Similarly, ABA induced rapid accumulation of several *HB6* transcripts, *ABI1* as well as other PP2C transcripts in *T. triangulare* (Fig. S5A, E, F). Assuming that (some) molecular aspects of ABA signalling are conserved between Arabidopsis and *T. triangulare*, we speculate that recruitment of additional protein-protein and/or DNA-protein interactions resulted in differentiation between stress signalling and a signalling cascade leading to CAM induction, both mediated by ABA. In this scenario, an imaginable evolutionary path towards differentiation of ABA signalling between stress response and CAM induction in a facultative CAM species is via acquisition of respective TFBS in promoters of CAM genes and/or regulators of CAM induction ("CAM switches"). Based on the TO-GCN, HB6 could regulate expression of *BCA, PPC, PPDK* and *NADP-ME* (Fig. 7), enzymes involved both in carboxylation and decarboxylation part of the CAM pathway as well as in PEP regeneration.

In Arabidopsis, physical interaction between HB6 and MATH-BTB DOMAIN PROTEINS BPM3 and BPM6 was confirmed. Both BPM proteins act as substrate-specific adapters of an E3 ubiquitin-protein ligase complex, thus regulating proteasome-mediated protein degradation, including degradation of PP2Cs (Julian *et al.*, 2019). Orthologs of *E3 UBIQUITIN PROTEIN LIGASEs* were present in the analysed part of the co-expression network (Fig. 7), which agrees with regulated protein degradation occurring during ABA response proposed before (Wu *et al.*, 2016*b*). We hypothesize that protein degradation also plays an important regulatory function during CAM induction to rapidly prevent competing C_3 protein activities and co-expression of *E3 UBIQUITIN PROTEIN LIGASEs* with several TFs is an encouraging finding for further investigations.

HSFs form a large TF family involved in response to variety of stress stimuli with a considerable crosstalk between them (Swindell *et al.*, 2007). ABA-induced transcript accumulation of several members

was observed both in Arabidopsis and willow *Salix suchowensis* (Goda *et al.*, 2008; Zhang *et al.*, 2015). ABA-responsive *HSF* transcripts of *T. triangulare* were orthologous to *HSFA5, HSFB1* or *HSFC1*, the latter identified together with *HSFA2, NF-YA9* and *NF-YB3* as a possible transcriptional regulator of CAM induction in our previous work based solely on extent of transcript accumulation in response to ABA (Maleckova *et al.*, 2019). In the TO-GCN, *HSF* orthologs showed connections to *GPT* and *VGT1* sugar transporters (Fig. 7). GPT mediates export of glucose-6-phosphate resulting from starch degradation from chloroplast and its transcript levels as well as activity change diurnally in *M. crystallinum* (Häusler *et al.*, 2000). Since transporters play an essential role in maintaining metabolite flow through the CAM cycle, it is imaginable that sugar transporters are among the early responders during CAM induction and possibly controlled directly via ABA-responsive TFs. In hexose-storing CAM species, the vacuole serves as a transitory storage site of soluble sugars besides accumulating malic acid at night (Holtum *et al.*, 2005). This is not the case for the starch-storing species *T. triangulare*, but VGT1 is important for turgor maintenance (Aluri and Büttner, 2007). Flexibility in sugar usage was demonstrated for CAM species *Aechmea* 'Maya' and *M. crystallinum* (Ceusters *et al.*, 2009; Taybi *et al.*, 2017), which is also imaginable during rapid CAM induction when sufficient PEP pool has to be generated first.

Transcripts encoding enzymes of starch turnover were co-expressed with stress/ABA-responsive TFs (Fig. 7). Transcripts of *STARCH SYNTHASE* accumulated, especially overnight (Fig. S5H), while transcripts of a starch degrading *BMY* were depleted (Fig. S5I). Not only starch metabolism but also glycolysis are subjected to diurnal regulation both in C_3 and CAM species (Geiger and Servaites, 1994; Holtum *et al.*, 2005; Borland *et al.*, 2016), and thus enzymes of both pathways are putative targets for regulation during CAM induction. The only glycolytic enzyme present in the analysed portion of the TO-GCN was *PFK* (Fig. 7). Besides supporting carbon flow via the CAM pathway, carbohydrate metabolism is central to maintenance, growth and reproduction of all living organisms (Kooke and Keurentjes, 2012; Borland *et al.*, 2016; Zakhartsev *et al.*, 2016; Sakr *et al.*, 2018). Due to its central role in primary metabolism, it is bidirectionally regulated with other metabolic pathways, which include the circadian clock, responsiveness to stressors, and phytohormone signalling (León and Sheen, 2003; Satake *et al.*, 2015; Zanella *et al.*, 2016). For example, starch degradation provides carbon backbones for proline synthesis under osmotic stress conditions (Zanella *et al.*, 2016) and BAM1 mediates starch degradation in an ABA-dependent manner during infection with *Plectosphaerella cucumerina* (Gamir *et al.*, 2018). For a successful engineering of inducible CAM, it is therefore required to further dissect the metabolic aspects of general stress response and events specific for CAM induction.

Taken together, motif analysis of promoter sequences of DEGs provided evidence that the coordinated transcriptional regulation during CAM induction might be achieved through the presence of certain TFBS in promoters of target genes. Independently, TO-GCN suggested that even a limited number of TFs might be sufficient to target components of diverse metabolic pathways including core CAM enzymes and enzymes and transporters of carbohydrate metabolism.

Promoters of several CAM genes, genes of carbohydrate metabolism and transporters are enriched for motifs specific to T. triangulare

TFBS enriched in promoters of genes with similar expression patterns could be identified in *T. triangulare* and TO-GCN revealed co-expression of ABA/stress-responsive TFs with several

components of CAM and carbohydrate metabolism. It was thus interesting to investigate, whether some aspects of the observed transcriptional responsiveness to exogenous ABA evolved specifically in the facultative CAM species *T. triangulare*. In addressing this point, we compared TFBS present in gene promoters of species utilizing various carbon assimilation strategies. However, two constraints arose: often unclear orthology relationships among proteins belonging to the same orthogroup and limited number of (facultative) CAM species with genome assemblies available. For these reasons, promoters of all orthologs in a given orthogroup were analysed and the comparison of *T. triangulare* promoters was restricted to the following species: C_3 model species Arabidopsis, more closely related C_3 species *B. vulgaris* and *S. oleracea*, C_3 monocot *H. vulgare*, C_4 species *Z. mays*, obligate CAM monocots *A. comosus* and *P. equestris*, obligate CAM dicots *K. fedtschenkoi* and *K. laxiflora*, and facultative CAM monocot *D. catenatum*.

Analysed orthogroups included those present in the depicted portion of the TO-GCN (Fig. 7) and those expected to play a role in the core CAM cycle or related reactions (*i.e.* carbohydrate metabolism and transporters). TFBS frequent in promoters of *T. triangulare* but not in promoters of orthologous genes of the other species were indeed found frequently and moreover, several cases were identified, where a single TF or a group of closely related TFs could possibly regulate expression of several target genes. This was the case for motifs recognized by FRS9, which were frequent in *MDH, FBA* and *ARF* promoters (Fig. 8A, B, E). FRS9 was identified as a negative regulator of PHYTOCHROME B (phyB) signalling in Arabidopsis (Lin and Wang, 2004). In obligate CAM species *K. fedtschenkoi (Bryophyllum)* phytochrome was identified as the major photoreceptor in entrainment of rhythmic CO_2 uptake (Harris and Wilkins, 1978). Given that phyB signalling plays a similar role also in *T. triangulare*, FRS9 might serve as a hub to coordinate complex metabolic re-programming during CAM induction by contributing to regulation of CO_2 uptake and at the same time targeting CAM cycle itself, tightly connected glycolysis and even ARF-dependent signalling.

NF-Y TFs act as trimers of NF-YA, NF-YB and NF-YC subunits. Several of them regulate flowering and other developmental processes, NF-YA and NF-YB TFs also play a role in drought tolerance as shown also for the NF-YA2-NF-YB2/5/6-NF-YC10 complex (Nelson *et al.*, 2007; Li *et al.*, 2008; Siefers *et al.*, 2009; Zhao *et al.*, 2017; Lian *et al.*, 2018). Transcript accumulation of *NF-YB* and *NF-YC* orthologs in ABA-treated leaves (Fig. S5C, D), co-expression of an *NF-YC* ortholog with numerous other transcripts in response to ABA (Fig. 7), and an enrichment of likely NF-YB4 binding sites in *PPC* and *ABCB* promoters in *T. triangulare* (Fig. 8A, D; Tab. S9) all point at the role of NF-Y TFs in the early response to ABA and it should be investigated whether and how they also contribute to CAM induction. Several ABCB transporters mediate cellular transport of auxin in Arabidopsis (Cho and Cho, 2013; Xu *et al.*, 2014*b*). Even though 2,4-dichlorophenoxyacetic acid treatment did not induce CAM in *M. crystallinum* (Dai *et al.*, 1994) and significant contribution of auxin to CAM induction upon water withdrawal in *T. triangulare* was excluded (Brilhaus *et al.*, 2016), *ARF2*, *ABCB* and auxin efflux carrier *ETHYLENE INSENSITIVE ROOT 1 (EIR1)* orthologs were integrated in the TO-GCN (Fig. 7). Besides that, *MONOPTEROS/AUXIN RESPONSE FACTOR 5 (MP/ARF5)* was identified as a candidate "CAM switch" based on a degree of ABA-induced change in transcript abundance (Maleckova *et al.*, 2019). In the work by Abraham *et al.* (2016) comparing expression profiles between Arabidopsis and *Agave americana*, *ARF4* was identified as a candidate TF regulating largely inversed gene expression between

the two species. It cannot be excluded that ABA signalling interacts with some components of auxin signalling cascade – as also suggested by enrichment of HSF TFBS in *ARF* promoters (Fig. 8E) – but whether such crosstalk is essential for CAM induction remains to be answered. It remains a speculation at the current state of our knowledge but the involvement of *ABCB14*, a malate importer localized in the guard cells, during CAM induction is an imaginable scenario. Regulation of stomatal behaviour in C_3 plants is understood in detail, revealing the importance of light signals. In contrast, regulation underlying inversed stomatal behaviour of CAM species remains elusive (Tallman *et al.*, 1997; Grams and Thiel, 2002). One possibility is control of stomatal aperture by metabolite levels upon initial input from the circadian clock and malate is a promising signalling molecule in this scenario as it can act as a sensor of internal CO_2 concentration (Lee, 2010; Borland *et al.*, 2014).

Evidence for the role of circadian clock in CAM induction

Thanks to the endogenous circadian clock, metabolic processes can be regulated relative to the anticipated light/dark cycle, including for example leaf movement, hypocotyl elongation, diurnal patterns of transcript abundance of stress-responsive genes, and light/dark cycle of starch turnover. In CAM species, the clock is crucial for synchronization of carboxylating and decarboxylating reactions to avoid futile cycling (Wilkins, 1992; Eriksson and Millar, 2003; Hartwell, 2005; Weise *et al.*, 2011). In the context of CAM induction, two possibly competing scenarios are imaginable. TFs of the circadian oscillator could have acquired transcriptional control over additional targets in (facultative) CAM species, leading to shifted gene expression. Alternatively, the TFs of the circadian oscillator could regulate similar genes in both C_3 and (facultative) CAM species, but the expression of clock genes themselves could be shifted in phase. The first option is supported by barely changed transcript levels of the clock genes both in drought- and ABA-induced CAM even though both treatments induced pronounced changes of gene expression (Brilhaus *et al.*, 2016; Maleckova *et al.*, 2019). Transcriptome analysis of *M. crystallinum* exposed to high salinity revealed stability of the circadian oscillator under stress conditions (Boxall *et al.*, 2005). In pineapple, rhythmically expressed gene cluster containing CAM enzymes was enrichment for the CCA1 binding site and sequence of the evening element (Wai *et al.*, 2017). Both studies thus provide additional evidence favouring the first scenario. The present work provides evidence for enrichment of clock-associated TFBS in several promoter sequences of the facultative CAM species *T. triangulare*.

A *MYB-RELATED* transcript of the circadian oscillator was co-expressed with several core CAM enzymes (Fig. 7) and clock-related motifs were enriched in genes up-regulated in response to ABA in *T. triangulare* as well as in several orthogroups in the interspecies comparison of promoter sequences. Unlike ABA- and stress-responsive TFs, the *T. triangulare MYB-RELATED* ortholog was low in abundance and somewhat down-regulated in ABA-treated leaves at the last time point (Fig. S5G), but formed co-expressed pairs with *MDH, PPC* and *PPDK* orthologs (Fig. 7). Connection between the clock and *PPC* transcripts in the TO-GCN was further confirmed in the interspecies comparison of promoter sequences, which revealed enrichment of A(C/T)AAAAGA motif in *PPC* promoters of *T. triangulare* (Fig. 8A). However, the clock transcript included in the TO-GCN was one of 17 distinct transcripts encoded by a single gene, which definitely requires further confirmation. If the unusually high number of splice variants is confirmed, it will be interesting to dissect their individual contribution to maintenance of the circadian rhythm and explore, whether any of these splice variants evolved specifically as a regulator of CAM induction.

Besides *PPC* promoters, clock-associated motifs were also frequent in *ABCB, SWEET* and *ARF* promoters of *T. triangulare* (Fig. 8). An intriguing possibility is the acquisition of clock-related TFBS in the *ABCB14* promoter, a malate importer of the guard cells contributing to turgor regulation and thus stomatal aperture. For CAM species, it was proposed that stomatal behaviour is controlled by metabolic status (Lee, 2010) or alternatively, by fluctuation in internal [CO_2] (Cockburn *et al.*, 1979). Both scenarios however assume already established CAM-specific cycling of metabolites. An independent regulatory mechanism, clock and/or ABA dependent, might be needed in facultative CAM species, since CAM pathway can be induced very rapidly.

In promoters of ABA-responsive genes in *T. triangulare*, TATTTT(G/C)A motif was frequent among up-regulated genes. Candidate TFs likely binding to this sequence include LCL1, EPR1/RVE7, LHY1, RVE1 or RVE8 (Fig. 6C, Tab. S8). These TFs regulate the expression of other clock components via a mechanism of interlinked transcriptional feedback loops (Kuno *et al.*, 2003; Farinas and Mas, 2011; Rawat *et al.*, 2011; Shalit-Kaneh *et al.*, 2018), but there is also accumulating evidence on their regulatory role in other processes. For example, RVE8, in addition to showing rhythmical transcript fluctuation, rapidly responds to heat stress and regulates expression of downstream TFs, such as *DEHYDRATION-RESPONSIVE ELEMENT BINDING PROTEIN 2A (DREB2A)* and *NF-YC2*. It also directly controls expression of two drought-induced ERF domain TFs, ERF53 and ERF54 (Farinas and Mas, 2011; Hsieh *et al.*, 2013; Li *et al.*, 2019). Given that CAM induction required adjustment to the circadian clock, RVE8 would be a good candidate to act as a mediator between drought stress response and the circadian oscillator.

Our results as well as the findings by Wai *et al.* (2017) are supportive of clock-regulated expression of *PPC* in CAM species. However, the studies performed with RNAi lines of *K. fedtschenkoi* and *K. laxiflora,* both obligate CAM species, suggest that the transcriptional regulation of *PPC* might be more complex. *PPCK* silencing affected diurnal fluctuation of *CCA1, PRR7, PRR37* and *PRR9* transcripts under free-running conditions (Boxall *et al.*, 2017). Similarly, loss of PPC1 (*i.e.* the CAM-associated isoform) activity resulted in up-regulation of *PRR7* and *PRR9*, and down-regulation of *RVE1-LIKE* and *EPR1* specifically toward the end of the dark period (Boxall *et al.*, 2020). It should be examined whether (some) core CAM genes are incorporated in transcriptional feedback loops of the circadian oscillator in CAM species and whether the clock regulation differs between obligate and facultative CAM species.

Last but not least, bidirectional regulation between ABA and clock has evolved as demonstrated by diurnally fluctuating levels of endogenous ABA (Novakova *et al.*, 2005; Hanano *et al.*, 2006). It cannot be excluded that some ABA cycling is also preserved in CAM species, thus contributing in part to maintenance of proper CAM cycling once the CAM pathway was fully activated or even in obligate CAM species, for example contributing to inversed stomatal behaviour. Besides the circadian clock, ABA signalling also interacts with light signalling. For example, ELONGATED HYPOCOTYL 5 (HY5), a positive regulator of light signalling, is a positive regulator of ABI5, a TF of the ABA signalling pathway (Chen *et al.*, 2008) and, a convergent amino acid exchange in one of its bZIP domains was identified in *K. fedtschenkoi* (Yang *et al.*, 2017). Only recently, physical interaction between GBF3 and HY5 was demonstrated (Kurihara *et al.*, 2020). Heterodimerisation of ABA- and light-responsive TFs could form a signalling hub to connect the two pathways or even change the binding specificity of both interaction partners (Xu *et al.*, 2014*a*). For CAM, this might be relevant especially during the induction phase when

numerous metabolic processes must switch to a diurnal pattern which is basically a reversal of the C_3 metabolic activities.

lncRNA are co-expressed with protein-coding transcripts

Besides protein-coding transcripts, candidate lncRNAs were quantified as well, revealing that 23.5% of mapped lncRNAs showed significantly altered abundance in response to ABA (Fig. 9). The proportion of ABA-responsive lncRNAs was comparable to the proportion of ABA-responsive protein-coding transcripts (21%) and moreover, both protein-coding transcripts and lncRNAs showed highly similar expression profiles. It has been shown that lncRNAs act as both *cis*- and *trans*-regulators of gene expression (Heo *et al.*, 2013; Wang *et al.*, 2019*a*). Diurnal cycling of lncRNA transcript was confirmed in green leaf tissue in pineapple and candidate lncRNAs possibly regulating expression of CAM-related genes were identified (Bai *et al.*, 2019). It will be of interest to further investigate potential interaction between lncRNA and mRNA in *T. triangulare* and evaluate the importance of lncRNA in the process of CAM induction.

Analysis of protein-coding sequences and changes associated with CAM

Phylogenetic placement of Talinum triangulare

Based on protein sequence comparison within individual orthogroups, a species tree was obtained (OrthoFinder analysis). In accordance with its confirmed phylogenetic placement based on plastidic and nuclear rDNA marker genes and comparative cytogenetics (Brockington *et al.*, 2009; Marinho *et al.*, 2019), *T. triangulare* was included within the "ACPT" clade of *Caryophyllales* (*Anacampserotaceae, Cactaceae, Portulacaceae* and *Talinacea*). *T. triangulare* was placed as a sister group to all *Portulaca* spp. contained in the analysis, including C_4-CAM species *P. oleracea* (Fig. 4A). Further relationships within the *Caryophyllales* family were in agreement with a phylogenetic tree obtained within a transcriptome sequencing study by Yang *et al.* (2015).

Orthogroups shared by facultative CAM species include genes encoding ABA-responsive transcripts in Talinum triangulare

It was proposed repeatedly that evolution of CAM photosynthesis in a C_3 ancestor species requires a mere adjustment of existing enzymatic activities and metabolite flows in a time-dependent manner (West-Eberhard *et al.*, 2011; Bräutigam *et al.*, 2017; Heyduk *et al.*, 2019; Wai *et al.*, 2019). The ways to achieve such metabolic re-programming would include changes in transcriptional control as discussed above and likely also modifications of enzymatic activities (e.g. by altering sensitivity to intermediates of the metabolism acting as inhibitors). However, truly facultative CAM species present a special case of extreme metabolic flexibility, regulation of which might indeed require emergence of new proteins acting as „CAM switches". Taking advantage of the increasing number of genomics resources, we compared proteomes of species including C_3, C_4 and C_3-C_4 intermediates as well as obligate and facultative CAM species (Tab. S1) with the aim to identify proteins (orthogroups) unique to facultative CAM species. There were indeed 2,100 such orthogroups (Fig. 9A), including 1,538 orthogroups shared by *T. triangulare* and at least one other facultative CAM species (Fig. 9B). The proteins possibly specific to facultative CAM species included 377 *T. triangulare* proteins, transcript levels of which were responsive to ABA.

Only a small fraction of these DEGs carried a recognizable protein domain in the translated CDS sequences, but these included several promising candidates acting as "CAM switches": an ABA/WDS induced protein in the early ABA-responsive up-regulated cluster 3, a putative Ser/Thr-protein kinase and a protein with an F-box-like domain in cluster 1 (down-regulated transcripts) and in up-regulated clusters 5 and 7, a basic helix-loop-helix (bHLH) protein and a protein with SANT/Myb domain, respectively (Tab. S10/11?). In *Pinus taeda*, a small WDS gene family was identified, transcripts of which accumulate especially in roots upon water withdrawal, and homologous genes were reported in other species (Padmanabhan et al., 1997). It can be speculated that WDS orthologs are present also in *T. triangulare* and involved either in drought stress regulation or even connect ABA signalling to CAM induction in response to drought. Many chromatin-remodelling complexes contain a SANT domain (Boyer et al., 2002) and F-box-containing proteins, genes of which are abundant in Arabidopsis genome, are involved in regulation of diverse cellular processes ranging from cell cycle control to transcriptional regulation and signal transduction (Kuroda et al., 2002). It is thus imaginable that that either of these DEGs contributes to transcriptional or post-translational changes associated with CAM induction and are worth further investigation.

PEPC and PPDK protein sequence comparison

Whole genome duplications and duplications of individual genes, providing a basis for neofunctionalization, occurred in at least some CAM lineages (Christin *et al.*, 2014; Silvera *et al.*, 2014; Yang *et al.*, 2017). PEPC and PPDK are two examples of C_4/CAM enzymes with detailed understanding of their activity, amino acid substitutions differentiating C_3 and C_4/CAM isoforms and sites of post-translation modifications (Westhoff and Gowik, 2004; Yang *et al.*, 2017). W

Whole-genome duplications as well as duplications of individual genes have occurred in all lineages of angiosperms and when retained in the genome, the duplicated copies are a source for acquisition of evolutionary novelties (Soltis *et al.*, 2003; Moriyama and Koshiba-Takeuchi, 2018). In the context of CAM evolution, neofunctionalization could lead to emerge of CAM-specific isoforms from their C_3 or non-photosynthetic ancestors as documented on the example of C_4 PEPC isoforms (Westhoff and Gowik, 2004) and CAM-specific *PPC* lineage in both eudicots (*e.g. Caryophyllales)* and monocots *(e.g. Oncidiinae)* (Christin *et al.*, 2014; Silvera *et al.*, 2014). This is however not the case for pineapple (Ming *et al.*, 2015). Preliminary analysis of number of orthologs of several CAM genes did not reveal any extremely expanded gene families in *T. triangulare* as it was the case of *K. laxiflora* or *Z. mays* (Fig. S6).

PEPC is one of the key enzymes of both C_4 and CAM and number of sites subjected to phosphorylation or associated with a certain photosynthetic type have been identified. These sites were therefore examined in predicted PEPC sequences of *T. triangulare.* Phosphorylation of Ser15 (numbering based on CAA33317 of *Z. mays*) desensitifies PEPC to its inhibitor malate (Nimmo, 2000). Exchange of Ser15 for Thr (*e.g.* in PB.6485_4; Fig. 11B) occurred frequently in *T. triangulare*, but it may not dramatically affect the phosphorylation status at this site. In cases when substitution with Cys, Leu or Trp occurred, the entire N-terminus differed largely from the remaining sequences of *T. triangulare* and other species (Fig. S7B). These proteins however contained Thr at the site homologous to His14 of *Z. mays* and five to six Ser and up to two Thr at the 28 sites following directly after the site homologous to Ser15 (Fig. S7B). It cannot be excluded that one or more of this alternative Ser/Thr sites are targeted by PPCK.

Gly890 also contributes to reduced inhibition by malate (Paulus *et al.*, 2013). Full-length PEPCs of *T. triangulare* contained Arg at this position, while several isoforms were truncated. It will be interesting to investigate the sensitivity to malate of these truncated isoforms.

Ser780 is not subjected to phosphorylation but determines PEP saturation kinetics. Ser780 is present in all C_4 enzymes analysed so far but replaced by Ala in C_3 and CAM PEPCs (Bläsing *et al.*, 2000; Svensson *et al.*, 2003). This was indeed the case for PEPC1/PEPC3 orthologs, where Ala was present even in several C_4 species. Among PEPC2 orthologs, Ser780 so far associated with C_4 isoforms only was present at the corresponding site in several of *T. triangulare* proteins, which is consistent with PEPC sequences obtained with Illumina sequencing (Brilhaus *et al.*, 2016), in C_4-CAM *P. oleracea* but not in relatively closely related *M. crystallinum* or any obligate CAM species analysed.

Predicted CAM-isoform of *K. fedtschenkoi* PEPC2 contains Asp509 (corresponds to His519 in CAA33317 of *Z. mays*). Owing to the negative charge of Asp and its localisation in proximity of the active site, Asp509 was predicted to mimic the effect of negatively charged glucose-6-phosphate, an allosteric activator of PEPC (Yang *et al.*, 2017). All *T. triangulare* PEPC sequence however contained ancestral His or Arg (Fig. 11B). Moreover, the homologous sites across a number of species turned out to be highly variable. Only Glu in sequence of CAM species *Cissus quadrangularis* and C_4 *Portulaca amilis* could have the same effect as Asp in PEPC2 of *K. fedtschenkoi* and *Phalaenopsis equestris* ortholog (Yang *et al.*, 2017; Fig. S4). Substantial diversity among PEPCs of various species exists (Fig. 11B, Fig. 13, Fig. S4), including obligate *vs* facultative CAM species, monocots *vs* dicots.

Several of the differences between PEP1 of *Z. mays* (C_4 isoform) and PEPC1/PEPC3 of *T. triangulare* were also preserved in C_4-CAM species *P. oleracea*, including presence of truncated isoforms lacking both Ser780 and Gly890. Truncated isoforms were also detected in *Agave deserti*, C_4 *Portulaca amilis* and *Aerva lanata*, facultative CAM *Basella alba* and *Vittaria lineata* but not it *Kalanchoë* orthologs (Fig. S6B). Since many of these predictions are based on shotgun assemblies, these observations need to be interpreted with caution. On the other hand, there was a marked difference between abundance of *PPC1, PPC3* and *PPC2* orthologs of *T. triangulare* (Fig. S7). It will be interesting to investigate whether truncated and less abundant isoforms act as first responders during CAM induction and high abundant *PPC2* isoform evolved as CAM-specific isoform with nocturnal activity in this species.

The action of PPDK is essential for PEP regeneration in both C_4 and CAM species, but as opposed to PEPC, it is much more conserved at the sites crucial for regulation of its activity. All analysed full-length sequences of *T. triangulare* carried Thr309, Ser506, Thr527 and Ser528 subjected to phosphorylation in *Z. mays*, as well as Gly525 (Fig. 11B; numbering based on CAAA33054 of *Z. mays*) important for PPDK-RP mediated phosphorylation of Thr257 and Ser528 (Chen *et al.*, 2014). Similar to PEPC, alternative splice variants lacking these sites of port-translation regulation were predicted for both PPDK orthologs of *T. triangulare*. The full-length variants however showed markedly higher abundance and accumulated at the transcript level in the early morning (Fig. 12C).

The similarity between PEPC2 orthologs of *T. triangulare* and *P. oleracea* as well as their clear distinction from other species was also confirmed by MDS analysis of the protein sequences and MDS was therefore applied on several other proteins of the CAM pathway to identify candidates with similar degree of sequence differentiation for future investigations. BCA1 and BCA2 orthologs of CAM

monocots differed from other species (Fig. 13D). Obligate CAM species are frequently strong CAM species (Winter, 2019) and increased nocturnal CO_2 fixation might indeed need adjustments of BCA activities, which however remain to be characterized. Chloroplastic NADP-ME4 differed between *T. triangulare* and obligate CAM monocots and dicots (Fig. 13F). Pineapple uses PEPCK as the decarboxylating enzyme instead on ME (Dittrich *et al.*, 1973) and *K. fedtschenkoi* is a NAD-ME species (Dever *et al.*, 2015), making it likely that NADP-MEs of these species are under different selection pressure than NADP-ME of *T. triangulare*. While PPDK was highly conserved (Fig. 12B, Fig. 13G), large diversity was observed for RP1 (Fig. 13H). Thus, any adjustments to PEP metabolism required for the CAM pathway might be achieved at the level of RP instead of altering PPDK itself.

Apart from PEPC, more proteins should be investigated for changes associated with certain modes of photosynthesis, while considering the diversity of plant lineages. Besides core CAM enzymes, transporters, enzymes of carbohydrate metabolism as well as proteins with regulatory functions should be considered and thus, an efficient approach to identify candidates for thorough investigations is needed. We propose MDS as one possibility and demonstrated this approach on CAM genes. In this analysis, BCA1 and BCA2 orthologs of obligate CAM monocots formed a separate cluster (Fig. S8D). Similarly, NADP-ME4 orthologs formed separate clusters for obligate CAM species and the facultative CAM species *T. triangulare* (Fig. S7F). Interestingly, NADP-ME1 orthologs could be identified in only a limited number of species, which did not include any obligate CAM species. *T. triangulare* orthologs were however clearly separated from orthologs of both C_3 and C_4 species (Fig. S8E). The high level of PPDK sequence conservation across all species considered was also apparent from the MDS analysis (Fig. S8G). In contrast to PPDK, considerable level of sequence differentiation was detected among orthologs of RP1 (Fig. S8H). Employing similar approach and considering effects of identified convergent substitutions on a protein's structure and kinetic properties in case of enzymes, it will be possible to faster identify isoforms suitable for an efficient CAM biodesign.

Conclusions and perspective

We generated a genome assembly of *T. triangulare,* to the best of our knowledge the first facultative CAM dicot sequenced. We improved the transcriptome reference for this species by employing long-read sequencing. Both assemblies are a rich source of information to deepen our understanding of molecular mechanisms controlling CAM induction in *T. triangulare* and contribute to comparative studies focusing on CAM evolution. We demonstrated the importance of interspecies comparisons in identifying *cis*-regulatory elements possibly recruited for regulation of CAM induction. Interspecies comparison of PEPC sequences revealed a large diversity of splice variants and amino acid substitutions at sites associated subjected to phosphorylation or associated with C_4/CAM isoforms. When the diversity of plants species and many forms of CAM is considered, amino acid substitutions associated with certain CAM forms or lineages might be identified and isoforms of key CAM compared to identify the best candidates for CAM engineering. Recent progress in the field of gene editing should make it possible to test such candidate genes experimentally.

References

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

Aluri S, Büttner M. 2007. Identification and functional expression of the *Arabidopsis thaliana* vacuolar glucose transporter 1 and its role in seed germination and flowering. Proceedings of the National Academy of Sciences of the United States of America **104**, 2537–2542.

Antoni R, Rodriguez L, Gonzalez-Guzman M, Pizzio GA, Rodriguez PL. 2011. News on ABA transport, protein degradation, and ABFs/WRKYs in ABA signaling. Current Opinion in Plant Biology **14**, 547–553.

Astley HM, Parsley K, Aubry S, Chastain CJ, Burnell JN, Webb ME, Hibberd M. 2011. The Pyruvate, Orthophosphate Dikinase Regulatory Proteins of Arabidopsis Are Both Bifunctional and Interact With the Catalytic and Nucleotide-Binding Domains of Pyruvate, Orthophosphate Dikinase. The Plant Journal **68**, 1070–1080.

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

Bailey TL. 2011. DREME: Motif discovery in transcription factor ChIP-seq data. Bioinformatics **27**, 1653–1659.

Baquar SR. 1986. Cytotaxonomic studies of the family *Portulacaceae* from Nigeria. Kromosomo **2**, 1255–1262.

Beaulieu JM, Leitch IJ, Patel S, Pendharkar A, Knight CA. 2008. Genome size is a strong predictor of cell size and stomatal density in angiosperms. New Phytologist **179**, 975–986.

Beier S, Himmelbach A, Colmsee C, *et al.* 2017. Data Descriptor: Construction of a map-based reference genome sequence for barley, *Hordeum vulgare* L. Scientific Data **4**, 170044.

Benzing DH, Givnish TJ, Bermudes D. 1985. Absorptive Trichomes in *Brocchinia reducta (Bromeliaceae)* and Their Evolutionary and Systematic Significance. Systematic Botany **10**, 81–91.

Bi Y, Xie H. 2015. C_3 Vegetation Mapping and CO_2 Fertilization Effect in the Arid Lower Heihe River Basin, Northwestern China. Remote Sensing **7**, 16384–16397.

Li B, Gao Z, Liu X, Sun D, Tang W. 2019. Transcriptional profiling reveals a time-of-day-specific role of REVEILLE 4/8 in regulating the first wave of heat shock–induced gene expression in Arabidopsis. Plant Cell **31**, 2353–2369.

Bianconi ME, Dunning LT, Moreno-villena JJ, Osborne CP, Christin P. 2018. Gene duplication and dosage effects during the early emergence of C 4 photosynthesis in the grass genus *Alloteropsis*. Journal of Experimental Botany **69**, 1967–1980.

Bläsing OE, Westhoff P, Svensson P. 2000. Evolution of C4 Phosphoenolpyruvate Carboxylase in *Flaveria*, a Conserved Serine Residue in the Carboxyl-terminal Part of the Enzyme Is a Major Determinant for C_4-specific Characteristics. Journal of Biological Chemistry **275**, 27917–27923.

Bodenhofer U, Bonatesta E, Horejš-Kainrath C, Hochreiter S. 2015. Msa: An R package for multiple sequence alignment. Bioinformatics **31**, 3997–3999.

Borland AM, Griffiths H, Hartwell J, Smith JAC. 2009. Exploiting the potential of plants with crassulacean acid metabolism for bioenergy production on marginal lands. Journal of Experimental Botany **60**, 2879–2896.

Borland AM, Guo HB, 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, Weston DJ, Schlauch KA, Tschaplinski TJ, Tuskan GA, Yang X, Cushman JC. 2014. Engineering crassulacean acid metabolism to improve water-use efficiency. Trends in Plant Science **19**, 327–338.

Bowman JL, Kohchi T, Yamato KT, *et al.* 2017. Insights into Land Plant Evolution Garnered from the *Marchantia polymorpha* Genome. The Cell **171**, 287–304.

Boxall SF, Dever L V, Knerova J, Gould PD, Hartwell J. 2017. Phosphorylation of Phosphoenolpyruvate 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.* The Plant Cell **29**, 2519–2536.

Boxall SF, Foster JM, Bohnert HJ, Cushman JC, Nimmo HG, Hartwell J. 2005. Conservation and

divergence of circadian clock operation in a stress-inducible Crassulacean acid metabolism species reveals clock compensation against stress. Plant Physiology **137**, 969–982.

Boxall SF, Kadu N, Dever L V, Kneřová 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.

Bräutigam A, Eisenhut M, Schlüter U, Gowik U. 2017. On the evolutionary origin of CAM photosynthesis. Plant Physiology **49**, pp.00195.2017.

Brilhaus D, Bräutigam A, Mettler-Altmann T, Winter K, Weber APM. 2016. Reversible Burst of Transcriptional Changes During Induction of Crassulacean Acid Metabolism (CAM) in *Talinum triangulare.* Plant Physiology **170**, 102–122.

Brkljacic J, Grotewold E, Scholl R, *et al.* 2011. *Brachypodium* as a Model for the Grasses: Today and the Future. Plant Physiology **157**, 3–13.

Brock G, Pihur V, Datta S, Datta S. 2008. clValid: An R Package for Cluster Validation. Journal of Statistical Software **25**, 1–22.

Brockington SF, Alexandre R, Ramdial J, Moore MJ, Crawley S, Dhingra A, Hilu K, Soltis DE, Soltis PS. 2009. Phylogeny of the *Caryophyllales sensu lato:* revisiting hypotheses on pollination biology and perianth differentiation in the core *Caryophyllales*. International Journal of Plant Sciences **170**, 627–643.

Brugnoli E, Lauteri M. 1991. Effects of Salinity on Stomatal Conductance, Photosynthetic Capacity, and Carbon Isotope Discrimination of Salt-Tolerant (*Gossypium hirsutum* L.) and Salt-Sensitive (*Phaseolus vulgaris* L.) C_3 Non-Halophytes. Plant Physiology **95**, 628–635.

Cai J, Liu X, Vanneste K, *et al.* 2014. The genome sequence of the orchid *Phalaenopsis equestris*. Nature Genetics **47**, 65–72.

Carpenter E, Matasci N, Ayyampalayam S, *et al.* 2019. Data supporting the 1000 plant (1KP) transcriptomes initiative GigaScience Database.

Ceusters J, Borland AM, Londers E, Verdoodt V, Godts C, De Proft MP. 2009. Differential usage of storage carbohydrates in the CAM bromeliad *Aechmea* 'Maya' during acclimation to drought and recovery from dehydration. Physiologia Plantarum **135**, 174–184.

Chan AP, Crabtree J, Zhao Q, *et al.* 2011. Draft genome sequence of the ricin-producing oilseed castor bean. Nature Biotechnology **28**, 951–956.

Chang C, Bowman JL, Meyerowitz EM. 2016. Field Guide to Plant Model Systems. Cell **167**, 325–339.

Chang YM, Lin HH, Liu WY, *et al.* 2019. Comparative transcriptomics method to infer gene coexpression networks and its applications to maize and rice leaf transcriptomes. Proceedings of the National Academy of Sciences of the United States of America **116**, 3091–3099.

Chen YB, Lu TC, Wang HX, Shen J, Bu TT, Chao Q, Gao ZF, Zhu XG, Wang YF, Wang BC. 2014. Posttranslational modification of maize chloroplast pyruvate orthophosphate dikinase reveals the precise regulatory mechanism of its enzymatic activity. Plant Physiology **165**, 534–549.

Chen H, Zhang J, Neff MM, Hong SW, Zhang H, Deng XW, Xiong L. 2008. Integration of light and abscisic acid signaling during seed germination and early seedling development. Proceedings of the National Academy of Sciences of the United States of America **105**, 4495–4500.

Cheng C, Krishnakumar V, Chan AP, Thibaud-Nissen F, Schobel S, Town CD. 2017. Araport11: a complete reannotation of the Arabidopsis thaliana reference genome. The Plant Journal **89**, 789–804.

Chiang CP, Yim WC, Sun YH, Ohnishi M, Mimura T, Cushman JC, Yen HE. 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**, 1–18.

Cho M, Cho H-T. 2013. The function of ABCB transporters in auxin transport. Plant Signaling & Behavior **8**, e22990.

Chou ML, Shih MC, Chan MT, Liao SY, Hsu CT, Haung YT, Chen JJW, Liao DC, Wu FH, Lin CS. 2013. Global transcriptome analysis and identification of a CONSTANS-like gene family in the orchid *Erycina pusilla*. Planta **237**, 1425–1441.

Choy-Sin H, Suan WY. 1974. Photosynthesis and Respiration of Ferns in Relation to Their Habitat. American Fern Journal **64**, 40–48.

Christin P-A, Arakaki M, Osborne CP, *et al.* 2014. Shared origins of a key enzyme during the evolution

of C_4 and CAM metabolism. Journal of Experimental Botany **65**, 3609–3621.

Christopher JT, Holtum J. 1996. Patterns of Carbon Partitioning in Leaves of Crassulacean Acid Metabolism Species during Deacidification. Plant physiology **112**, 393–399.

Clouse JW, Adhikary D, Page JT, Ramaraj T, Deyholos MK, Udall JA, Fairbanks DJ, Jellen EN, Maughan PJ. 2016. The Amaranth Genome: Genome, Transcriptome, and Physical Map Assembly. The Plant Genome **9**, 1–14.

Cockburn W, Ting IP, Sternberg LO. 1979. Relationships between Stomatal Behavior and Internal Carbon Dioxide Concentration in Crassulacean Acid Metabolism Plants. Plant Physiology **63**, 1029–1032.

Cooper EA, Brenton ZW, Flinn BS, *et al.* 2019. A new reference genome for *Sorghum bicolor* reveals high levels of sequence similarity between sweet and grain genotypes: implications for the genetics of sugar metabolism. BMC Genomics **20**, 420.

Cushman JC. 1992. Characterization and expression of a NADP-malic enzyme cDNA induced by salt stress from the facultative crassulacean acid metabolism plant, *Mesembryanthemum crystallinum*. European Journal of Biochemistry **208**, 259–266.

Cushman JJC, Bohnert HHJ. 1999. Crassulacean Acid Metabolism: Molecular Genetics. Annual Review of Plant Biology **50**, 305–332.

Cushman JC, Tillett RL, Wood JA, Branco JM, Schlauch KA. 2008. Large-scale mRNA expression profiling in the common ice plant, *Mesembryanthemum crystallinum*, performing C_3 photosynthesis and Crassulacean acid metabolism (CAM). Journal of Experimental Botany **59**, 1875–1894.

Dai Z, Edwards GE, Ku MSB. 1992. Control of Photosynthesis and Stomatal Conductance in *Ricinus communis* L. (Castor Bean) by Leaf to Air Vapor Pressure Deficit. Plant Physiology **99**, 1426–1434.

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.

Davis ASC, Simpson J, Del K, Gil C, Nicholas A, Tongerlo E Van, Castano NH, Dever L V. 2019. Undervalued potential of crassulacean acid metabolism (CAM) for current and future agricultural production. Journal of Experimental Botany **70**, 6521–6537.

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.

DiMario RJ, Clayton H, Mukherjee A, Ludwig M, Moroney J V. 2017. Plant Carbonic Anhydrases: Structures, Locations, Evolution, and Physiological Roles. Molecular Plant **10**, 30–46.

Dittrich P, Campbell WH, Black CC. 1973. Phosphoenolpyruvate Carboxykinase in Plants Exhibiting Crassulacean Acid Metabolism. Plant Physiology **52**, 357–361.

Dohm JC, Minoche AE, Holtgräwe D, *et al.* 2014. The genome of the recently domesticated crop plant sugar beet (*Beta vulgaris*). Nature **505**, 546–549.

Doležel J, Bartoš J. 2005. Plant DNA flow cytometry and estimation of nuclear genome size. Annals of Botany **95**, 99–110.

Doležel J, Bartoš J, Voglmayr H, Greilhuber J. 2003. Nuclear DNA content and genome size of trout and human. Cytometry **51A**, 127–128.

Doležel J, Greilhuber J, Suda J. 2007. Estimation of nuclear DNA content in plants using flow cytometry. Nature Protocols **2**, 2233–2244.

Du Q, Wang L, Yang X, Gong C, Zhang D. 2015. *Populus* endo-β-1,4-glucanases gene family: genomic organization, phylogenetic analysis, expression profiles and association mapping. Planta **241**, 1417–1434.

Emms DM, Kelly S. 2019. OrthoFinder: Phylogenetic orthology inference for comparative genomics. Genome Biology **20**, 1–14.

Eriksson ME, Millar AJ. 2003. The circadian clock. A plant's best friend in a spinning world. Plant Physiology **132**, 732–738.

Farinas B, Mas P. 2011. Histone acetylation and the circadian clock: A role for the MYB transcription factor RVE8/LCL5. Plant Signaling and Behavior **6**, 541–543.

Flavell R. 1980. The Molecular Characterization and Organization of Plant Chromosomal DNA Sequences. Annual Review of Plant Physiology **31**, 569–596.

Fu Y, Li L, Hao S, *et al.* 2017. Draft genome sequence of the Tibetan medicinal herb *Rhodiola crenulata*. GigaScience **6**, 1–5.

Fujita Y, Nakashima K, Yoshida T, *et al.* 2009. Three SnRK2 Protein Kinases are the Main Positive Regulators of Abscisic Acid Signaling in Response to Water Stress in Arabidopsis. Plant and Cell Physiology **50**, 2123–2132.

Galbraith DW, Harkins KR, Maddox JM, Ayres NM, Sharma DP, Firoozabady E. 1983. Rapid flow cytometric analysis of the cell cycle in intact plant tissues. Science **220**, 1049–1051.

Gamir J, Pastor V, Sánchez-Bel P, Agut B, Mateu D, García-Andrade J, Flors V. 2018. Starch degradation, abscisic acid and vesicular trafficking are important elements in callose priming by indole-3-carboxylic acid in response to *Plectosphaerella cucumerina* infection. The Plant Journal **96**, 518–531.

Geiger DR, Servaites JC. 1994. Diurnal regulation of photosynthetic carbon metabolism in C_3 plants. Annual Review of Plant Physiology and Plant Molecular Biology **45**, 235–256.

Giraud E, Ng S, Carrie C, Duncan O, Low J, Lee CP, van Aken O, Harvey Millar A, Murcha M, Whelan J. 2010. TCP transcription factors link the regulation of genes encoding mitochondrial proteins with the circadian clock in *Arabidopsis Thaliana*. Plant Cell **22**, 3921–3934.

Goda H, Sasaki E, Akiyama K, *et al.* 2008. The AtGenExpress hormone and chemical treatment data set: Experimental design, data evaluation, model data analysis and data access. Plant Journal **55**, 526–542.

Grams TEE, Thiel S. 2002. High light-induced switch from C_3-photosynthesis to Crassulacean acid metabolism is mediated by UV-A/blue light. Journal of Experimental Botany **53**, 1475–1483.

Gregory BD. 2018. Shedding some blue light on alternative promoter usage in plants. Proceedings of the National Academy of Sciences of the United States of America **115**, 7654–7656.

Gross SM, Martin JA, Simpson J, Abraham-Juarez MJ, Wang Z, Visel A. 2013. *De novo* transcriptome assembly of drought tolerant CAM plants, *Agave deserti* and *Agave tequilana*. BMC Genomics **14**, 1–14.

Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble WS. 2007. Quantifying similarity between motifs. Genome Biology **8**, R24.1-R24.9.

Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: Quality assessment tool for genome assemblies. Bioinformatics **29**, 1072–1075.

Hanano S, Domagalska MA, Nagy F, Davis SJ. 2006. Multiple phytohormones influence distinct parameters of the plant circadian clock. Genes to Cells **11**, 1381–1392.

Harriet ES. 1946. *Sedum pulchellum*: A Physiological and Morphological Comparison of Diploid, Tetraploid, and Hexaploid Races. Bulletin of the Torrey Botanical Club **73**, 495–541.

Harris FS, Carter ME, Martin CE. 1993. Relationship between ploidy level and CAM in five species of *Talinum (Portulacaceae).* American Journal of Botany **80**, 533–536.

Harris PJC, Wilkins MB. 1978. Evidence of Phytochrome Involvement in the Entrainment of the Circadian Rhythm of Carbon Dioxide Metabolism in *Bryophyllum*. Planta **138**, 271–278.

Hartwell J. 2005. The co-ordination of central plant metabolism by the circadian clock. Biochemical Society Transactions **33**, 945–948.

Häusler RE, Baur B, Scharte J, Teichmann T, Eicks M, Fischer KL, Flügge UI, Schubert S, Weber A, Fischer K. 2000. Plastidic metabolite transporters and their physiological functions in the inducible crassulacean acid metabolism plant *Mesembryanthemum crystallinum*. Plant Journal **24**, 285–296.

Heo JB, Lee YS, Sung S. 2013. Epigenetic regulation by long noncoding RNAs in plants. Chromosome Research **21**, 685–693.

Hernandez-Garcia CM, Finer JJ. 2014. Identification and validation of promoters and cis-acting regulatory elements. Plant Science **217**–**218**, 109–119.

Heyduk K, Burrell N, Lalani F, Leebens-mack J. 2016. Gas exchange and leaf anatomy of a C_3-CAM hybrid, *Yucca gloriosa (Asparagaceae).* Journal of Experimental Botany **67**, 1369–1379.

Heyduk K, Hwang M, Albert V, Silvera K, Lan T, Farr K, Chang T, Chan M, Winter K, Leebens-mack J. 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.

Himmelbach A, Hoffmann T, Leube M, Höhener B, Grill E. 2002. Homeodomain protein ATHB6 is a target of the protein phosphatase ABI1 and regulates hormone responses in Arabidopsis. EMBO Journal

21, 3029–3038.

HMMER. URL http://hmmer.org/

Holtum JAM, Smith JAC, Neuhaus HE. 2005. Intracellular transport and pathways of carbon flow in plants with crassulacean acid metabolism. Functional Plant Biology **32**, 429.

Hsieh EJ, Cheng MC, Lin TP. 2013. Functional characterization of an abiotic stress-inducible transcription factor AtERF53 in *Arabidopsis thaliana*. Plant Molecular Biology **82**, 223–237.

Huson DH, Beier S, Flade I, Górska A, El-Hadidi M, Mitra S, Ruscheweyh HJ, Tappu R. 2016. MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. PLoS Computational Biology **12**, 1–12.

Ibarra-Laclette E, Zamudio-Hernández F, Pérez-Torres CA, *et al.* 2015. *De novo* sequencing and analysis of *Lophophora williamsii* transcriptome, and searching for putative genes involved in mescaline biosynthesis. BMC Genomics **16**, 657.

Jarvis DE, Ho YS, Lightfoot DJ, *et al.* 2017. The genome of *Chenopodium quinoa*. Nature **542**, 307–312.

Jin J, Tian F, Yang DC, Meng YQ, Kong L, Luo J, Gao G. 2017. PlantTFDB 4.0: Toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Research **45**, D1040–D1045.

Jiu S, Xu Y, Wang J, *et al.* 2019. Genome-Wide Identification, Characterization, and Transcript Analysis of the TCP Transcription Factors in *Vitis vinifera*. Frontiers in Genetics **20**, 786.

Jones P, Binns D, Chang HY, *et al.* 2014. InterProScan 5: Genome-scale protein function classification. Bioinformatics **30**, 1236–1240.

Julian J, Coego A, Lozano-Juste J, *et al.* 2019. The MATH-BTB BPM3 and BPM5 subunits of Cullin3-RING E3 ubiquitin ligases target PP2CA and other clade A PP2Cs for degradation. Proceedings of the National Academy of Sciences of the United States of America **116**, 15725–15734.

Jungnick N, Ma Y, Mukherjee B, Cronan JC, Speed DJ, Laborde SM, Longstreth DJ, Moroney J V. 2014. The Carbon Concentrating Mechanism in *Chlamydomonas Reinhardtii*: Finding the Missing Pieces. Photosynthesis Research **121**, 159–173.

Kang YJ, Yang DC, Kong L, Hou M, Meng YQ, Wei L, Gao G. 2017. CPC2: A fast and accurate coding potential calculator based on sequence intrinsic features. Nucleic Acids Research **45**, W12–W16.

Kent WJ. 2002. BLAT — The BLAST-Like Alignment Tool. Genome Research **12**, 656–664.

Kim S, Kang JY, Cho DI, Park JH, Soo YK. 2004. ABF2, an ABRE-binding bZIP factor, is an essential component of glucose signaling and its overexpression affects multiple stress tolerance. The Plant Journal **40**, 75–87.

Kim D, Langmead B, Salzberg1 SL. 2015. HISAT: a fast spliced aligner with low memory requirements Daehwan HHS Public Access. Nature Methods **12**, 357–360.

Kocacinar F, Sage RF. 2003. Photosynthetic pathway alters xylem structure and hydraulic function in herbaceous plants. Plant, Cell and Environment **26**, 2015–2026.

Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. Nature Biotechnology **37**, 540–546.

Kooke R, Keurentjes JJB. 2012. Multi-dimensional regulation of metabolic networks shaping plant development and performance. Journal of Experimental Botany **63**, 3353–3365.

Kool A. 2012. Desert Plants and Deserted Islands: Systematics and Ethnobotany in *Caryophyllaceae*.

Kuno N, Møller SG, Shinomura T, Xu XM, Chua NH, Furuya M. 2003. The Novel MYB Protein EARLY-PHYTOCHROME-RESPONSIVE1 Is a Component of a Slave Circadian Oscillator in Arabidopsis. Plant Cell **15**, 2476–2488.

Kurihara Y, Makita Y, Shimohira H, Matsui M. 2020. Time-Course Transcriptome Study Reveals Mode of bZIP Transcription Factors on Light Exposure in Arabidopsis. International Journal of Molecular Sciences **21**, 1993.

Lara M V., Drincovich MF, Andreo CS. 2004. Induction of a crassulacean acid-like metabolism in the C_4 succulent plant, *Portulaca oleracea* L.: Study of enzymes involved in carbon fixation and carbohydrate metabolism. Plant and Cell Physiology **45**, 618–626.

Lee JS. 2010. Stomatal Opening Mechanism of CAM Plants. Journal of Plant Biology **53**, 19–23.

Lee JT. 2012. Epigenetic regulation by long noncoding RNAs. Science **338**, 1435–1439.

Lee HO, Davidson JM, Duronio RJ. 2009. Endoreplication: polyploidy with purpose. Genes & Development **23**, 2461–2477.

León P, Sheen J. 2003. Sugar and hormone connections. Trends in Plant Science **8**, 110–116.

Leung J, Merlot S, J G. 1997. A*BI1* Genes Encode Homologous Protein Phosphatases 2C Involved in Abscisic Acid Signal Transduction. The Plant cell **9**, 759–771.

Li S. 2015. The *Arabidopsis thaliana* TCP transcription factors: A broadening horizon beyond development. Plant Signaling and Behavior **10**, 1–12.

Li B, Gao Z, Liu X, Sun D, Tang W. 2019. Transcriptional profiling reveals a time-of-day-specific role of REVEILLE 4/8 in regulating the first wave of heat shock–induced gene expression in Arabidopsis. Plant Cell **31**, 2353–2369.

Li G, Jain R, Chern M, *et al.* 2017. The Sequences of 1504 Mutants in the Model Rice Variety Kitaake Facilitate Rapid Functional Genomic Studies. The Plant Cell **29**, 1218–1231.

Li WX, Oono Y, Zhu J, He XJ, Wu JM, Iida K, Lu XY, Cui X, Jin H, Zhu JK. 2008. The Arabidopsis NFYA5 transcription factor is regulated transcriptionally and posttranscriptionally to promote drought resistance. Plant Cell **20**, 2238–2251.

Lian C, Li Q, Yao K, Zhang Y, Meng S, Yin W, Xia X. 2018. *Populus trichocarpa* PtNF-YA9, A Multifunctional Transcription Factor, Regulates Seed Germination, Abiotic Stress, Plant Growth and Development in Arabidopsis. Frontiers in Plant Science **9**, 1–15.

Lin ZF, Ehleringer J. 1983. Photosynthetic characteristics of *Amaranthus tricolor,* a C_4 tropical leafy vegetable. Photosynthesis Research **4**, 171–178.

Lin R, Wang H. 2004. Arabidopsis *FHY3*/*FAR1* gene family and distinct roles of its members in light control of Arabidopsis development. Plant Physiology **136**, 4010–4022.

Liu D, Palla KJ, Hu R, *et al.* 2018. Perspectives on the basic and applied aspects of crassulacean acid metabolism (CAM) research. Plant Science **274**, 394–401.

Liu Z-Q, Yan L, Wu Z, Mei C, Lu K, Yu Y-T, Liang S, Zhang X-F, Wang X-F, Da-Peng Z. 2012. Cooperation of three WRKY-domain transcription factors methylation and chromatin patterning WRKY18, WRKY40, and WRKY60 in repressing two ABA- responsive genes ABI4 and ABI5 in Arabidopsis. Journal of Experimental Botany **63**, 6371–6392.

Lu G, Paul AL, McCarty DR, Ferl RJ. 1996. Transcription factor veracity: Is GBF3 responsible for ABA-regulated expression of Arabidopsis *Adh*? Plant Cell **8**, 847–857.

Lüttge U. 2002. CO_2-concentrating: consequences in crassulacean acid metabolism. Journal of Experimental Botany **53**, 2131–2142.

Ma L, Bajic VB, Zhang Z. 2013. On the classification of long non-coding RNAs. RNA Biology **10**, 924–933.

Mach J. 2017. The long-noncoding RNA ELENA1 functions in plant immunity. Plant Cell **29**, 916.

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.

Mallappa C, Yadav V, Negi P, Chattopadhyay S. 2006. A basic leucine zipper transcription factor, G-box-binding factor 1, regulates blue light-mediated photomorphogenic growth in Arabidopsis. Journal of Biological Chemistry **281**, 22190–22199.

Marinho MAO, Souza G, Felix LP, De Carvalho R. 2019. Comparative cytogenetics of the ACPT clade (*Anacampserotaceae, Cactaceae, Portulacaceae,* and *Talinaceae*): a very diverse group of the suborder *Cactineae, Caryophyllales*. Protoplasma **256**, 805–814.

Martin JA, Wang Z. 2011. Next-generation transcriptome assembly. Nature Reviews Genetics **12**, 671–682.

Matasci N, Hung LH, Yan Z, *et al.* 2014. Data access for the 1,000 Plants (1KP) project. GigaScience **3**, 1–10.

Minardi BD, Voytena APL, Santos M, Randi ÁM. 2014. Water stress and abscisic acid treatments induce the CAM pathway in the epiphytic fern *Vittaria lineata* (L.) Smith. Photosynthetica **52**, 404–412.

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

Ming R, Wai CM, Guyot R. 2016. Pineapple Genome: A Reference for Monocots and CAM Photosynthesis. Trends in Genetics **32**, 690–696.

Miyao M. 2003. Molecular evolution and genetic engineering of C_4 photosynthetic enzymes. Journal of Experimental Botany **54**, 179–189.

Montero E, Francisco AM, Montes E, Herrera A. 2018. Salinity induction of recycling Crassulacean acid metabolism and salt tolerance in plants of *Talinum triangulare*. Annals of Botany, 1–10.

Moriyama Y, Koshiba-Takeuchi K. 2018. Significance of whole-genome duplications on the emergence of evolutionary novelties. Briefings in Functional Genomics **17**, 329–338.

Moseley RC, Motta F, Tuskan GA, Haase S, Yang X. 2019. Inference of Gene Regulatory Network Uncovers the Linkage Between Circadian Clock and Crassulacean Acid Metabolism in *Kalanchoë fedtschenkoi.* bioRxiv, 745893.

Nelson P, Kiriakidou M, Sharma A, Maniataki E, Mourelatos Z. 2003. The microRNA world: Small is mighty. Trends in Biochemical Sciences **28**, 534–540.

Nelson DE, Repetti PP, Adams TR, *et al.* 2007. Plant nuclear factor Y (NF-Y) B subunits confer drought tolerance and lead to improved corn yields on water-limited acres. Proceedings of the National Academy of Sciences of the United States of America **104**, 16450–16455.

Nimmo HG. 2000. The regulation of phosphoenolpyruvate carboxylase in CAM plants. Trends in Plant Science **5**, 75–80.

Novakova M, Motyka V, Dobrev PI, Malbeck J, Gaudinova A, Vankova R. 2005. Diurnal variation of cytokinin, auxin and abscisic acid levels in tobacco leaves. Journal of Experimental Botany **56**, 2877–2883.

Nozawa A, Matsubara Y, Tanaka Y, Takahashi H, Akagi T, Seki M, Shinozaki K, Endo Y, Sawasaki T. 2009. Construction of a protein library of Arabidopsis transcription factors using a wheat cell-free protein production system and its application for DNA binding analysis. Bioscience, Biotechnology and Biochemistry **73**, 1661–1664.

Nyananyo BL, Olowokudejo JD. 1986. Taxonomic studies in the genus *Talinum (Portulacaceae)* in Nigeria. Willdenowia **15**, 455–463.

O'Malley RC, Huang S shan C, Song L, Lewsey MG, Bartlett A, Nery JR, Galli M, Gallavotti A, Ecker JR. 2016. Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. Cell **165**, 1280–1292.

Odee DW, Wilson J, Omondi S, Perry A, Cavers S. 2015. Rangewide ploidy variation and evolution in *Acacia senegal*: A north-south divide? AoB PLANTS **7**.

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

Paulus JK, Niehus C, Groth G. 2013. Evolution of C_4 phosphoenolpyruvate carboxylase: Enhanced feedback inhibitor tolerance is determined by a single residue. Molecular Plant **6**, 1996–1999.

Pedersen O, Rich SM, Pulido C, Cawthray GR, Colmer TD. 2011. Crassulacean acid metabolism enhances underwater photosynthesis and diminishes photorespiration in the aquatic plant *Isoetes australis.* New Phytologist **190**, 332–339.

Pelé J, Bécu JM, Abdi H, Chabbert M. 2012. Bios2mds: An R package for comparing orthologous protein families by metric multidimensional scaling. BMC Bioinformatics **13**, 1.

Pimentel H, Bray NL, Puente S, Melsted P, Pachter L. 2017. Differential analysis of RNA-seq incorporating quantification uncertainty. Nature Methods **14**, 687–690.

Raghavendra AS, Gerst U, Heber U. 1995. Oscillations in photosynthetic carbon assimilation and chlorophyll fluorescence are different in *Amaranthus caudatus*, a C_4 plant, and *Spinacia oleracea*, a C_3 plant. Planta **195**, 471–477.

Rai MI, Alam M, Lightfoot DA, Gurha P, Afzal AJ. 2019. Classification and experimental identification of plant long non-coding RNAs. Genomics **111**, 997–1005.

Ramegowda V, Gill US, Sivalingam PN, *et al.* 2017. GBF3 transcription factor imparts drought tolerance in *Arabidopsis thaliana*. Scientific Reports **7**, 1–13.

Rawat R, Takahashi N, Hsu PY, Jones MA, Schwartz J, Salemi MR, Phinney BS, Harmer SL. 2011. REVEILLE8 and PSEUDO-REPONSE REGULATOR5 Form a Negative Feedback Loop within the

Arabidopsis Circadian Clock. PLoS Genetics **7**.

Rayder L, Ting IP. 1981. Carbon Metabolism in Two Species of *Pereskia (Cactaceae).* Plant Physiology **68**, 139–142.

De Rocher EJ, Harkins KR, Galbraith DW, Bohnert HJ. 1990. Developmentally regulated systemic endopolyploidy in succulents with small genomes. Science **250**, 99–101.

Ruan J. 2016. Ultra-fast de novo assembler using long noisy reads.

Sage RF, Christin PA, Edwards EJ. 2011. The C_4 plant lineages of planet Earth. Journal of Experimental Botany **62**, 3155–3169.

Sage RF, Sage TL, Pearcy RW, Borsch T. 2007. The taxonomic distribution of C_4 photosynthesis in Amaranthaceae sensu stricto. American Journal of Botany **94**, 1992–2003.

Sakr S, Wang M, Dédaldéchamp F, Perez-Garcia M-D, Ogé L, Hamama L, Atanassova R. 2018. The Sugar-Signaling Hub: Overview of Regulators and Interaction with the Hormonal and Metabolic Network. International Journal of Molecular Sciences **19**, 2506.

Satake A, Sakurai G, Kinoshita T. 2015. Modeling strategies for plant survival, growth and reproduction. Plant and Cell Physiology **56**, 583–585.

Schmutz J, Cannon SB, Schlueter J, *et al.* 2010. Genome sequence of the palaeopolyploid soybean. Nature **463**, 178–183.

Schnable PS, Pasternak S, Liang C, *et al.* 2009. The B73 Maize Genome: Complexity, Diversity, and Dynamics. Science **326**, 1112–1115.

Schuler ML, Mantegazza O, Weber APM. 2016. Engineering C_4 photosynthesis into C_3 chassis in the synthetic biology age. The Plant Journal **87**, 51–65.

Shalit-Kaneh A, Kumimoto RW, Filkov V, Harmer SL. 2018. Multiple feedback loops of the Arabidopsis circadian clock provide rhythmic robustness across environmental conditions. Proceedings of the National Academy of Sciences of the United States of America **115**, 7147–7152.

Shang Y, Yan L, Liu ZQ, *et al.* 2010. The Mg-chelatase H subunit of Arabidopsis antagonizes a group of WRKY transcription repressors to relieve ABA-responsive genes of inhibition. Plant Cell **22**, 1909–1935.

Shin R, Burch AY, Huppert KA, Tiwari SB, Murphy AS, Guilfoyle TJ, Schachtman DP. 2007. The Arabidopsis transcription factor MYB77 modulates auxin signal transduction. Plant Cell **19**, 2440–2453.

Siefers N, Dang KK, Kumimoto RW, Bynum IV WE, Tayrose G, Holt BF. 2009. Tissue-specific expression patterns of Arabidopsis NF-Y transcription factors suggest potential for extensive combinatorial complexity. Plant Physiology **149**, 625–641.

Silvera K, Neubig KM, Whitten WM, Williams NH, Winter K, Cushman JC. 2010. Evolution along the crassulacean acid metabolism continuum. Functional Plant Biology **37**, 995–1010.

Silvera K, Winter K, Rodriguez BL, Albion RL, Cushman JC. 2014. Multiple isoforms of phosphoenolpyruvate carboxylase in the *Orchidaceae* (subtribe *Oncidiinae*): implications for the evolution of crassulacean acid metabolism. Journal of Experimental Botany **65**, 3623–3636.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E V., Zdobnov EM. 2015. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics **31**, 3210–3212.

Simon SA, Zhai J, Nandety RS, McCormick KP, Zeng J, Mejia D, Meyers BC. 2009. Short-Read Sequencing Technologies for Transcriptional Analyses. Annual Review of Plant Biology **60**, 305–333.

Söderman E, Hjellström M, Fahleson J, Engström P. 1999. The HD-Zip gene *ATHB6* in *Arabidopsis* is expressed in developing leaves, roots and carpels and up-regulated by water deficit conditions. Plant Molecular Biology **40**, 1073–1083.

Soltis DE, Soltis PS, Bennett MD, Leitch IJ. 2003. Evolution of genome size in the angiosperms. American Journal of Botany **90**, 1596–1603.

Song L, Huang SC, Wise A, Castanon R, Nery JR, Chen H, Watanabe M, Thomas J, Bar-Joseph Z, Ecker JR. 2016. A transcription factor hierarchy defines an environmental stress response network. Science **354**, aag15510-1.

Stata M, Sage TL, Rennie TD, Khoshravesh R, Sultmanis S, Khaikin Y, Ludwig M, Sage RF. 2014. Mesophyll cells of C_4 plants have fewer chloroplasts than those of closely related C_3 plants. Plant, Cell and Environment **37**, 2587–2600.

Steiner E, Efroni I, Gopalraj M, Saathoff K, Tseng TS, Kieffer M, Eshed Y, Olszewski N, Weiss D. 2012. The Arabidopsis O-linked N-acetylglucosamine transferase SPINDLY interacts with class I TCPs to facilitate cytokinin responses in leaves and flowers. Plant Cell **24**, 96–108.

Straube J, Gorse A-D, PROOF Centre of Excellence Team, Huang BE, Lê Cao K-A. 2015. A Linear Mixed Model Spline Framework for Analysing Time Course 'Omics' Data. PLos ONE **10**, 1–20.

Sugimoto-Shirasu K, Roberts K. 2003. 'Big it up': Endoreduplication and cell-size control in plants. Current Opinion in Plant Biology **6**, 544–553.

Sun L, Luo H, Bu D, Zhao G, Yu K, Zhang C, Liu Y, Chen R, Zhao Y. 2013. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. Nucleic Acids Research **41**, e166.

Svensson P, Bläsing OE, Westhoff P. 2003. Evolution of C_4 phosphoenolpyruvate carboxylase. Archives of Biochemistry and Biophysics **414**, 180–188.

Swindell WR, Huebner M, Weber AP. 2007. Transcriptional profiling of Arabidopsis heat shock proteins and transcription factors reveals extensive overlap between heat and non-heat stress response pathways. BMC Genomics **8**, 1–15.

Tallman G, Zhu J, Mawson BT, Amodeo G, Nouhi Z, Levy K, 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 and Cell Physiology **38**, 236–242.

Tang H, Krishnakumar V, Bidwell S, *et al.* 2014. An improved genome release (version Mt4.0) for the model legume *Medicago truncatula.* BMC Ge **15**, 312.

Taybi T, Cushman JC, Borland AM. 2017. Leaf carbohydrates influence transcriptional and post-transcriptional regulation of nocturnal carboxylation and starch degradation in the facultative CAM plant, *Mesembryanthemum crystallinum.* Journal of Plant Physiology **218**, 144–154.

Team RC. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

The French-Italian Public Consortium for Grapevine Genome Characterization. 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature **449**, 463–468.

The International Brachypodium Initiative. 2010. Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature **463**, 763–768.

The International Wheat Genome Sequencing Consortium. 2014. A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. Science **345**, 1251788.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research **22**, 4673–4680.

Tian Y, Wen H, Qi X, *et al.* 2019. Characterization of Full-Length Transcriptome Sequences and Splice Variants of Lateolabrax maculatus by Single-Molecule Long-Read Sequencing and Their Involvement in Salinity Regulation. Frontiers in Genetics **10**, 1–19.

TransDecoder. URL https://github.com/TransDecoder/TransDecoder

Tsukagoshi H, Suzuki T, Nishikawa K, Agarie S, Ishiguro S, Higashiyama T. 2015. RNA-Seq analysis of the response of the halophyte, *Mesembryanthemum crystallinum* (ice plant) to high salinity. PLoS ONE **10**, 1–16.

The UniProt Consortium. 2019. UniProt: A worldwide hub of protein knowledge. Nucleic Acids Research **47**, D506–D515.

Uno Y, Furihata T, Abe H, Yoshida R, Shinozaki K, Yamaguchi-Shinozaki K. 2000. Arabidopsis basic leucine zipper transcription factors involved in an abscisic acid-dependent signal transduction pathway under drought and high-salinity conditions. Proceedings of the National Academy of Sciences of the United States of America **97**, 11632–11637.

Venturini L, Caim S, Kaithakottil GG, Mapleson DL, Swarbreck D. 2018. Leveraging multiple transcriptome assembly methods for improved gene structure annotation. GigaScience **7**, 1–15.

Vovides AP, Etherington JR, Dresser PQ, Groenhof A, Ramirez CIJF. 2002. CAM-cycling in the cycad *Dioon edule* Lindl. in its natural tropical deciduous forest habitat in central Veracruz, Mexico. Botanical Journal of the Linnean Society **138**, 155–162.

Voznesenskaya E V, Koteyeva NK, Edwards GE, Ocampo G. 2017. Unique photosynthetic phenotypes

in *Portulaca (Portulacaceae):* C_3-C_4 intermediates and NAD-ME C_4 species with Pilosoid-type Kranz anatomy. Journal of Experimental Botany **68**, 225–239.

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

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*.

Walker BJ, Abeel T, Shea T, *et al.* 2014. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS ONE **9**.

Wang P, Fouracre J, Kelly S, *et al.* 2013. Evolution of GOLDEN2-LIKE gene function in C_3 and C_4 plants. Planta **237**, 481–495.

Wang A, Hu J, Gao C, Chen G, Wang B, Lin C, Song L, Ding Y, Zhou G. 2019*a*. Genome-wide analysis of long non-coding RNAs unveils the regulatory roles in the heat tolerance of Chinese cabbage (*Brassica rapa* ssp. chinensis). Scientific Reports **9**, 5002.

Wang L, Ma G, Wang H, *et al.* 2019*b*. A draft genome assembly of halophyte *Suaeda aralocaspica*, a plant that performs C4 photosynthesis within individual cells. GigaScience **8**, 1–9.

Wang D, Qu Z, Yang L, *et al.* 2017. Transposable elements (TEs) contribute to stress-related long intergenic noncoding RNAs in plants. Plant Journal **90**, 133–146.

Wang G, Yin H, Li B, *et al.* 2019*c*. Characterization and identification of long non-coding RNAs based on feature relationship. Bioinformatics **35**, 2949–2956.

Wedding RT, Black MK, Meyer CR. 1990. Inhibition of Phosphoenolpyruvate Carboxylase by Malate. Plant Physiology **92**, 456–461.

Weise SE, Van Wijk KJ, Sharkey TD. 2011. The role of transitory starch in C_3, CAM, and C_4 metabolism and opportunities for engineering leaf starch accumulation. Journal of Experimental Botany **62**, 3109–3118.

West-Eberhard MJ, Smith JAC, Winter K. 2011. Photosynthesis, Reorganized. Science **332**, 311–312.

Westhoff P, Gowik U. 2004. Evolution of C_4 phosphoenolpyruvate carboxylase. Genes and proteins: A case study with the genus *Flaveria*. Annals of Botany **93**, 13–23.

Wilkins MB. 1992. Tansley Review No. 37 Circadian rhythms: their origin and control. New Phytologist **121**, 347–375.

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

Winter K, Holtum JAM. 2014. Facultative crassulacean acid metabolism (CAM) plants: Powerful tools for unravelling the functional elements of CAM photosynthesis. Journal of Experimental Botany **65**, 3425–3441.

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

Wolfe-Bellin KS, He JH, Bazzaz FA. 2006. Leaf-Level Physiology, Biomass, and Reproduction of *Phytolacca americana* under Conditions of Elevated Carbon Dioxide and Increased Nocturnal Temperature. International Journal of Plant Sciences **167**, 1011–1020.

Wu JF, Tsai HL, Joanito I, *et al.* 2016*a*. LWD-TCP complex activates the morning gene CCA1 in Arabidopsis. Nature Communications **7**, 13181.

Wu Q, Zhang X, Peirats-Llobet M, Belda-Palazon B, Wang X, Cui S, Yu X, Rodriguez PL, An C. 2016*b*. Ubiquitin Ligases RGLG1 and RGLG5 Regulate Abscisic Acid Signaling by Controlling the Turnover of Phosphatase PP2CA. The Plant Cell **28**, 2178–2196.

Xu D, Li J, Gangappa SN, Hettiarachchi C, Lin F, Andersson MX, Jiang Y, Deng XW, Holm M. 2014*a*. Convergence of Light and ABA Signaling on the *ABI5* Promoter. PLoS Genetics **10**, e1004197.

Xu YX, Liu Y, Chen ST, Li XQ, Xu LG, Qi YH, Jiang DA, Jin SH. 2014*b*. The B subfamily of plant ATP binding cassette transporters and their roles in auxin transport. Biologia Plantarum **58**, 401–410.

Yang X, Cushman JC, Borland AM, *et al.* 2015*a*. 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, *et al.* 2017. The *Kalanchoë* genome provides insights into convergent evolution and

building blocks of crassulacean acid metabolism. Nature Communications **8**, 1899.

Yang X, Liu D, Tschaplinski TJ, Tuskan GA. 2019. Comparative genomics can provide new insights into the evolutionary mechanisms and gene function in CAM plants. Journal of Experimental Botany **70**, 6539–6547.

Yang Y, Moore MJ, Brockington SF, *et al.* 2015*b*. Dissecting molecular evolution in the highly diverse plant clade *Caryophyllales* using transcriptome sequencing. Molecular Biology and Evolution **32**, 2001–2014.

Yang Y, Moore MJ, Brockington SF, Mikenas J, Olivieri J, Walker JF, Smith SA. 2018. Improved transcriptome sampling pinpoints 26 ancient and more recent polyploidy events in *Caryophyllales*, including two allopolyploidy events. New Phytologist **217**, 855–870.

Yates AD, Achuthan P, Akanni W, *et al.* 2020. Ensembl 2020. Nucleic Acids Research **48**, D682–D688.

Zakhartsev M, Medvedeva I, Orlov Y, Akberdin I, Krebs O, Schulze WX. 2016. Metabolic model of central carbon and energy metabolisms of growing *Arabidopsis thaliana* in relation to sucrose translocation. BMC Plant Biology **16**.

Zanella M, Borghi GL, Pirone C, Thalmann M, Pazmino D, Costa A, Santelia D, Trost P, Sparla F. 2016. β-amylase 1 (BAM1) degrades transitory starch to sustain proline biosynthesis during drought stress. Journal of Experimental Botany **67**, 1819–1826.

Zhang R, Calixto CPG, Marquez Y, *et al.* 2017. A high quality Arabidopsis transcriptome for accurate transcript-level analysis of alternative splicing. Nucleic Acids Research **45**, 5061–5073.

Zhang J, Li Y, Jia HX, Li JB, Huang J, Lu MZ, Hu JJ. 2015. The heat shock factor gene family in *Salix suchowensis*: A genome-wide survey and expression profiling during development and abiotic stresses. Frontiers in Plant Science **6**, 748.

Zhang GQ, Xu Q, Bian C, *et al.* 2016. The *Dendrobium catenatum* Lindl. genome sequence provides insights into polysaccharide synthase, floral development and adaptive evolution. Scientific Reports **6**, 1–10.

Zhao H, Wu D, Kong F, Lin K, Zhang H, Li G. 2017. The *Arabidopsis* thaliana Nuclear Factor Y Transcription Factors. Frontiers in Plant Science **07**, 1–11.

Zhao Y, Xing L, Wang X, Hou YJ, Gao J, Wang P, Duan CG, Zhu X, Zhu JK. 2014. The ABA receptor PYL8 promotes lateral root growth by enhancing MYB77-dependent transcription of auxin-responsive genes. Science Signaling **7**, 1–12.

Ziegler H, Batanouny KH, Sankhla N, Vyas OP, Stichler W. 1981. The Photosynthetic Pathway Types of some Desert Plants from India, Saudi Arabia, Egypt, and Iraq. Oecologia **48**, 93–99.

Zonneveld BJ. 2002. Genome size analysis of selected species of *Aloe (Aloaceae)* reveals the most primitive species and results in some new combinations. Bradleya **20**, 5–12.

Supplementary figures

Figure S1. Flow cytometry for genome size estimation.

Figure S2. Completeness assessment of the transcriptome assembly upon removal of all open reading frames tagged as “suspicious”.

Figure S3. OrthoFinder analysis: Comparative genomics statistics and orthogroup occupancy.

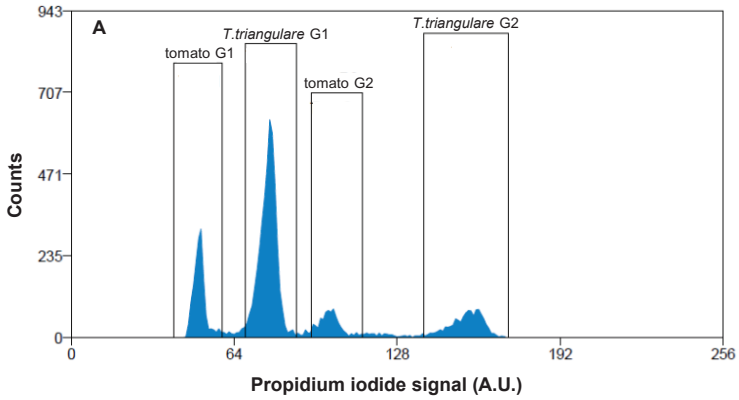
Figures S4. A schematic representation of multiple sequence alignment of PEPC protein sequences in the proximity of His519 and Ser780, numbering is based on CAA33317 of *Zea mays*.

Figure S5. Abundance of selected transcripts.

Figure S6. Number of orthologs of CAM genes in *Ananas comosus, Beta vulgaris, Chenopodium quinoa, Dendrobium catenatum, Hordeum vulgare, Kalanchoë laxiflora, K. fedtschenkoi, Phalaenopsis equestris, Spinacia oleracea, Talinum triangulare*, *Zea mays*.

Figure S7. Multiple sequence alignment of PEPC.

Figure S8. Metric multidimensional scaling (MDS) for several CAM genes.



sample 2C = reference 2C * (sample 2C mean peak position / reference 2C mean peak position)
sample 2C = 1.96 * (145 / 71)
sample 2C = 3.97 pg ~ 3.89 Gbp

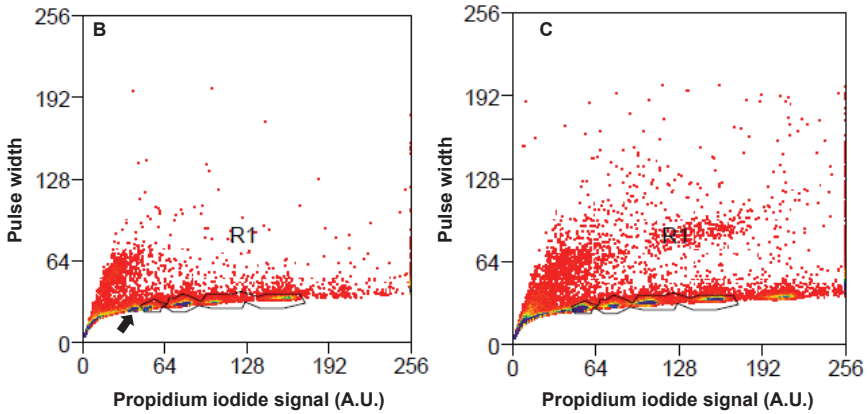


Figure S1. Flow cytometry for genome size estimation. (A) A representative histogram with G1 and G2 peaks of *Solanum lycopersicum* L. 'Stupické polní rané' used as internal standard and *Talinum triangulare*. Genome size calculation according to (Doležel and Bartoš, 2005). (B) Nuclei population of leaf tissue, one-month old *Talinum triangulare*. The arrow marks a putative sub-population of diploid nuclei (C) Nuclei population of leaf tissue, *Talinum triangular* more than two months old.

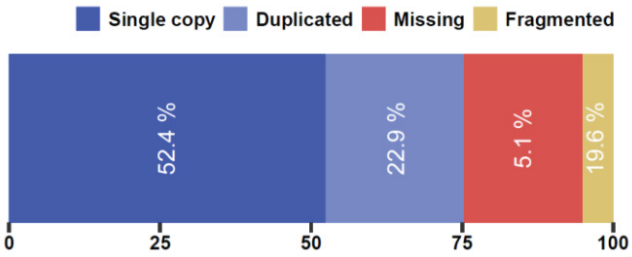


Figure S2. Completeness assessment of the transcriptome assembly upon removal of all open reading frames tagged as "suspicious". BUSCO analysis with *Embryophyta* single-copy orthologs. The proportion of missing BUSCOs increased to 19.6% (as compared to 3.8% for all predicted open reading frames; Fig. 3A).

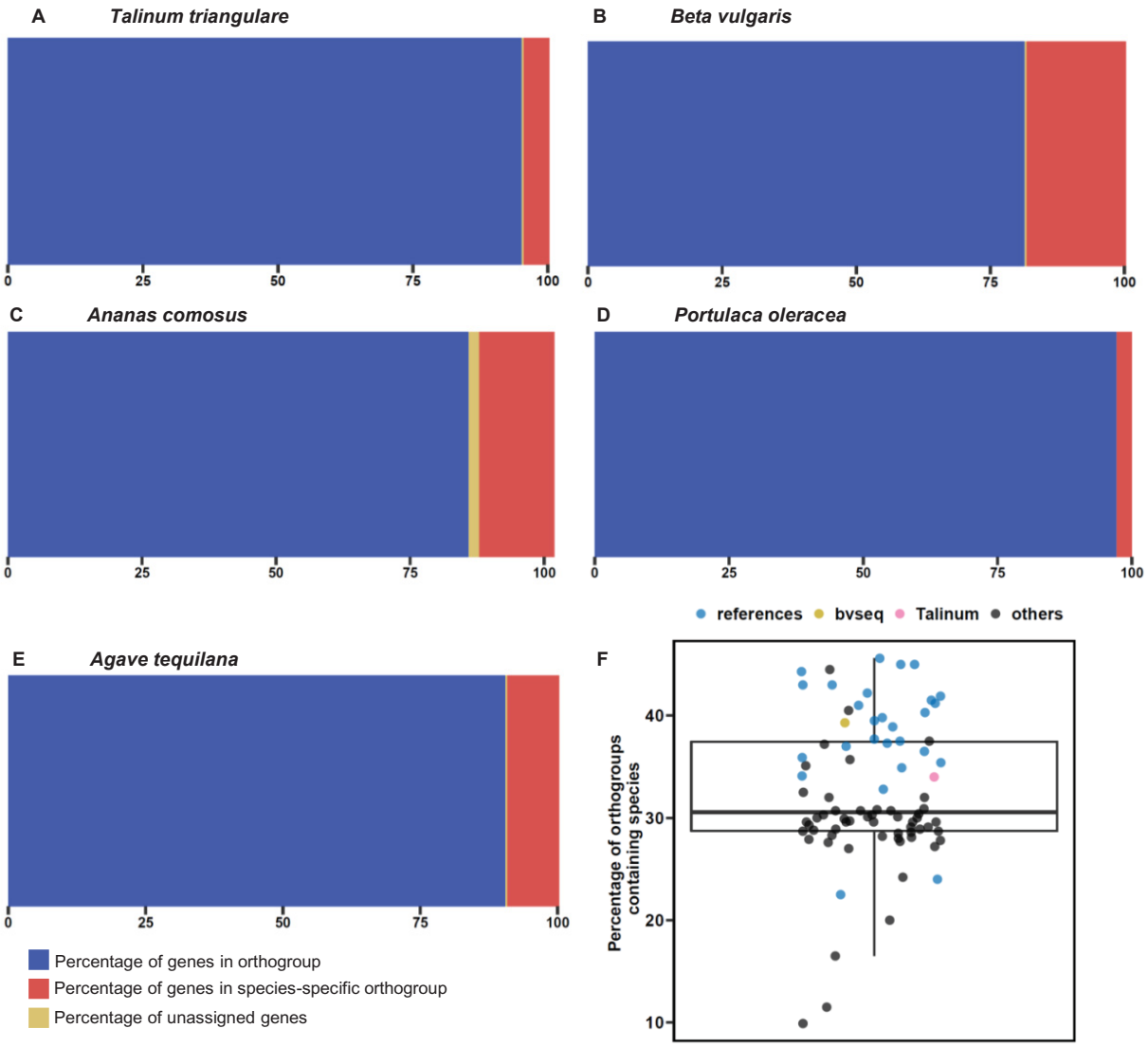


Figure S3. OrthoFinder analysis: Comparative genomics statistics and orthogroup occupancy. (A) *Talinum triangulare* transcriptome assembly generated in this work. (B) *Beta vulgaris*, transcriptome based on reference genome. (C) *Ananas comosus*, transcriptome based on reference genome. (D) *Portulaca oleracea*, a shotgun transcriptome assembly. (E) *Agave tequilana*, a shotgun transcriptome assembly. (F) Percentage of orthogroups occupied by proteins of a given species. "Reference" species refer to species with a reference genome available, *Beta vulgaris* (bvseq) is highlighted because it is the closest sequenced relative of *Talinum triangulare*, "others" are shotgut assemblies.

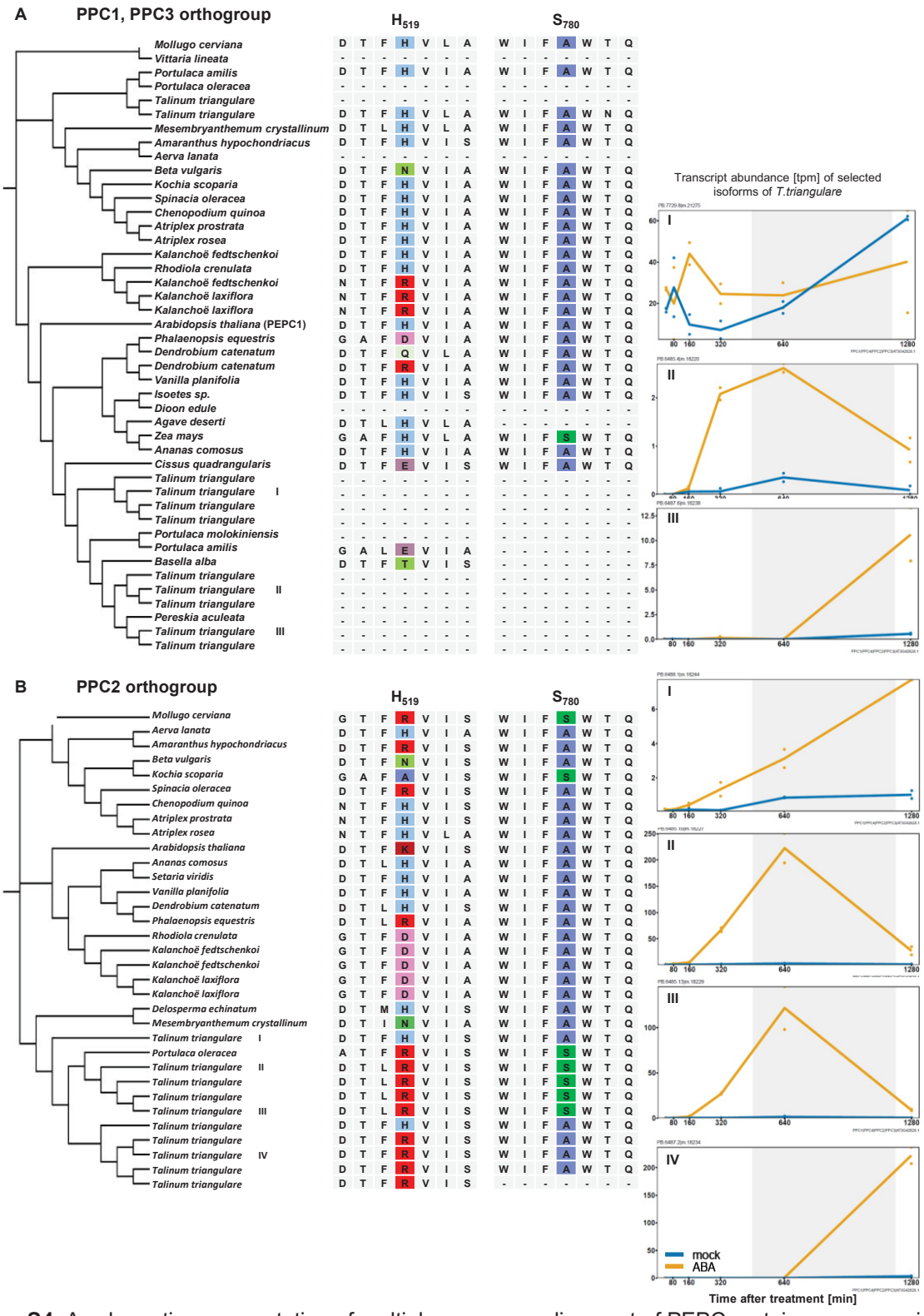


Figure S4. A schematic representation of multiple sequence alignment of PEPC protein sequences in the proximity of His519 and Ser780, numbering is based on CAA33317 of *Zea mays*. The site of interest is colour-coded. For selected isoforms of *Talinum triangulare*, transcript abundance patterns are shown and labelled with roman numerals to match patterns with respective isoform in the phylogenetic tree.

Figure S4. (*Continued.*)
Phylogenetic trees are approximately-maximum-likelihood trees obtained with FastTree v2.1 (default settings) (A) Orthologs of Arabidopsis PEPC1 and PEPC3, contains *Zea mays* PEP1 (B) Orthologs of Arabidopsis PEPC2, contains *Kalanchoë fedtschenkoi* PEPC2.

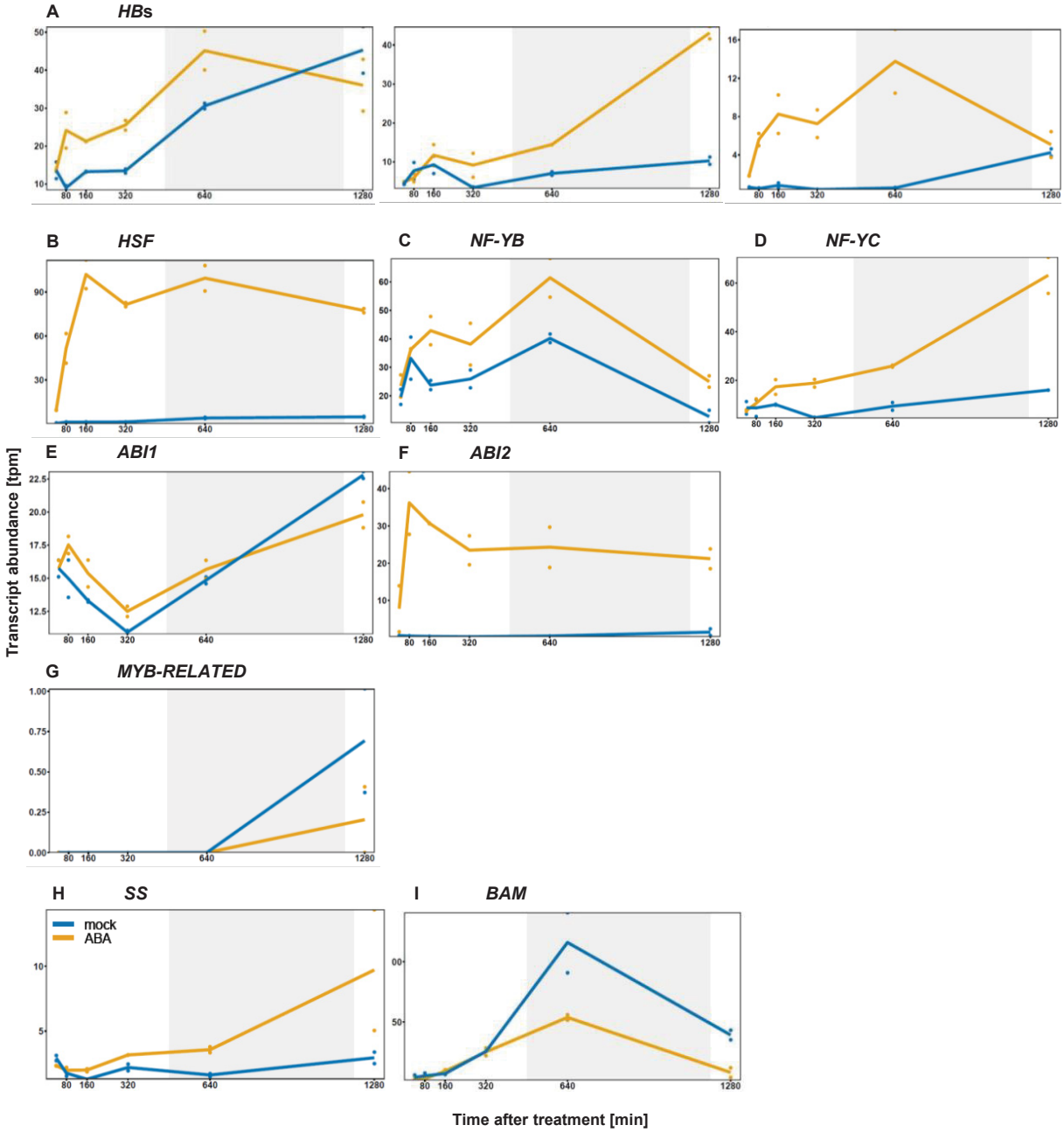


Figure S5. Abundance of selected transcripts. (A) *Homeobox proteins (HB).* (B) *Heat shock factor (HSF).* (C) *Nuclear factor Y, subunit B (NF-YB).* (D) *Nuclear factor Y, subunit C (NF-YC).* (E) *ABA insensitive 1 (ABI1),* a 2C class of protein serine/threonine phosphatase (PP2C). (F) *ABA insensitive 2 (ABI2),* a PP2C. (G) A MYB-related transcription factor, ortholog of components of the circadian oscillator. (H) *Starch synthase (SS).* (I) *β-amylase (BAM).* Dark phase is marked with a grey background.

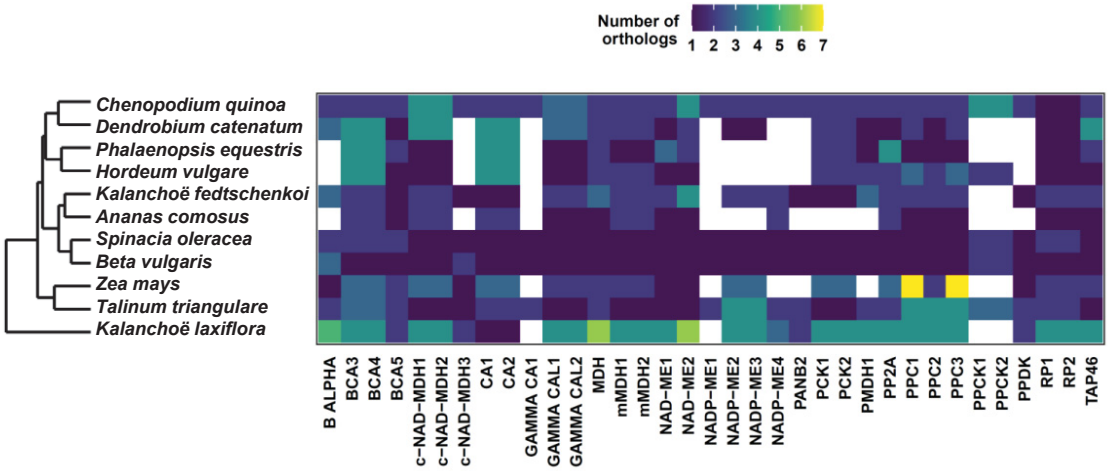


Figure S6. Number of orthologs of CAM genes in *Ananas comosus, Beta vulgaris, Chenopodium quinoa, Dendrobium catenatum, Hordeum vulgare, Kalanchoë laxiflora, K. fedtschenkoi, Phalaenopsis equestris, Spinacia oleracea, Talinum triangulare, Zea mays.* Orthologs were identified based on their homology to Arabidopsis genes.

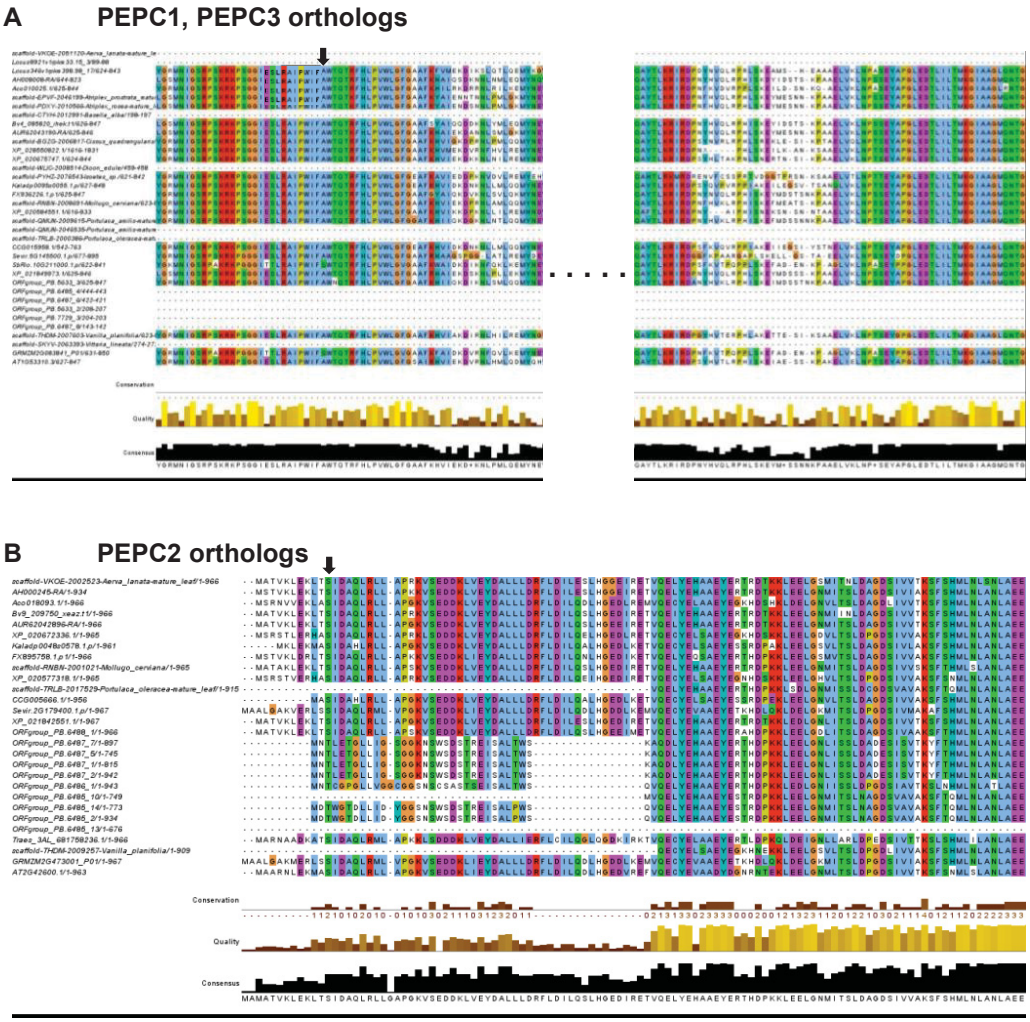


Figure S7. Multiple sequence alignment of PEPC. (A) PEPC1 and PEPC3 orthologs, the arrow points at the position homologous to Ser780 in PEP1 of *Zea mays* (CAA33317). C-terminus of PEPC is truncated in several species, including *Aerva lanate, Agave deserti, Portulaca oleracea* and *Talinum triangulare*. (B) PEPC2 orthologs, the arrow points at the position homologous to Ser15 in PEP1 of *Zea mays* (CAA33317). Several isoforms of *Talinum triangulare* show highly diversified N-terminus compared to phylogenetically diverse C_3, C_4 and CAM species.

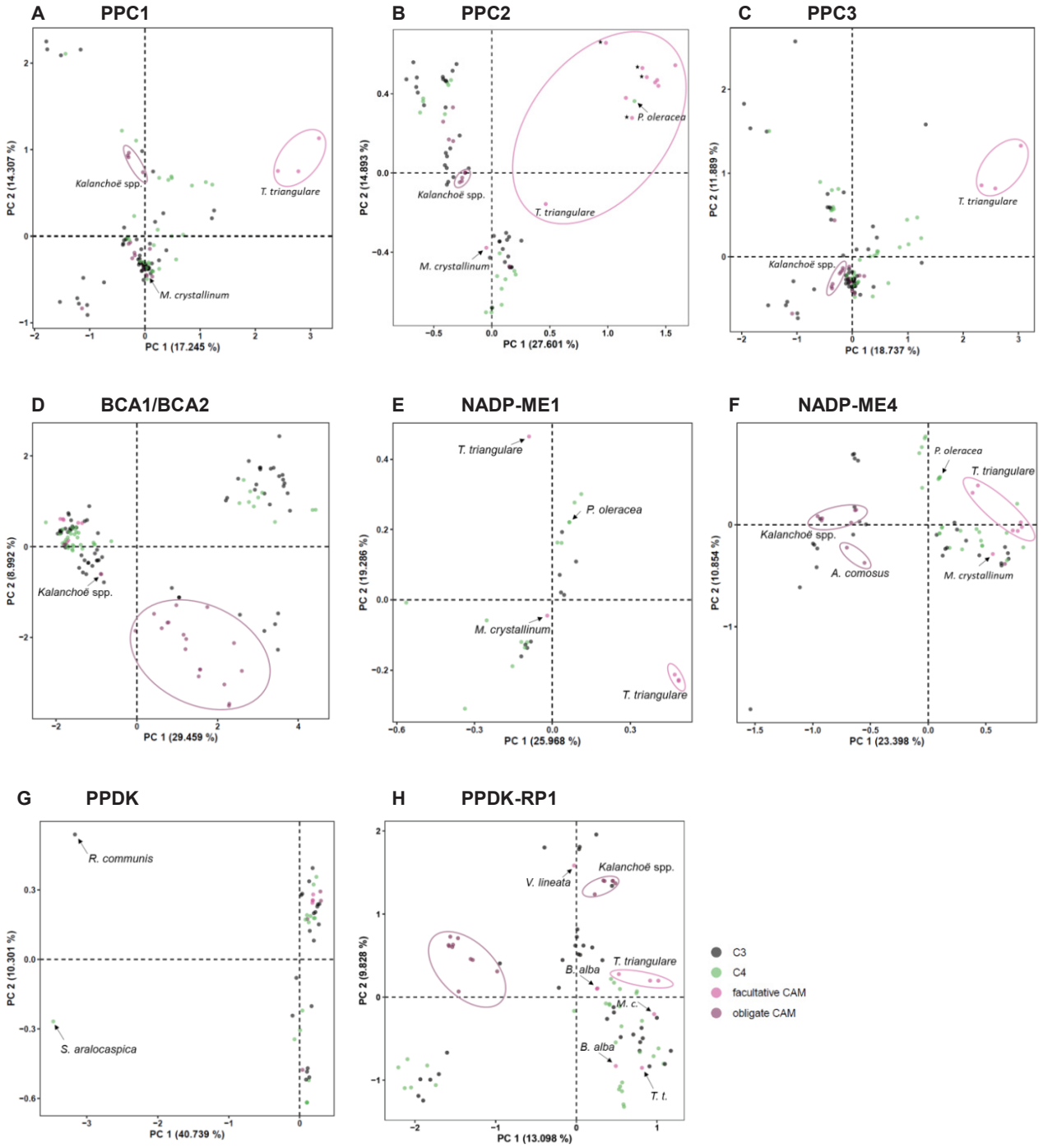


Figure S8. Metric multidimensional scaling (MDS) for several CAM genes. Multiple sequence alignments served as basis to construct distance matrices, which can then be visualized in a low dimensional space by MDS as available within the bios2mds package for R. Orthologs were identified based on their homology to Arabidopsis sequences (A) PEPC1 orthologs. (B) PEPC2 orthologs. (C) PEPC3 orthologs. (D) β-carbonic anhydrase 1 and β-carbonic anhydrase 2 orthologs (could not be differentiated unambiguously). (E) NADP-malic enzyme 1. (F) NADP-malic enzyme 4. (G) Pyruvate, phosphate dikinase (PPDK). (H) PPDK regulatory protein 1.

Supplementary tables

Table S1. Species included in OrthoFinder analysis, photosynthesis information and data source.

Table S2. Statistics of raw Nanopore reads used for de novo genome assembly of Talinum triangulare

Table S3. Statistics of the genome assembly (Flye & Pilon) obtained with QUAST.

Table S4. RNA-seq mappings statistics on the genome assembly obtained with HISAT2. All details about RNA-sequencing and samples available in Maleckova et al. (2019)

Table S5. Statistic for Iso-Seq (PacBio) transcriptome sequencing and identification of high-quality transcripts with the isoseq3 pipeline.

Table S6. BLAST results for high quality transcripts (obtained with isoseq3 pipeline) with non-plant hits.

Table S7. RNA-seq mappings statistics on the transcriptome assembly

Table S8. Enriched motifs in promoter regions of differentially expressed genes ($q < 0.01$) as identified with the lmms package.

Table S9. Enriched motifs in promoter regions of selected orthogroups.

Table S10. Orthologs of *Talinum triangulare* and at least one other facultative CAM species (OrthoFinder analysis) and their assignment to clusters with altered temporal patterns upon ABA treatment.

Table S11. Statistics for genome assembly obtained with SMARTdenovo

Table S1. Species included in OrthoFinder analysis, photosynthesis information and data source.

Species	PS type	Family	Order	Data source		Photosynthesis information
Aerva lanata	C_3	*Amaranthaceae*	*Caryophyllales*	1KP Transcriptomes	Matasci *et al.* (2014)	Ziegler *et al.* (1981)
Aerva persica	C_4	*Amaranthaceae*	*Caryophyllales*	1KP Transcriptomes	Matasci *et al.* (2014)	Stata *et al.* (2014)
Agave deserti	obligate CAM	*Agavaceae*	*Asparagales*	www.datadrayd.org	Gross *et al.* (2013)	Winter *et al.* (2015)
Agave tequilana	obligate CAM	*Agavaceae*	*Asparagales*	www.datadrayd.org	Gross *et al.* (2013)	Winter *et al.* (2015)
Aloe vera	obligate CAM	*Asphodelaceae*	*Asparagales*	1KP Transcriptomes	Matasci *et al.* (2014)	Davis *et al.* (2019)
Alternanthera brasiliana	C_3	*Amaranthaceae*	*Caryophyllales*	1KP Transcriptomes	Matasci *et al.* (2014)	Sage *et al.* (2007)
Alternanthera caracasana	C4	*Amaranthaceae*	*Caryophyllales*	1KP Transcriptomes	Matasci *et al.* (2014)	Sage *et al.* (2007)
Alternanthera sessilis	C_3	*Amaranthaceae*	*Caryophyllales*	1KP Transcriptomes	Matasci *et al.* (2014)	Sage *et al.* (2007)
Alternanthera tenella	C_3-C_4	*Amaranthaceae*	*Caryophyllales*	1KP Transcriptomes	Matasci *et al.* (2014)	Sage *et al.* (2007)
Amaranthus hypochondriacus	C_4	*Amaranthaceae*	*Caryophyllales*	Phytozome 12	Clouse *et al.* (2016)	Lin and Ehleringer (1983)
Ananas comosus	obligate CAM	*Bromeliaceae*	*Poales*	Phytozome 12	Ming *et al.* (2015)	Ming *et al.* (2015)
Arabidopsis thaliana	C_3	*Brassicaceae*	*Brassicales*	Phytozome 12	Cheng *et al.* (2017)	Chang *et al.* (2016)
Asplenium nidus	C_3	*Aspleniaceae*	*Polypodiales*	1KP Transcriptomes	Matasci *et al.* (2014)	Choy-Sin and Suan (1974)
Atriplex prostrata	C_3	*Amaranthaceae*	*Caryophyllales*	1KP Transcriptomes	Matasci *et al.* (2014)	Stata *et al.* (2014)
Atriplex rosea	C_4	*Chenopodiaceae*	*Caryophyllales*	1KP Transcriptomes	Matasci *et al.* (2014)	Stata *et al.* (2014)
Basella alba	facultative CAM	*Basellaceae*	*Caryophyllales*	1KP Transcriptomes	Matasci *et al.* (2014)	Winter 2019
Beta maritima	C_3	*Amaranthaceae*	*Caryophyllales*	1KP Transcriptomes	Matasci *et al.* (2014)	
Beta vulgaris	C_3	*Amaranthaceae*	*Caryophyllales*	www.http://bvseq.boku.ac.at/	Dohm *et al.* (2014)	Schuler *et al.*, (2016)
Blutaparon vermiculare	C_4	*Amaranthaceae*	*Caryophyllales*	1KP Transcriptomes	Matasci *et al.* (2014)	Sage *et al.* (2007)
Boerhavia burbidgeana	C_4	*Nyctaginaceae*	*Caryophyllales*	1KP Transcriptomes	Matasci *et al.* (2014)	Stata *et al.* (2014)
Boerhavia coccinea	C_4	*Nyctaginaceae*	*Caryophyllales*	1KP Transcriptomes	Matasci *et al.* (2014)	Kocacinar and Sage (2015)
Bougainvillea spectabilis	C_4	*Nyctaginaceae*	*Caryophyllales*	1KP Transcriptomes	Matasci *et al.* (2014)	
Brachypodium distachyon	C_3	*Amaranthaceae*	*Caryophyllales*	Phytozome 12	The International Brachypodium Initiative (2010)	Brkljacic *et al.* (2011)
Brocchinia reducta	C_3	*Bromeliaceae*	*Poales*	1KP Transcriptomes	Matasci *et al.* (2014)	Benzing *et al.* (1985)

Table S1. (*Continued.*)

Species	PS type	Family	Order	Data source		Photosynthesis information
Chenopodium quinoa	C_3	*Amaranthaceae*	*Caryophyllales*	Phytozome 12	Jarvis *et al.* (2017)	Geissler 2015
Chlamydomonas reinhardtii	CCM	*Chlamydomonadaceae*	*Chlamydomonadales*	Phytozome 12	Jarvis *et al.* (2017)	Jungnick *et al.* (2014)
Cissus quadrangularis	CAM	*Vitaceae*	*Vitales*	1KP Transcriptomes	Matasci *et al.* (2014)	Ziegler *et al.* (1981)
Delosperma echinatum	CAM	*Aizoaceae*	*Caryophyllales*	1KP Transcriptomes	Matasci *et al.* (2014)	
Dendrobium catenatum	facultative CAM	*Orchidaceae*	*Asparagales*	NCBI	Zhang *et al.*, (2016)	Zhang *et al.* (2016)
Dianthus caryophyllus	C_3	*Caryophyllaceae*	*Caryophyllales*	1KP Transcriptomes	Matasci *et al.*, (2014)	Martin 1984
Dioon edule	CAM	*Zamiaceae*	*Cycadales*	1KP Transcriptomes	Matasci *et al.* (2014)	Vovides *et al.* (2002)
Glycine max	C_3	*Fabaceae*	*Fabales*	Phytozome 12	Schmutz *et al.* (2010)	Miyao (2003)
Gossypium hirsutum	C_3	*Malvaceae*	*Malvales*	Phytozome 12	JGI	Brugnoli and Lauteri (1991)
Hordeum vulgare	C_3	*Poaceae*	*Poales*	Phytozome 12	Beier *et al.* (2017)	Brkljacic *et al.* (2011)
Isoetes sp.	obligate CAM	*Isoetaceae*	*Isoetales*	1KP Transcriptomes	Matasci *et al.* (2014)	Pedersen *et al.* (2011)
Kalanchoe fedtschenkoi	obligate CAM	*Crassulaceae*	*Saxifragales*	Phytozome 12	Yang *et al.* (2017)	Yang *et al.* (2017)
Kalanchoe laxiflora	obligate CAM	*Crassulaceae*	*Saxifragales*	Phytozome 12	JGI	Boxall *et al.* (2020)
Kochia scoparia	C4	*Amaranthaceae*	*Caryophyllales*	1KP Transcriptomes	Matasci *et al.* (2014)	Kocacinar and Sage (2003)
Lophophora williamsii	obligate CAM	*Cactaceae*	*Caryophyllales*	1KP Transcriptomes	Matasci *et al.* (2014)	Ibarra-laclette *et al.* (2015)
Marchantia polymorpha	C_3	*Marchantiaceae*	*Marchantiales*	Phytozome 12	Bowman *et al.* (2017)	
Medicago truncatula	C_3	*Fabaceae*	*Fabales*	Phytozome 12	(Tang *et al.*, 2014)	DiMario *et al.* (2017)
Mesembryanthemum crystallinum	facultative CAM	*Aizoaceae*	*Caryophyllales*	1KP Transcriptomes	Matasci *et al.* (2014)	Cushman (1992)
Mirabilis jalapa	C_3	*Nyctaginaceae*	*Caryophyllales*	1KP Transcriptomes	Matasci *et al.* (2014)	Stata *et al.* (2014)
Mollugo cerviana	C_4	*Molluginaceae*	*Caryophyllales*	1KP Transcriptomes	Matasci *et al.* (2014)	Stata *et al.* (2014)
Mollugo nudicaulis	C_2	*Molluginaceae*	*Caryophyllales*	1KP Transcriptomes	Matasci *et al.* (2014)	Stata *et al.* (2014)
Mollugo pentaphylla	C_3	*Molluginaceae*	*Caryophyllales*	1KP Transcriptomes	Matasci *et al.* (2014)	Stata *et al.* (2014)
Mollugo verticillata	C_2	*Molluginaceae*	*Caryophyllales*	1KP Transcriptomes	Matasci *et al.* (2014)	Stata *et al.* (2014)
Oncidium sphacelatum	CAM	*Orchidaceae*	*Asparagales*	1KP Transcriptomes	Matasci *et al.* (2014)	Silvera *et al.* (2014)
Oryza sativa	C_3	*Poaceae*	*Poales*	Phytozome 12	Li *et al.* (2017)	Brkljacic *et al.* (2011)

CCM, carbon concentrating mechanism

Table S1. (*Continued.*)

Species	PS type	Family	Order	Data source		Photosynthesis information
Phalaenopsis equestris	obligate CAM	*Orchidaceae*	*Asparagales*	NCBI	Cai *et al.* (2014)	Cai *et al.* (2014)
Physcomitrella patens	C_3	*Funariaceae*	*Funariales*	Phytozome 12	Matasci *et al.* (2014)	Wang *et al.* (2013)
Physena madagascariensis	no data	*Physenaceae*	*Caryophyllales*	1KP Transcriptomes	Matasci *et al.* (2014)	
Phytolacca americana	C_3	*Phytolaccaceae*	*Caryophyllales*	1KP Transcriptomes	Matasci *et al.* (2014)	Wolfe-Bellin *et al.* (2006)
Phytolacca bogotensis	no data	*Phytolaccaceae*	*Caryophyllales*	1KP Transcriptomes	Matasci *et al.* (2014)	
Polycarpaea repens	C_3	*Polygonaceae*	*Caryophyllales*	1KP Transcriptomes	Matasci *et al.* (2014)	Kool (2012)
Polygonum convolvulus	C_3	*Polygonaceae*	*Caryophyllales*	1KP Transcriptomes	Matasci *et al.* (2014)	Advances in Agronomy (Google Books)
Populus trichocarpa	C_3	*Salicaceae*	*Malpighiales*	Phytozome 12	Du *et al.* (2015)	Yang *et al.* (2015*a*)
Portulaca amilis	C_4	*Portulacaceae*	*Caryophyllales*	1KP Transcriptomes	Matasci *et al.* (2014)	Voznesenskaya *et al.* (2017)
Portulaca cryptopetala	C_3-C_4	*Portulacaceae*	*Caryophyllales*	1KP Transcriptomes	Matasci *et al.* (2014)	Voznesenskaya *et al.* (2017)
Portulaca grandiflora	C_4	*Portulacaceae*	*Caryophyllales*	1KP Transcriptomes	Matasci *et al.* (2014)	Voznesenskaya *et al.* (2017)
Portulaca molokiniensis	C_4	*Portulacaceae*	*Caryophyllales*	1KP Transcriptomes	Matasci *et al.* (2014)	Voznesenskaya *et al.* (2017)
Portulaca oleracea	C_4-CAM	*Portulacaceae*	*Caryophyllales*	1KP Transcriptomes	Matasci *et al.* (2014)	Lara *et al.* (2004)
Portulaca suffrutescens	C_4	*Portulacaceae*	*Caryophyllales*	1KP Transcriptomes	Matasci *et al.* (2014)	Voznesenskaya *et al.* (2017)
Portulaca umbraticola	C_4	*Portulacaceae*	*Caryophyllales*	1KP Transcriptomes	Matasci *et al.* (2014)	Voznesenskaya *et al.* (2017)
Rhodiola crenulata	C_3	*Crassulaceae*	*Saxifragales*	www.gigadb.org	Fu *et al.* (2017)	https://eol.org/pages/2886822
Ricinus communis	C_3	*Euphorbiaceae*	*Malpighiales*	Phytozome 12	Chan *et al.* (2011)	Dai *et al.* (1992)
Sesuvium verrucosum	C_3	*Aizoaceae*	*Caryophyllales*	1KP Transcriptomes	Matasci *et al.* (2014)	Stata *et al.* (2014)
Setaria viridis	C4	*Poaceae*	*Poales*	1KP Transcriptomes	Matasci *et al.* (2014)	Brkljacic 2011
Silene latifolia	C_3	*Caryophyllaceae*	*Caryophyllales*	1KP Transcriptomes	Matasci *et al.* (2014)	https://eol.org/pages/590246
Sorghum bicolor	C_4	*Poaceae*	*Poales*	Phytozome 12	Cooper *et al.* (2019)	Brkljacic *et al.* (2011)
Spinacia oleracea	C_3	*Amaranthaceae*	*Caryophyllales*	NCBI	Dohm *et al.* (2014)	Raghavendra *et al.* (1995)
Suaeda aralocaspica	C_4	*Amaranthaceae*	*Caryophyllales*	www.gigadb.org	Wang *et al.* (2019*b*)	Wang *et al.* (2019*b*)
Tamarix chinensis	C_3	*Tamaricaceae*	*Caryophyllales*	1KP Transcriptomes	Matasci *et al.* (2014)	Bi and Xie (2015)

Table S1. (*Continued.*)

Species	PS type	Family	Order	Data source		Photosynthesis information
Trianthema portulacastrum	C_4	*Aizoaceae*	*Caryophyllales*	1KP Transcriptomes	Matasci *et al.* (2014)	Stata *et al.* (2014)
Triticum aestivum	C_3	*Poaceae*	*Poales*	Phytozome 12	The International Wheat Genome Sequencing Consortium (2014)	Brkljacic *et al.* (2011)
Vanilla planifolia	obligate CAM	*Orchidaceae*	*Asparagales*	1KP Transcriptomes	Matasci *et al.* (2014)	Christopher and Holtum (1996)
Vitis vinifera	C_3	*Vitaceae*	*Vitales*	Phytozome 12	The French-Italian Public Consortium for Grapevine Genome Characterization, (2007)	DiMario *et al.* (2017)
Vittaria lineata	facultative CAM	*Pteridaceae*	*Polypodiales*	1KP Transcriptomes	Matasci *et al.* (2014)	Minardi *et al.* (2014)
Yucca aloifolia	obligate CAM	*Asparagaceae*	*Asparagales*	Phytozome 12	JGI	Heyduk *et al.* (2016)
Zea mays	C_4	*Poaceae*	*Poales*	Phytozome 12	Schnable *et al.* (2009)	Brkljacic *et al.* (2011)

Table S2. Statistics of raw Nanopore reads used for *de novo* genome assembly of *Talinum triangulare*.

Flow cell	Pass reads	Bases	N_{50} length
1	1,962,503	16,645,767,699	22,988
2	1,235,269	13,045,496,067	28,747
3	1,265,588	14,370,375,600	30,510
4	1,167,309	15,569,237,462	38,440
5	537,363	5,152,770,000	30,900
Average	1,233,606	12,956,729,366	30,317
Total	6,168,032	64,783,646,828	-

Table S3. Statistics of the genome assembly (Flye & Pilon) obtained with QUAST.

Length parameters	**Length [bp]**
Total length	608,541,345
Largest contig	47,709,168
Total length (>= 0 bp)	608,590,492
Total length (>= 1000 bp)	608,580,813
Total length (>= 5000 bp)	608,380,484
Total length (>= 10000 bp)	608,135,854
Total length (>= 25000 bp)	607,347,907
Total length (>= 50000 bp)	606,788,954
N50	21,613,741
N75	11,389,798

Coverage parameters	
Average coverage	111x
Coverage >= 1x (%)	99.99
Number of Ns per 100 kbp	0.03

Count parameters	**Count**
Total number of contigs	234
Number of contigs (>= 0 bp)	234
Number of contigs (>= 1000 bp)	221
Number of contigs (>= 5000 bp)	162
Number of contigs (>= 10000 bp)	126
Number of contigs (>= 25000 bp)	77
Number of contigs (>= 50000 bp)	61
L50	9
L75	19

Table S4. RNA-seq mappings statistics on the genome assembly obtained with HISAT2. All details about RNA-sequencing and samples available in *Maleckova et al.* (2019)

Sample	Number of RNA-seq reads	Alignment rate (%)	Aligned more than once (%)
311	23,396,038	98.70	7.56
316	32,305,179	97.30	12.42
319	25,324,104	98.67	7.89
322	3,236,593	98.59	8.54
324	36,294,440	98.55	6.50
330	32,337,182	98.37	8.06
334	28,574,580	92.80	5.68
Average		**97.57**	**8.09**

Table S5. Statistics for Iso-Seq (PacBio) transcriptome sequencing and identification of high-quality transcripts with the isoseq3 pipeline.

SMRT cell	Raw data						isoseq3 pipeline	
	Total output [GB]	Total number of polymerase reads	Proportion of ZMWs with high quality reads	Mean insert length [kbp]	Mean polymerase read length [kbp]	N_{50} polymerase read length [kbp]	Number of circular consensus sequences (CCS)	Number of full-length non-chimeric reads (FLNC)
1	22.74	698,480	68.9%	3.87	32.55	70.71	503,782	364,162
2	29.22	750,071	74.0%	4.12	38.95	82.52	564,316	418,220
3	28.90	686,031	67.6%	4.29	42.13	87.23	521,743	391,336
4	29.43	732,905	72.2%	4.30	40.15	82.27	552,819	420,449
Average	27.57	716,872	70.7%	4.14	38.45	80.68	535,665	398,542
Total	110.29	2,867,487	-	-	-	-	2,142,660	1,594,167

Table S6. BLAST results for high quality transcripts (obtained with isoseq3 pipeline) with non-plant hits.

Transcript ID	Plant hit(s)	Max score	Total score	Query coverage [%]	E-value	Identitity [%]	Accession
transcript/4350 full_length_coverage=2;length=6331 ;num_subreads=19	yes	52.8	52.8	0	1.10E-01	96.77	LR597558.1
transcript/19461 full_length_coverage=2;length=4008 ;num_subreads=31	no	52.8	105	0	7.20E-02	100.00	LR597470.1
transcript/20367 full_length_coverage=2;length=3932 ;num_subreads=11	no	56.5	56.5	0	5.00E-03	100.00	CP017821.1
transcript/20536 full_length_coverage=10;length=387 3;num_subreads=60	no	58.4	58.4	0	2.00E-03	97.06	CP032601.1
transcript/22154 full_length_coverage=4;length=3744 ;num_subreads=60	no	58.4	58.4	0	1.00E-03	97.06	CP032601.1
transcript/22907 full_length_coverage=7;length=3724 ;num_subreads=60	no	60.2	60.2	1	4.00E-04	94.87	CP032601.1
transcript/24152 full_length_coverage=4;length=3567 ;num_subreads=50	no	60.2	60.2	1	4.00E-04	94.87	CP032601.1
transcript/41706 full_length_coverage=2;length=2847 ;num_subreads=21	no	60.2	60.2	1	3.00E-04	94.87	CP032601.1
transcript/44293 full_length_coverage=3;length=2807 ;num_subreads=47	no	60.2	60.2	1	3.00E-04	94.87	CP032601.1
transcript/47107 full_length_coverage=5;length=2726 ;num_subreads=60	no	56.5	56.5	1	4.00E-03	100.00	HE806324.1
transcript/47126 full_length_coverage=8;length=2723 ;num_subreads=60	no	56.5	56.5	1	4.00E-03	96.97	AB567734.1
transcript/47724 full_length_coverage=2;length=2704 ;num_subreads=37	yes	52.8	52.8	1	4.90E-02	100.00	LR132000.1
transcript/48846 full_length_coverage=8;length=2639 ;num_subreads=60	yes	52.8	52.8	1	4.70E-02	100.00	LR132000.1
transcript/49197 full_length_coverage=2;length=2677 ;num_subreads=59	no	56.5	56.5	1	7.00E-03	96.97	AB567734.1
transcript/49254 full_length_coverage=9;length=2691 ;num_subreads=60	no	56.5	56.5	1	7.00E-03	96.97	AB567734.1
transcript/53607 full_length_coverage=2;length=2534 ;num_subreads=60	no	56.5	56.5	1	7.00E-03	96.97	AB567734.1
transcript/56751 full_length_coverage=6;length=2467 ;num_subreads=60	no	54.7	54.7	1	2.40E-02	100.00	CP024765.1

Table S7. RNA-seq mappings statistics on the transcriptome assembly

Sample	Number of RNA-seq reads	Alignment rate (%)	Uniquely aligned (%)
311	32,305,179	77.2	21.4
312	28,147,275	78.9	19.0
313	32,860,874	78.7	18.7
314	30,117,770	78.4	18.6
315	29,684,734	77.9	19.2
316	32,365,938	78.0	18.9
317	36,065,244	78.5	19.9
318	28,928,321	76.8	19.4
319	23,396,038	76.7	19.9
320	33,736,335	77.0	20.2
321	27,317,333	77.1	20.0
322	25,324,104	70.7	17.1
323	22,793,855	71.0	18.5
324	36,294,440	70.9	18.4
325	21,457,579	69.6	18.3
326	28,953,152	69.2	18.0
327	29,388,774	64.7	18.4
328	31,108,522	65.5	18.0
329	31,764,569	65.9	18.4
330	32,337,182	64.4	17.9
331	29,662,032	68.7	15.3
332	3,458,718	64.4	16.4
333	30,598,746	67.6	15.1
334	28,574,580	67.7	17.1
Average		**72.31**	**18.42**

Table S8. Enriched motifs in promoter regions of differentially expressed genes ($q < 0.01$) as identified with the lmms package.
Enriched motifs were identified with DREME and subsequently annotated with Tomtom using the Plant Cistrome Database. False discovery rate adjustment was done according to Benjamini and Hochberg (1995). Only the motifs with q-value < 0.01 both for DREME and Tomtom are shown.
[a]Cluster, cluster based on lmms analysis and *k*-means clustering
[b]Pos, number of counts in the positive group (*i.e.* genes within an analysed cluster)
[c]Neg, number of counts in the negative group (*i.e.* randomly selected promoters of *Talinum triangulare*)
[d]Target ID, transcription factor predicted to bind to the given motif (Plant Cistrome Database nomenclature)

Cluster[a]	Motif	Pos[b]	Neg[c]	*q*-value	Target ID[d]	Optimal offset	*q*-value	Overlap
1	CCACAM	461	474	8.91E-07	TCP_tnt.TCP16_colamp_a_m1	5	5.27E-03	6
1	CCACAM	461	474	8.91E-07	TCP_tnt.At1g69690_colamp_a_m1	7	5.27E-03	6
1	CCACAM	461	474	8.91E-07	TCP_tnt.At1g72010_colamp_a_m1	7	5.27E-03	6
1	CCACAM	461	474	8.91E-07	TCP_tnt.PTF1_colamp_a_m1	7	5.27E-03	6
1	CCACAM	461	474	8.91E-07	TCP_tnt.At1g72010_col_m1	7	5.27E-03	6
1	CCACAM	461	474	8.91E-07	TCP_tnt.At5g08330_colamp_a_m1	7	5.27E-03	5
1	CCACAM	461	474	8.91E-07	TCP_tnt.PTF1_col_a_m1	7	5.27E-03	6
1	CCACAM	461	474	8.91E-07	TCP_tnt.TCP3_col_m1	7	5.27E-03	6
1	CCACAM	461	474	8.91E-07	TCP_tnt.At1g69690_col_a_m1	7	5.27E-03	6
1	CCACAM	461	474	8.91E-07	C3H_tnt.U2AF35B_col_a_m1	1	5.27E-03	6
1	CCACAM	461	474	8.91E-07	TCP_tnt.TCP20_colamp_a_m1	7	5.37E-03	5
1	CCACAM	461	474	8.91E-07	TCP_tnt.TCP3_colamp_a_m1	7	5.44E-03	5
1	CCACAM	461	474	8.91E-07	TCP_tnt.At5g08330_col_a_m1	9	6.02E-03	5
1	CCACAM	461	474	8.91E-07	TCP_tnt.TCP14_col_b_m1	7	6.29E-03	6
1	CCACAM	461	474	8.91E-07	TCP_tnt.TCP24_col_m1	7	6.31E-03	5
1	CCACAM	461	474	8.91E-07	TCP_tnt.TCP16_col_a_m1	7	6.75E-03	4
1	CCACAM	461	474	8.91E-07	TCP_tnt.At2g45680_colamp_a_m1	7	7.16E-03	4
1	CCACAM	461	474	8.91E-07	C3H_tnt.TZF9_col_a_m1	1	7.75E-03	5
1	CCACAM	461	474	8.91E-07	TCP_tnt.TCP7_col_a_m1	7	9.18E-03	4
1	CCACAM	461	474	8.91E-07	AP2EREBP_tnt.AT3G16280_colamp_a_m1	4	9.47E-03	6
1	CCACAM	461	474	8.91E-07	TCP_tnt.TCP24_colamp_a_m1	7	9.73E-03	4
1	AAAWGGA	520	551	8.91E-07	MADS_tnt.AGL25_colamp_a_m1	5	5.27E-03	7
1	AAAWGGA	520	551	8.91E-07	MADS_tnt.AGL6_col_a_m1	7	5.27E-03	7
1	AAAWGGA	520	551	8.91E-07	MADS_tnt.AGL16_colamp_a_m1	11	5.27E-03	7
1	AAAWGGA	520	551	8.91E-07	MADS_tnt.SVP_col_v3b_m1	4	5.27E-03	7
1	AAAWGGA	520	551	8.91E-07	MADS_tnt.AGL15_col_a_m1	9	5.27E-03	7
1	AAAWGGA	520	551	8.91E-07	MADS_tnt.AGL25_col_a_m1	7	5.27E-03	7
1	AAAWGGA	520	551	8.91E-07	MADS_tnt.SVP_colamp_v3b_m1	6	5.27E-03	7
1	AAAWGGA	520	551	8.91E-07	MADS_tnt.AGL16_col_a_m1	6	5.27E-03	7
1	AAAWGGA	520	551	8.91E-07	MADS_tnt.AGL13_col_b_m1	7	5.27E-03	7
1	AAAWGGA	520	551	8.91E-07	MADS_tnt.AGL15_colamp_a_m1	9	5.44E-03	7
1	AAAWGGA	520	551	8.91E-07	C2C2dof_tnt.dof42_col_a_m1	6	6.02E-03	7
1	AAAWGGA	520	551	8.91E-07	MADS_tnt.AGL63_col_a_m1	6	7.44E-03	7
1	CCATAW	623	696	3.18E-06	GeBP_tnt.AT1G66420_col_a_m1	8	6.38E-03	6

Table S8. (*Continued.*)

Cluster[a]	Motif	Pos[b]	Neg[c]	*q*-value	Target ID[d]	Optimal offset	*q*-value	Overlap
1	CCATAW	623	696	3.18E-06	MYBrelated_tnt.AT5G56840_col_a_m1	8	9.50E-03	6
1	CCATAW	623	696	3.18E-06	GeBP_tnt.AT1G66420_col_a_m1	8	6.38E-03	6
1	CCATAW	623	696	3.18E-06	MYBrelated_tnt.AT5G56840_col_a_m1	8	9.50E-03	6
1	CACCWC	335	338	9.30E-06	HMG_tnt.3XHMGBOX1_colamp_a_m1	8	5.27E-03	6
1	CACCWC	335	338	9.30E-06	AP2EREBP_tnt.CEJ1_col_a_m1	1	5.27E-03	6
1	CACCWC	335	338	9.30E-06	AP2EREBP_tnt.DREB26_col_a_m1	1	6.20E-03	6
1	CACCWC	335	338	9.30E-06	MYB_tnt.MYB55_col_m1	7	7.35E-03	6
1	CACCWC	335	338	9.30E-06	DBP_tnt.AT3G51470_col_a_m1	7	7.35E-03	6
1	CACCWC	335	338	9.30E-06	AP2EREBP_tnt.At1g77640_col_a_m1	14	8.11E-03	6
1	CACCWC	335	338	9.30E-06	AP2EREBP_tnt.AT1G44830_col_a_m1	1	8.11E-03	6
1	CACCWC	335	338	9.30E-06	AP2EREBP_tnt.DREB2_col_a_m1	1	8.97E-03	6
1	CACCWC	335	338	9.30E-06	AP2EREBP_tnt.At1g75490_col_a_m1	1	8.97E-03	6
1	CACCWC	335	338	9.30E-06	C2H2_tnt.At2g48100_col_b_m1	1	8.97E-03	5
1	CACCWC	335	338	9.30E-06	MADS_tnt.AGL55_col_a_m1	1	8.97E-03	5
1	CACCWC	335	338	9.30E-06	MADS_tnt.FEM111_col_a_m1	1	8.97E-03	5
1	CACCWC	335	338	9.30E-06	AP2EREBP_tnt.At5g65130_col_a_m1	1	9.47E-03	6
1	CACCWC	335	338	9.30E-06	AP2EREBP_tnt.At1g22810_col_m1	8	9.47E-03	6
1	CACCWC	335	338	9.30E-06	bHLH_tnt.bHLH31_col_m1	15	9.92E-03	6
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY17_colamp_a_m1	4	5.27E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY15_colamp_a_m1	3	5.27E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY28_colamp_a_m1	3	5.27E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY45_colamp_a_m1	3	5.27E-03	8
1	ACTCAASG	32	10	9.30E-06	zfGRF_tnt.AT3G42860_col_a_m1	3	5.27E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY65_colamp_a_m1	3	5.27E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY46_col_a_m1	3	5.27E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY70_col_m1	3	5.27E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY21_col_m1	3	5.27E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY30_col_a_m1	2	5.27E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY30_colamp_a_m1	3	5.27E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY17_col_a_m1	4	5.27E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY45_col_a_m1	4	5.27E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY71_col_a_m1	3	5.27E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY75_colamp_a_m1	3	5.27E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY7_col_m1	4	5.27E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY46_colamp_a_m1	3	5.27E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY7_colamp_a_m1	12	5.27E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY65_col_a_m1	3	5.27E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY8_col_m1	5	5.27E-03	8

Table S8. (*Continued.*)

Cluster[a]	Motif	Pos[b]	Neg[c]	*q*-value	Target ID[d]	Optimal offset	*q*-value	Overlap
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY14_colamp_a_m1	3	5.27E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY21_colamp_a_m1	3	5.27E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY24_colamp_a_m1	3	5.27E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY27_col_a_m1	3	5.27E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY15_col_b_m1	3	5.27E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY18_col_a_m1	3	5.27E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY33_col_a_m1	3	5.27E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY8_colamp_a_m1	4	5.27E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY14_col_a_m1	3	5.27E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY75_col_a_m1	3	5.27E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY55_col_a_m1	2	5.27E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY26_colamp_a_m1	3	5.27E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY26_col_a_m1	3	5.27E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY3_col_a_m1	3	5.27E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY11_col_a_m1	4	5.27E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY24_col_a_m1	5	5.27E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY43_col_a_m1	4	5.27E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY40_colamp_a_m1	5	5.37E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY25_col_a_m1	3	5.44E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY18_colamp_a_m1	5	5.44E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY22_colamp_a_m1	3	5.44E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY29_col_a_m1	3	5.51E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY43_colamp_a_m1	6	5.51E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY22_col_m1	3	5.51E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY50_col_a_m1	5	5.51E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY40_col_m1	5	5.92E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY27_colamp_a_m1	3	5.92E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY28_col_a_m1	3	6.29E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY31_colamp_a_m1	3	7.64E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY29_colamp_a_m1	3	8.18E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY20_col_a_m1	3	8.18E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY25_colamp_a_m1	5	8.46E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY6_colamp_a_m1	5	8.46E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY47_col_a_m1	1	8.83E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY31_col_a_m1	11	8.97E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY6_col_a_m1	11	9.14E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY50_colamp_a_m1	3	9.47E-03	8
1	ACTCAASG	32	10	9.30E-06	WRKY_tnt.WRKY59_col_a_m1	4	9.62E-03	8

Table S8. (*Continued.*)

Cluster[a]	Motif	Pos[b]	Neg[c]	*q*-value	Target ID[d]	Optimal offset	*q*-value	Overlap
1	CTGCCYAC	22	4	1.10E-05	MYB_tnt.MYB116_colamp_a_m1	1	5.84E-03	8
1	CTGCCYAC	22	4	1.10E-05	AP2EREBP_tnt.Rap210_colamp_a_m1	3	8.97E-03	8
3	TATTTTSA	104	300	3.78E-06	MYBrelated_tnt.LCL1_col_m1	3	5.27E-03	7
3	TATTTTSA	104	300	3.78E-06	MYBrelated_tnt.LCL1_colamp_a_m1	3	5.27E-03	8
3	TATTTTSA	104	300	3.78E-06	MYBrelated_tnt.EPR1_col_m1	5	5.27E-03	7
3	TATTTTSA	104	300	3.78E-06	MYBrelated_tnt.At3g09600_col_a_m1	3	5.27E-03	8
3	TATTTTSA	104	300	3.78E-06	MYBrelated_tnt.At3g09600_colamp_a_m1	3	5.27E-03	8
3	TATTTTSA	104	300	3.78E-06	MYBrelated_tnt.At5g52660_col_a_m1	3	5.27E-03	8
3	TATTTTSA	104	300	3.78E-06	CCAATHAP3_tnt.NFYB4_col_a_m1	-1	5.27E-03	7
3	TATTTTSA	104	300	3.78E-06	MYBrelated_tnt.At5g52660_colamp_a_m1	3	5.27E-03	8
3	TATTTTSA	104	300	3.78E-06	ND_tnt.AT2G28920_col_a_m1	3	5.27E-03	7
3	TATTTTSA	104	300	3.78E-06	MYBrelated_tnt.At4g01280_col_a_m1	3	5.27E-03	8
3	TATTTTSA	104	300	3.78E-06	MYBrelated_tnt.EPR1_colamp_a_m1	3	5.27E-03	8
3	TATTTTSA	104	300	3.78E-06	MYBrelated_tnt.LHY1_col_a_m1	3	5.27E-03	7
3	TATTTTSA	104	300	3.78E-06	MYBrelated_tnt.At4g01280_colamp_a_m1	3	5.27E-03	8
3	TATTTTSA	104	300	3.78E-06	MYBrelated_tnt.AT3G10113_col_a_m1	5	5.27E-03	8
3	TATTTTSA	104	300	3.78E-06	C2H2_tnt.At2g41835_col_b_m1	-2	5.27E-03	6
3	TATTTTSA	104	300	3.78E-06	MYBrelated_tnt.LHY1_colamp_a_m1	3	5.27E-03	8
3	TATTTTSA	104	300	3.78E-06	MYBrelated_tnt.RVE1_col_a_m1	5	5.27E-03	8
3	TATTTTSA	104	300	3.78E-06	MYBrelated_tnt.RVE1_colamp_a_m1	3	5.37E-03	7
3	TATTTTSA	104	300	3.78E-06	ND_tnt.AT1G63040_col_a_m1	-3	7.41E-03	5
4	GCCACR	139	211	8.91E-07	bZIP_tnt.bZIP16_col_v3a_m1	1	5.27E-03	6
4	GCCACR	139	211	8.91E-07	bZIP_tnt.GBF3_col_m1	1	5.27E-03	6
4	GCCACR	139	211	8.91E-07	bZIP_tnt.bZIP16_colamp_a_m1	1	5.27E-03	6
4	GCCACR	139	211	8.91E-07	bZIP_tnt.ABI5_colamp_v3b_m1	1	5.27E-03	6
4	GCCACR	139	211	8.91E-07	bZIP_tnt.AREB3_colamp_a_m1	1	5.27E-03	6
4	GCCACR	139	211	8.91E-07	bZIP_tnt.AREB3_col_v31_m1	1	5.27E-03	6
4	GCCACR	139	211	8.91E-07	bZIP_tnt.ABI5_col_v3h_m1	3	5.27E-03	6
4	GCCACR	139	211	8.91E-07	bZIP_tnt.bZIP68_col_a_m1	1	5.27E-03	6
4	GCCACR	139	211	8.91E-07	bZIP_tnt.bZIP42_colamp_a_m1	1	5.27E-03	6
4	GCCACR	139	211	8.91E-07	bZIP_tnt.bZIP44_colamp_a_m1	0	5.27E-03	6
4	GCCACR	139	211	8.91E-07	bZIP_tnt.ABF2_col_v3a_m1	6	5.27E-03	6
4	GCCACR	139	211	8.91E-07	bZIP_tnt.bZIP42_col_a_m1	0	5.27E-03	6
4	GCCACR	139	211	8.91E-07	E2FDP_tnt.E2FA_col_a_m1	7	5.27E-03	6
4	GCCACR	139	211	8.91E-07	bZIP_tnt.GBF5_colamp_a_m1	1	5.27E-03	6
4	GCCACR	139	211	8.91E-07	bZIP_tnt.bZIP43_col_a_m1	0	5.27E-03	6
4	GCCACR	139	211	8.91E-07	bZIP_tnt.bZIP48_col_a_m1	0	5.27E-03	6
4	GCCACR	139	211	8.91E-07	bZIP_tnt.bZIP44_col_m1	1	5.37E-03	6

Table S8. (*Continued.*)

Cluster[a]	Motif	Pos[b]	Neg[c]	*q*-value	Target ID[d]	Optimal offset	*q*-value	Overlap
4	GCCACR	139	211	8.91E-07	E2FDP_tnt.E2FA_colamp_a_m1	7	5.44E-03	6
4	GCCACR	139	211	8.91E-07	bZIP_tnt.GBF5_col_v3a_m1	1	5.44E-03	6
4	GCCACR	139	211	8.91E-07	bZIP_tnt.bZIP28_col_a_m1	1	5.44E-03	6
4	GCCACR	139	211	8.91E-07	bZIP_tnt.bZIP3_col_a_m1	1	5.44E-03	6
4	GCCACR	139	211	8.91E-07	bZIP_tnt.bZIP48_colamp_a_m1	1	6.02E-03	6
4	GCCACR	139	211	8.91E-07	bZIP_tnt.bZIP53_colamp_a_m1	1	6.02E-03	6
4	GCCACR	139	211	8.91E-07	bZIP_tnt.GBF3_colamp_m1	1	6.69E-03	6
4	GCCACR	139	211	8.91E-07	bZIP_tnt.bZIP53_col_m1	1	6.69E-03	6
4	GCCACR	139	211	8.91E-07	bZIP_tnt.GBF6_col_m1	1	7.35E-03	6
4	GCCACR	139	211	8.91E-07	TCP_tnt.TCP16_colamp_a_m1	4	7.35E-03	6
4	GCCACR	139	211	8.91E-07	bHLH_tnt.PIF7_col_a_m1	6	8.73E-03	6
4	GCCACR	139	211	8.91E-07	TCP_tnt.TCP16_col_a_m1	6	8.97E-03	5
4	GCCACR	139	211	8.91E-07	TCP_tnt.At1g72010_colamp_a_m1	6	9.00E-03	6
4	GCCACR	139	211	8.91E-07	bZIP_tnt.GBF6_colamp_a_m1	1	9.18E-03	6
4	GCCACR	139	211	8.91E-07	TCP_tnt.At5g08330_colamp_a_m1	6	9.47E-03	6
4	GCCACR	139	211	8.91E-07	E2FDP_tnt.E2FC_col_a_m1	8	9.47E-03	6
4	GCCACR	139	211	8.91E-07	TCP_tnt.At1g72010_col_m1	6	9.86E-03	6
4	GAGTGGGW	35	25	2.29E-06	TCP_tnt.At5g08330_colamp_a_m1	-1	5.27E-03	7
4	GAGTGGGW	35	25	2.29E-06	TCP_tnt.TCP20_colamp_a_m1	-1	5.27E-03	7
4	GAGTGGGW	35	25	2.29E-06	TCP_tnt.At1g72010_col_m1	0	5.27E-03	8
4	GAGTGGGW	35	25	2.29E-06	TCP_tnt.At1g72010_colamp_a_m1	0	5.27E-03	8
4	GAGTGGGW	35	25	2.29E-06	TCP_tnt.At2g45680_colamp_a_m1	-2	5.27E-03	6
4	GAGTGGGW	35	25	2.29E-06	TCP_tnt.TCP7_col_a_m1	-2	5.27E-03	6
4	GAGTGGGW	35	25	2.29E-06	TCP_tnt.At1g69690_colamp_a_m1	0	5.27E-03	8
4	GAGTGGGW	35	25	2.29E-06	TCP_tnt.TCP14_col_b_m1	7	5.27E-03	8
4	GAGTGGGW	35	25	2.29E-06	TCP_tnt.At5g08330_col_a_m1	-1	5.60E-03	7
4	GAGTGGGW	35	25	2.29E-06	TCP_tnt.At1g69690_col_a_m1	0	5.92E-03	8
4	GAGTGGGW	35	25	2.29E-06	TCP_tnt.At2g45680_col_b_m1	-1	7.82E-03	7
4	CCASTA	189	331	4.16E-06	bHLH_tnt.bHLH64_col_a_m1	1	5.92E-03	5
4	AGRTGGTC	26	16	8.14E-06	TCP_tnt.PTF1_colamp_a_m1	0	5.27E-03	8
4	AGRTGGTC	26	16	8.14E-06	TCP_tnt.TCP3_col_m1	0	5.27E-03	8
4	AGRTGGTC	26	16	8.14E-06	TCP_tnt.TCP3_colamp_a_m1	-1	5.27E-03	7
4	AGRTGGTC	26	16	8.14E-06	TCP_tnt.TCP24_col_m1	-1	5.27E-03	7
4	AGRTGGTC	26	16	8.14E-06	TCP_tnt.TCP17_colamp_a_m1	-2	5.27E-03	6
4	AGRTGGTC	26	16	8.14E-06	TCP_tnt.TCP24_colamp_a_m1	-2	5.27E-03	6
4	AGRTGGTC	26	16	8.14E-06	TCP_tnt.PTF1_col_a_m1	0	5.51E-03	8
4	AGRTGGTC	26	16	8.14E-06	TCP_tnt.TCP17_col_a_m1	-2	5.92E-03	6
4	GGYAGTCA	17	6	9.30E-06	WRKY_tnt.WRKY18_col_a_m1	0	7.08E-03	8

Table S8. (*Continued.*)

Cluster[a]	Motif	Pos[b]	Neg[c]	*q*-value	Target ID[d]	Optimal offset	*q*-value	Overlap
4	GGYAGTCA	17	6	9.30E-06	zfGRF_tnt.AT3G42860_col_a_m1	0	7.35E-03	8
4	GGYAGTCA	17	6	9.30E-06	WRKY_tnt.WRKY25_col_a_m1	0	8.57E-03	8
4	GGYAGTCA	17	6	9.30E-06	WRKY_tnt.WRKY65_col_a_m1	0	8.57E-03	8
4	GGYAGTCA	17	6	9.30E-06	WRKY_tnt.WRKY15_col_b_m1	0	8.90E-03	8
4	GGYAGTCA	17	6	9.30E-06	WRKY_tnt.WRKY31_colamp_a_m1	0	8.97E-03	8
4	GGYAGTCA	17	6	9.30E-06	WRKY_tnt.WRKY33_col_a_m1	0	8.97E-03	8
4	GGYAGTCA	17	6	9.30E-06	WRKY_tnt.WRKY26_colamp_a_m1	0	8.97E-03	8
4	GGYAGTCA	17	6	9.30E-06	WRKY_tnt.WRKY30_colamp_a_m1	0	9.14E-03	8
4	GGYAGTCA	17	6	9.30E-06	WRKY_tnt.WRKY18_colamp_a_m1	2	9.14E-03	8
4	GGYAGTCA	17	6	9.30E-06	WRKY_tnt.WRKY30_col_a_m1	-1	9.18E-03	7
4	GGYAGTCA	17	6	9.30E-06	WRKY_tnt.WRKY59_col_a_m1	1	9.18E-03	8
4	GGYAGTCA	17	6	9.30E-06	WRKY_tnt.WRKY29_col_a_m1	0	9.47E-03	8
4	GGYAGTCA	17	6	9.30E-06	WRKY_tnt.WRKY3_col_a_m1	0	9.47E-03	8
4	GGYAGTCA	17	6	9.30E-06	WRKY_tnt.WRKY50_colamp_a_m1	0	9.47E-03	8
4	GGYAGTCA	17	6	9.30E-06	WRKY_tnt.WRKY20_col_a_m1	0	9.50E-03	8
4	GGYAGTCA	17	6	9.30E-06	WRKY_tnt.WRKY55_col_a_m1	-1	9.61E-03	7
4	CWGGAATA	21	11	1.01E-05	G2like_tnt.KAN2_colamp_a_m1	0	5.27E-03	8
4	CWGGAATA	21	11	1.01E-05	G2like_tnt.AT2G20400_colamp_a_m1	0	5.27E-03	8
4	CWGGAATA	21	11	1.01E-05	G2like_tnt.KAN2_col_a_m1	0	5.27E-03	8
4	CWGGAATA	21	11	1.01E-05	Homeobox_tnt.PDF2_col_m1	0	5.27E-03	8
4	CWGGAATA	21	11	1.01E-05	G2like_tnt.At3g24120_colamp_a_m1	1	5.27E-03	8
4	CWGGAATA	21	11	1.01E-05	G2like_tnt.At5g29000_col_a_m1	1	5.37E-03	8
4	CWGGAATA	21	11	1.01E-05	G2like_tnt.At2g01060_col_m1	1	5.37E-03	8
4	CWGGAATA	21	11	1.01E-05	G2like_tnt.At1g68670_colamp_a_d1	-2	5.68E-03	6
4	CWGGAATA	21	11	1.01E-05	G2like_tnt.AT5G45580_col_a_m1	1	5.68E-03	8
4	CWGGAATA	21	11	1.01E-05	G2like_tnt.At2g03500_col_a_m1	-2	5.82E-03	6
4	CWGGAATA	21	11	1.01E-05	G2like_tnt.At3g12730_colamp_a_m1	1	6.76E-03	8
4	CWGGAATA	21	11	1.01E-05	G2like_tnt.At3g04030_colamp_a_m1	3	7.08E-03	8
4	CWGGAATA	21	11	1.01E-05	G2like_tnt.At3g12730_col_a_m1	2	7.35E-03	8
4	CWGGAATA	21	11	1.01E-05	HSF_tnt.HSF3_colamp_a_d1	0	7.35E-03	6
4	CWGGAATA	21	11	1.01E-05	G2like_tnt.AT5G45580_colamp_a_m1	2	7.85E-03	8
4	CWGGAATA	21	11	1.01E-05	E2FDP_tnt.DEL2_colamp_a_m1	7	8.11E-03	8
4	CWGGAATA	21	11	1.01E-05	G2like_tnt.At3g24120_col_a_m1	-2	8.46E-03	6
4	CWGGAATA	21	11	1.01E-05	E2FDP_tnt.DEL1_colamp_a_m1	7	8.70E-03	8
4	CWGGAATA	21	11	1.01E-05	G2like_tnt.At3g04030_col_a_m1	1	9.10E-03	8
4	CWGGAATA	21	11	1.01E-05	G2like_tnt.At2g01060_colamp_a_m1	3	9.73E-03	8
4	GATTCTSA	38	35	1.01E-05	G2like_tnt.At1g68670_colamp_a_d1	1	5.27E-03	6
4	GATTCTSA	38	35	1.01E-05	AP2EREBP_tnt.CRF10_colamp_a_m1	9	5.27E-03	8

Table S8. (*Continued.*)

Cluster[a]	Motif	Pos[b]	Neg[c]	*q*-value	Target ID[d]	Optimal offset	*q*-value	Overlap
4	GATTCTSA	38	35	1.01E-05	G2like_tnt.AT4G37180_col_a_m1	5	5.27E-03	6
4	GATTCTSA	38	35	1.01E-05	G2like_tnt.At3g24120_col_a_m1	2	5.27E-03	6
4	GATTCTSA	38	35	1.01E-05	G2like_tnt.AT4G37180_colamp_a_m1	5	5.27E-03	6
4	GATTCTSA	38	35	1.01E-05	G2like_tnt.AT1G49560_col_a_m1	6	5.27E-03	5
4	GATTCTSA	38	35	1.01E-05	G2like_tnt.AT1G49560_colamp_a_m1	9	5.27E-03	6
4	GATTCTSA	38	35	1.01E-05	G2like_tnt.At1g25550_col_m1	5	6.17E-03	6
4	GATTCTSA	38	35	1.01E-05	G2like_tnt.At2g03500_colamp_a_m1	8	6.48E-03	6
4	GCCATTSA	32	26	1.10E-05	bHLH_tnt.BIM2_colamp_m1	5	5.80E-03	8
5	ACSATTAC	22	12	9.30E-06	MYB_tnt.MYB81_colamp_a_m1	3	5.27E-03	8
5	ACSATTAC	22	12	9.30E-06	MYB_tnt.MYB118_colamp_a_m1	2	5.44E-03	8
5	ACSATTAC	22	12	9.30E-06	HB_tnt.ATHB15_col_a_m1	3	5.92E-03	8
5	ACSATTAC	22	12	9.30E-06	MYB_tnt.MYB119_col_a_m1	2	7.35E-03	8
5	ACSATTAC	22	12	9.30E-06	C2H2_tnt.AT2G15740_col_a_m1	1	7.56E-03	8
5	ACSATTAC	22	12	9.30E-06	HB_tnt.PHV_col_a_m1	3	7.76E-03	8
5	ACSATTAC	22	12	9.30E-06	MYB_tnt.MYB118_col_a_m1	6	8.46E-03	8
5	ACSATTAC	22	12	9.30E-06	MYB_tnt.MYB119_colamp_a_m1	2	8.75E-03	8
5	ACSATTAC	22	12	9.30E-06	MYB_tnt.MYB77_colamp_a_m1	8	9.04E-03	8
5	ACSATTAC	22	12	9.30E-06	MYB_tnt.MYB65_col200_a_m1	2	9.50E-03	8
6	GTSTTAAA	60	58	8.91E-07	Trihelix_tnt.GT3a_col_a_m1	5	5.27E-03	8
6	GTSTTAAA	60	58	8.91E-07	NAC_tnt.ANAC079_colamp_a_m1	4	5.27E-03	8
6	GTSTTAAA	60	58	8.91E-07	NAC_tnt.CUC1_colamp_a_m1	7	5.92E-03	8
6	GTSTTAAA	60	58	8.91E-07	NAC_tnt.NAM_col_v3a_m1	4	8.97E-03	8
6	GTSTTAAA	60	58	8.91E-07	NAC_tnt.ANAC034_col_m1	7	9.50E-03	8
6	ATTGATTS	69	80	3.18E-06	Homeobox_tnt.HAT1_colamp_a_m1	2	5.27E-03	8
6	ATTGATTS	69	80	3.18E-06	Homeobox_tnt.ATHB18_colamp_a_m1	2	5.27E-03	8
6	ATTGATTS	69	80	3.18E-06	Homeobox_tnt.WUS1_colamp_a_m1	2	5.27E-03	8
6	ATTGATTS	69	80	3.18E-06	Homeobox_ecoli.HAT2_col_v31_m1	1	5.27E-03	8
6	ATTGATTS	69	80	3.18E-06	Homeobox_tnt.HAT22_col_a_m1	0	5.27E-03	8
6	ATTGATTS	69	80	3.18E-06	HB_tnt.ANL2_colamp_a_m1	2	5.27E-03	8
6	ATTGATTS	69	80	3.18E-06	HB_tnt.ANL2_col_a_m1	2	5.27E-03	8
6	ATTGATTS	69	80	3.18E-06	Homeobox_tnt.ATHB7_col_a_m1	2	5.27E-03	8
6	ATTGATTS	69	80	3.18E-06	Homeobox_tnt.HAT1_col_a_m1	2	5.27E-03	8
6	ATTGATTS	69	80	3.18E-06	Homeobox_tnt.HDG1_colamp_a_m1	2	5.27E-03	8
6	ATTGATTS	69	80	3.18E-06	Homeobox_tnt.WUS1_col_a_m1	2	5.27E-03	8
6	ATTGATTS	69	80	3.18E-06	HB_tnt.HDG7_col_a_m1	2	5.27E-03	8
6	ATTGATTS	69	80	3.18E-06	Homeobox_tnt.HDG1_col100_a_m1	2	5.27E-03	8
6	ATTGATTS	69	80	3.18E-06	Homeobox_tnt.HAT2_colamp_a_m1	2	5.27E-03	8
6	ATTGATTS	69	80	3.18E-06	Homeobox_tnt.ATHB18_col_a_m1	2	5.27E-03	8

Table S8. (*Continued.*)

Cluster[a]	Motif	Pos[b]	Neg[c]	*q*-value	Target ID[d]	Optimal offset	*q*-value	Overlap
6	ATTGATTS	69	80	3.18E-06	HB_tnt.ATHB5_colamp_a_m1	1	5.37E-03	8
6	ATTGATTS	69	80	3.18E-06	Homeobox_tnt.EDT1_col_m1	2	5.68E-03	8
6	ATTGATTS	69	80	3.18E-06	HB_tnt.ATHB40_colamp_a_m1	6	7.34E-03	8
6	ATTGATTS	69	80	3.18E-06	HB_tnt.ATHB53_colamp_a_m1	7	7.66E-03	8
6	ATTGATTS	69	80	3.18E-06	HB_tnt.LMI1_colamp_a_m1	6	8.11E-03	7
6	ATTGATTS	69	80	3.18E-06	HB_tnt.ATHB21_col_a_m1	6	8.43E-03	8
6	ATTGATTS	69	80	3.18E-06	Homeobox_tnt.ATHB20_colamp_a_m1	2	8.46E-03	8
6	ATTGATTS	69	80	3.18E-06	HB_tnt.ATHB5_col_a_m1	2	8.46E-03	8
6	ATTGATTS	69	80	3.18E-06	ARID_tnt.At1g76110_col_a_m1	1	8.97E-03	8
6	ATTGATTS	69	80	3.18E-06	Homeobox_tnt.ATHB6_colamp_a_m1	2	9.47E-03	8
6	ATTGATTS	69	80	3.18E-06	HB_tnt.ATHB21_colamp_a_m1	12	9.50E-03	8
6	ATTGATTS	69	80	3.18E-06	Homeobox_tnt.ATHB6_col_a_m1	3	9.50E-03	8
6	TAMAAGGA	60	67	6.36E-06	G2like_tnt.AT5G45580_colamp_a_m1	-1	9.58E-03	7
6	TAMAAGGA	60	67	6.36E-06	MADS_tnt.SVP_col_v3b_m1	3	9.92E-03	8
7	AASATGGT	55	29	9.30E-06	GeBP_tnt.AT1G66420_col_a_m1	0	8.00E-03	7
7	AASATGGT	55	29	9.30E-06	MYBrelated_tnt.AT1G18960_col_a_m1	13	9.86E-03	7
7	RCTATGAA	71	44	1.01E-05	NA	NA	NA	NA

Table S9. Enriched motifs in promoter regions of selected orthogroups.
Enriched motifs were identified with DREME and subsequently annotated with Tomtom using the Plant Cistrome Database. False discovery rate adjustment was done according to Benjamini and Hochberg (1995). Only the motifs with *q*-value < 0.01 both for DREME and Tomtom are shown.
[a]OG, orthogroup
[b]Pos, number of counts in the positive group (*i.e.* genes within an analysed cluster)
[c]Neg, number of counts in the negative group (*i.e.* randomly selected promoters of *Talinum triangulare*)
[d]Target ID, transcription factor predicted to bind to the given motif (Plant Cistrome Database nomenclature)

OG[a]	Motif	Pos[a]	Neg[c]	*q*-value	Target ID[d]	Optimal offset	*q*-value	Overlap
PPC	ACACTMA	7	7	3.70E-05	C2C2YABBY_tnt.CRC_colamp_a_m1	21	2.74E-03	7
PPC	AYAAAAGA	5	1	4.26E-05	C2H2_tnt.SGR5_col_a_m1	1	5.48E-03	8
PPC	AYAAAAGA	5	1	4.26E-05	C2H2_tnt.At5g66730_colamp_a_m1	9	5.48E-03	8
PPC	AYAAAAGA	5	1	4.26E-05	C2H2_tnt.At5g66730_col_m1	9	5.48E-03	8
PPC	AYAAAAGA	5	1	4.26E-05	C2H2_tnt.MGP_colamp_a_m1	9	6.37E-03	8
PPC	AYAAAAGA	5	1	4.26E-05	C2H2_tnt.MGP_col_a_m1	9	6.44E-03	8
PPC	AYAAAAGA	5	1	4.26E-05	C2H2_tnt.IDD7_col_a_m1	9	6.85E-03	8
PPC	AYAAAAGA	5	1	4.26E-05	C2H2_tnt.IDD4_col_a_m1	9	6.92E-03	8
PPC	AYAAAAGA	5	1	4.26E-05	G2like_tnt.AT4G37180_col_a_m1	-1	7.00E-03	7
PPC	AYAAAAGA	5	1	4.26E-05	ABI3VP1_tnt.VRN1_colamp_a_m1	6	7.05E-03	8
PPC	AYAAAAGA	5	1	4.26E-05	C2H2_tnt.JKD_col_a_m1	9	7.06E-03	8
PPC	AYAAAAGA	5	1	4.26E-05	C2H2_tnt.NUC_colamp_a_m1	9	7.06E-03	8
PPC	AYAAAAGA	5	1	4.26E-05	C2H2_tnt.IDD2_colamp_a_m1	9	7.37E-03	8
PPC	AYAAAAGA	5	1	4.26E-05	C2H2_tnt.SGR5_colamp_a_m1	2	7.37E-03	8
PPC	AYAAAAGA	5	1	4.26E-05	ABI3VP1_tnt.VRN1_col_a_m1	7	7.60E-03	8
PPC	AYAAAAGA	5	1	4.26E-05	C2H2_tnt.At1g14580_col_a_m1	9	7.68E-03	8
PPC	AYAAAAGA	5	1	4.26E-05	C2H2_tnt.AtIDD11_colamp_a_m1	9	7.83E-03	8
PPC	AYAAAAGA	5	1	4.26E-05	REM_tnt.REM19_colamp_a_m1	7	8.01E-03	8
PPC	AYAAAAGA	5	1	4.26E-05	C2H2_tnt.NUC_col_a_m1	9	8.31E-03	8
PPC	AYAAAAGA	5	1	4.26E-05	C2C2dof_tnt.AT5G02460_col_a_m1	0	8.31E-03	8
PPC	AYAAAAGA	5	1	4.26E-05	C2H2_tnt.At1g14580_colamp_a_m1	9	8.31E-03	8
PPC	AYAAAAGA	5	1	4.26E-05	C2C2dof_tnt.OBP3_colamp_a_m1	0	8.53E-03	8
PPC	AYAAAAGA	5	1	4.26E-05	REM_tnt.REM19_col_a_m1	5	8.53E-03	8
PPC	AYAAAAGA	5	1	4.26E-05	C2C2dof_tnt.AT1G47655_colamp_a_m1	1	8.53E-03	8
PPC	AYAAAAGA	5	1	4.26E-05	C2C2dof_tnt.OBP3_col_a_m1	10	8.56E-03	8
PPC	AYAAAAGA	5	1	4.26E-05	MYBrelated_tnt.At4g01280_col_a_m1	0	8.58E-03	8
PPC	AYAAAAGA	5	1	4.26E-05	MYBrelated_tnt.LCL1_colamp_a_m1	0	8.58E-03	8
PPC	AYAAAAGA	5	1	4.26E-05	MYBrelated_tnt.At5g52660_col_a_m1	0	8.97E-03	8
PPC	AYAAAAGA	5	1	4.26E-05	G2like_tnt.AT4G37180_colamp_a_m1	-1	8.97E-03	7
PPC	AYAAAAGA	5	1	4.26E-05	C2C2dof_tnt.dof42_col_a_m1	4	9.18E-03	8
PPC	AYAAAAGA	5	1	4.26E-05	C2C2dof_tnt.dof43_colamp_a_m1	-1	9.18E-03	7
PPC	AYAAAAGA	5	1	4.26E-05	MYBrelated_tnt.EPR1_colamp_a_m1	0	9.18E-03	8
PPC	AYAAAAGA	5	1	4.26E-05	MYBrelated_tnt.LCL1_col_m1	-1	9.26E-03	7
PPC	AYAAAAGA	5	1	4.26E-05	C2H2_tnt.IDD5_colamp_a_m1	9	9.76E-03	8

Table S9. (*Continued.*)

OG[a]	Motif	Pos[a]	Neg[c]	*q*-value	Target ID[d]	Optimal offset	*q*-value	Overlap
PPC	AYAAAAGA	5	1	4.26E-05	G2like_tnt.AT1G49560_col_a_m1	0	9.76E-03	8
PPC	TGTCAWA	7	10	5.90E-05	GRF_tnt.GRF9_colamp_a_m1	0	4.07E-03	7
PPC	CATGAAAA	4	0	7.63E-05	CCAATHAP3_tnt.NFYB4_col_a_m1	-2	2.74E-03	6
PPC	CATGAAAA	4	0	7.63E-05	C2H2_tnt.At2g41835_col_b_m1	-1	4.07E-03	7
PPC	CATGAAAA	4	0	7.63E-05	G2like_tnt.At3g13040_col_b_m1	-1	4.07E-03	7
PPCK	AACATTA	3	0	7.63E-05	GeBP_tnt.AT1G66420_col_a_m1	0	6.24E-03	7
PPCK	AACATTA	3	0	7.63E-05	MYBrelated_tnt.AT1G18960_col_a_m1	13	7.83E-03	7
PPCK	AACATTA	3	0	7.63E-05	HB_tnt.PHV_col_a_m1	3	9.18E-03	7
PPCK	AACATTA	3	0	7.63E-05	GeBP_tnt.AT4G00250_col_a_m1	0	9.50E-03	7
PPCK	AACATTA	3	0	7.63E-05	HB_tnt.ATHB15_col_a_m1	3	9.50E-03	7
PPCK	AACATTA	3	0	7.63E-05	MYB_tnt.MYB67_col_a_m1	7	9.87E-03	7
PPCK	AATAAAAC	3	0	7.63E-05	G2like_tnt.At3g13040_col_b_m1	0	4.07E-03	8
PPCK	AATAAAAC	3	0	7.63E-05	Trihelix_tnt.AT1G76870_col_b_m1	1	4.07E-03	8
PPCK	AATAAAAC	3	0	7.63E-05	ABI3VP1_tnt.VRN1_colamp_a_m1	9	9.32E-03	8
APL	CAATAAC	4	3	7.63E-05	HB_tnt.LMI1_colamp_a_m1	3	4.07E-03	7
APL	CAATAAC	4	3	7.63E-05	HB_tnt.LMI1_col_a_m1	2	4.07E-03	7
APL	CAATAAC	4	3	7.63E-05	HB_tnt.ATHB53_col_a_m1	1	4.07E-03	7
APL	CAATAAC	4	3	7.63E-05	C2C2YABBY_tnt.CRC_colamp_a_m1	22	4.07E-03	7
APL	CAATAAC	4	3	7.63E-05	Homeobox_tnt.HAT5_col_a_m1	1	4.07E-03	7
APL	CAATAAC	4	3	7.63E-05	Homeobox_tnt.HAT5_colamp_a_m1	2	4.07E-03	7
APL	CAATAAC	4	3	7.63E-05	Homeobox_tnt.ATHB6_colamp_a_m1	1	4.53E-03	7
APL	CAATAAC	4	3	7.63E-05	Homeobox_tnt.ATHB13_colamp_a_m1	1	4.69E-03	7
APL	CAATAAC	4	3	7.63E-05	Homeobox_tnt.ATHB20_col_a_m1	1	4.69E-03	7
APL	CAATAAC	4	3	7.63E-05	HB_tnt.ATHB21_col_a_m1	4	5.48E-03	7
APL	CAATAAC	4	3	7.63E-05	Homeobox_tnt.ATHB13_col_a_m1	1	5.48E-03	7
APL	CAATAAC	4	3	7.63E-05	MYB_tnt.MYB56_col_a_m1	2	5.54E-03	7
APL	CAATAAC	4	3	7.63E-05	HB_tnt.ATHB53_colamp_a_m1	4	5.75E-03	7
APL	CAATAAC	4	3	7.63E-05	ARID_tnt.At3g13350_col_a_d1	0	5.90E-03	6
APL	CAATAAC	4	3	7.63E-05	HB_tnt.ATHB40_col_a_m1	4	6.03E-03	7
APL	CAATAAC	4	3	7.63E-05	HB_tnt.ATHB40_colamp_a_m1	4	6.22E-03	7
APL	CAATAAC	4	3	7.63E-05	Homeobox_tnt.ATHB20_colamp_a_m1	1	6.24E-03	7
APL	CAATAAC	4	3	7.63E-05	MYB_tnt.MYB56_colamp_a_m1	1	6.58E-03	7
APL	CAATAAC	4	3	7.63E-05	HB_tnt.ATHB5_col_a_m1	1	6.92E-03	7
APL	CAATAAC	4	3	7.63E-05	MYBrelated_tnt.AT1G18960_col_a_m1	9	7.20E-03	7
APL	CAATAAC	4	3	7.63E-05	HB_tnt.ATHB21_colamp_a_m1	4	7.37E-03	7
APL	CAATAAC	4	3	7.63E-05	HB_tnt.ATHB5_colamp_a_m1	1	7.83E-03	7
BAM	AAGACC	5	4	6.93E-05	NA	NA	NA	NA
ENO	CTCACTY	4	4	8.10E-05	C2H2_tnt.AT3G46070_col_a_m1	5	4.07E-03	7

Table S9. (*Continued.*)

OG[a]	Motif	Pos[a]	Neg[c]	*q*-value	Target ID[d]	Optimal offset	*q*-value	Overlap
FBA	TAGAGTWA	4	0	4.26E-05	ND_tnt.FRS9_colamp_a_m1	16	7.06E-03	8
FBA	TAGAGTWA	4	0	4.26E-05	C3H_tnt.At1g74370_col_a_m1	2	7.68E-03	8
PFK	AACCTAWT	4	2	5.22E-05	MYB_tnt.MYB57_colamp_a_d1	-1	5.10E-03	7
PFK	AACCTAWT	4	2	5.22E-05	MYB_tnt.MYB81_col_a_m1	2	5.75E-03	8
PFK	AACCTAWT	4	2	5.22E-05	MYB_tnt.MYB27_col_a_m1	2	8.31E-03	8
PFK	AACCTAWT	4	2	5.22E-05	GeBP_tnt.AT4G00250_col_a_m1	0	9.09E-03	8
PFK	AACCTAWT	4	2	5.22E-05	MYB_tnt.MYB27_colamp_a_m1	3	9.18E-03	8
PFK	AACCTAWT	4	2	5.22E-05	MYBrelated_tnt.AT5G56840_col_a_m1	0	9.90E-03	8
ABCB	TTTTCGR	10	27	4.26E-05	MYBrelated_tnt.AT3G10580_colamp_a_m1	8	4.07E-03	7
ABCB	TTTTCGR	10	27	4.26E-05	C3H_tnt.CDM1_colamp_a_m1	8	4.07E-03	7
ABCB	TTTTCGR	10	27	4.26E-05	G2like_tnt.At3g13040_col_b_m1	2	4.07E-03	7
ABCB	TTTTCGR	10	27	4.26E-05	CCAATHAP3_tnt.NFYB4_col_a_m1	1	4.07E-03	6
ABCB	TTTTCGR	10	27	4.26E-05	LOBAS2_tnt.LBD2_colamp_a_m1	8	4.07E-03	7
ABCB	TTTTCGR	10	27	4.26E-05	C2H2_tnt.At2g41835_col_b_m1	0	4.07E-03	7
ABCB	CARAGGTA	6	6	4.26E-05	EIL_tnt.EIN3_colamp_a_d1	0	9.50E-03	7
ABCB	AGCYAAAT	7	11	4.26E-05	RWPRK_tnt.NLP7_col_a_m1	7	6.80E-03	8
ABCB	TATCTTYA	7	11	4.26E-05	MYBrelated_tnt.EPR1_col_m1	5	4.07E-03	7
ABCB	TATCTTYA	7	11	4.26E-05	MYBrelated_tnt.At3g11280_col_a_m1	5	4.07E-03	8
ABCB	TATCTTYA	7	11	4.26E-05	MYBrelated_tnt.At5g05790_col_a_m1	5	4.07E-03	8
ABCB	TATCTTYA	7	11	4.26E-05	MYBrelated_tnt.At5g08520_colamp_a_m1	3	4.07E-03	8
ABCB	TATCTTYA	7	11	4.26E-05	MYBrelated_tnt.At3g11280_colamp_a_m1	8	4.07E-03	8
ABCB	TATCTTYA	7	11	4.26E-05	MYBrelated_tnt.LCL1_col_m1	5	4.07E-03	5
ABCB	TATCTTYA	7	11	4.26E-05	MYBrelated_tnt.LCL1_colamp_a_m1	6	4.53E-03	5
ABCB	TATCTTYA	7	11	4.26E-05	MYBrelated_tnt.At5g58900_colamp_a_m1	5	4.69E-03	8
ABCB	TATCTTYA	7	11	4.26E-05	MYBrelated_tnt.LHY1_col_a_m1	5	4.75E-03	5
ABCB	TATCTTYA	7	11	4.26E-05	MYBrelated_tnt.RVE1_col_a_m1	7	5.03E-03	7
ABCB	TATCTTYA	7	11	4.26E-05	ND_tnt.AT2G28920_col_a_m1	5	5.03E-03	5
ABCB	TATCTTYA	7	11	4.26E-05	MYBrelated_tnt.At5g52660_col_a_m1	6	5.06E-03	5
ABCB	TATCTTYA	7	11	4.26E-05	MYBrelated_tnt.At5g52660_colamp_a_m1	7	5.15E-03	5
ABCB	TATCTTYA	7	11	4.26E-05	MYBrelated_tnt.At4g01280_col_a_m1	6	5.15E-03	5
ABCB	TATCTTYA	7	11	4.26E-05	MYBrelated_tnt.At5g08520_col_b_m1	3	5.17E-03	8
ABCB	TATCTTYA	7	11	4.26E-05	MYBrelated_tnt.At3g09600_colamp_a_m1	6	5.40E-03	5
ABCB	TATCTTYA	7	11	4.26E-05	MYBrelated_tnt.At3g09600_col_a_m1	7	5.48E-03	5
ABCB	TATCTTYA	7	11	4.26E-05	G2like_tnt.AT4G37180_colamp_a_m1	2	5.48E-03	8
ABCB	TATCTTYA	7	11	4.26E-05	MYBrelated_tnt.EPR1_colamp_a_m1	6	5.48E-03	5
ABCB	TATCTTYA	7	11	4.26E-05	MYBrelated_tnt.At5g58900_col_a_m1	7	5.48E-03	8
SWEET	AAKAAATA	15	56	5.24E-06	MYBrelated_tnt.EPR1_colamp_a_m1	0	4.07E-03	8
SWEET	AAKAAATA	15	56	5.24E-06	MYBrelated_tnt.LCL1_colamp_a_m1	0	4.07E-03	8

Table S9. (*Continued.*)

OG[a]	Motif	Pos[a]	Neg[c]	*q*-value	Target ID[d]	Optimal offset	*q*-value	Overlap
SWEET	AAKAAATA	15	56	5.24E-06	MYBrelated_tnt.At5g52660_col_a_m1	0	4.07E-03	8
SWEET	AAKAAATA	15	56	5.24E-06	MYBrelated_tnt.At4g01280_col_a_m1	0	4.07E-03	8
SWEET	AAKAAATA	15	56	5.24E-06	MYBrelated_tnt.At5g52660_colamp_a_m1	1	4.07E-03	8
SWEET	AAKAAATA	15	56	5.24E-06	ND_tnt.AT2G28920_col_a_m1	-1	4.07E-03	7
SWEET	AAKAAATA	15	56	5.24E-06	MYBrelated_tnt.At3g09600_col_a_m1	1	4.07E-03	8
SWEET	AAKAAATA	15	56	5.24E-06	MYBrelated_tnt.At3g09600_colamp_a_m1	0	4.07E-03	8
SWEET	AAKAAATA	15	56	5.24E-06	MYBrelated_tnt.LHY1_col_a_m1	-1	4.07E-03	7
SWEET	AAKAAATA	15	56	5.24E-06	MYBrelated_tnt.AT3G10113_col_a_m1	2	4.07E-03	8
SWEET	AAKAAATA	15	56	5.24E-06	MYBrelated_tnt.LHY1_colamp_a_m1	0	4.07E-03	8
SWEET	AAKAAATA	15	56	5.24E-06	MYBrelated_tnt.LCL1_col_m1	-1	4.07E-03	7
SWEET	AAKAAATA	15	56	5.24E-06	Homeobox_tnt.ATHB20_col_a_m1	2	4.07E-03	8
SWEET	AAKAAATA	15	56	5.24E-06	MYBrelated_tnt.At4g01280_colamp_a_m1	0	4.07E-03	8
SWEET	AAKAAATA	15	56	5.24E-06	MYBrelated_tnt.RVE1_col_a_m1	1	4.07E-03	8
SWEET	AAKAAATA	15	56	5.24E-06	MYBrelated_tnt.EPR1_col_m1	-1	4.07E-03	7
SWEET	AAKAAATA	15	56	5.24E-06	Homeobox_tnt.HAT5_col_a_m1	2	4.83E-03	8
SWEET	AAKAAATA	15	56	5.24E-06	ARID_tnt.At3g13350_col_a_d1	1	5.48E-03	5
SWEET	AAKAAATA	15	56	5.24E-06	Homeobox_tnt.ATHB13_colamp_a_m1	2	5.48E-03	8
SWEET	AAKAAATA	15	56	5.24E-06	HB_tnt.ATHB53_col_a_m1	2	5.54E-03	8
NHX	CTTTTGCA	4	0	3.55E-05	C2C2dof_tnt.COG1_colamp_a_m1	0	9.87E-03	8
ARF	TTCYAAA	13	61	1.44E-05	HSF_tnt.HSFB3_colamp_a_m1	4	4.07E-03	7
ARF	TTCYAAA	13	61	1.44E-05	HSF_tnt.HSF6_col_a_m1	5	4.07E-03	7
ARF	TTCYAAA	13	61	1.44E-05	MADS_tnt.AGL13_col_b_m1	1	5.75E-03	7
ARF	TTCYAAA	13	61	1.44E-05	HSF_tnt.HSFC1_colamp_a_m1	3	5.97E-03	7
ARF	TTCYAAA	13	61	1.44E-05	C2H2_tnt.At2g41835_col_b_m1	0	6.03E-03	7
ARF	TTCYAAA	13	61	1.44E-05	HSF_tnt.HSFA6A_colamp_a_m1	0	6.10E-03	7
ARF	TTCYAAA	13	61	1.44E-05	HSF_tnt.HSFB4_col_a_m1	1	6.41E-03	7
ARF	TTCYAAA	13	61	1.44E-05	BSD_tnt.AT1G10720_col_a_m1	6	6.92E-03	7
ARF	TTCYAAA	13	61	1.44E-05	HSF_tnt.HSF7_col_a_m1	0	6.92E-03	7
ARF	TTCYAAA	13	61	1.44E-05	HSF_tnt.HSF7_colamp_a_m1	6	6.92E-03	7
ARF	TTCYAAA	13	61	1.44E-05	HSF_tnt.HSF3_colamp_a_d1	0	7.06E-03	6
ARF	TTCYAAA	13	61	1.44E-05	ND_tnt.FRS9_colamp_a_m1	9	7.06E-03	7
ARF	TTCYAAA	13	61	1.44E-05	HSF_tnt.HSFA1E_col_a_m1	6	7.06E-03	7
ARF	TTCYAAA	13	61	1.44E-05	HSF_tnt.HSFA6B_colamp_a_m1	6	7.06E-03	7
ARF	TTCYAAA	13	61	1.44E-05	MADS_tnt.AGL15_col_a_m1	3	7.23E-03	7
ARF	TTCYAAA	13	61	1.44E-05	Orphan_tnt.AT1G23810_col_a_m1	0	7.37E-03	7
ARF	TTCYAAA	13	61	1.44E-05	MADS_tnt.AGL63_col_a_m1	0	7.37E-03	7
ARF	TTCYAAA	13	61	1.44E-05	HSF_tnt.HSFA6A_col_a_m1	5	7.86E-03	7
ARF	TTCYAAA	13	61	1.44E-05	ND_tnt.AT1G63040_col_a_m1	0	8.25E-03	6

Table S9. (*Continued.*)

OG[a]	Motif	Pos[a]	Neg[c]	*q*-value	Target ID[d]	Optimal offset	*q*-value	Overlap
ARF	AACTTMTA	5	1	1.15E-05	NAC_tnt.ANAC005_col_a_m1	1	5.47E-03	8
ARF	AACTTMTA	5	1	1.15E-05	HSF_tnt.HSF3_col_a_m1	2	6.00E-03	8
ARF	AACTTMTA	5	1	1.15E-05	HSF_tnt.HSF6_col_a_m1	2	6.00E-03	8
ARF	AACTTMTA	5	1	1.15E-05	HSF_tnt.HSFA1E_col_a_m1	3	6.03E-03	8
ARF	AACTTMTA	5	1	1.15E-05	HSF_tnt.HSFC1_col_a_m1	2	6.19E-03	8
ARF	AACTTMTA	5	1	1.15E-05	HSF_tnt.HSF7_col_a_m1	3	6.25E-03	8
ARF	AACTTMTA	5	1	1.15E-05	HSF_tnt.HSFB4_col_a_m1	3	6.45E-03	8
ARF	AACTTMTA	5	1	1.15E-05	HSF_tnt.HSFA1E_colamp_a_m1	3	6.80E-03	8
ARF	AACTTMTA	5	1	1.15E-05	HSF_tnt.HSFB3_colamp_a_m1	1	6.80E-03	8
ARF	AACTTMTA	5	1	1.15E-05	HSF_tnt.HSF7_colamp_a_m1	3	6.81E-03	8
ARF	AACTTMTA	5	1	1.15E-05	HSF_tnt.HSFB3_col_a_m1	3	7.06E-03	8
ARF	AACTTMTA	5	1	1.15E-05	Orphan_tnt.AT1G23810_col_a_m1	7	8.54E-03	8
ARF	AACTTMTA	5	1	1.15E-05	HSF_tnt.HSFC1_colamp_a_m1	3	8.97E-03	8
ARF	AACTTMTA	5	1	1.15E-05	HSF_tnt.HSFA6B_colamp_a_m1	3	9.21E-03	8
ARF	AACTTMTA	5	1	1.15E-05	ND_tnt.AGL95_col_a_m1	3	9.35E-03	8
ARF	AACTTMTA	5	1	1.15E-05	S1Falike_tnt.AT3G09735_col_a_m1	3	9.72E-03	8
ARF	ACTAATTA	4	1	4.26E-05	ZFHD_tnt.ATHB23_colamp_a_m1	1	3.16E-03	8
ARF	ACTAATTA	4	1	4.26E-05	ZFHD_tnt.ATHB25_col_a_m1	2	4.07E-03	8
ARF	ACTAATTA	4	1	4.26E-05	HB_tnt.WOX11_col_a_m1	0	4.07E-03	8
ARF	ACTAATTA	4	1	4.26E-05	ZFHD_tnt.ATHB25_colamp_a_m1	1	4.07E-03	8
ARF	ACTAATTA	4	1	4.26E-05	ZFHD_tnt.ATHB24_colamp_a_m1	2	4.07E-03	8
ARF	ACTAATTA	4	1	4.26E-05	ZFHD_tnt.ATHB33_col_a_m1	0	4.07E-03	8
ARF	ACTAATTA	4	1	4.26E-05	ZFHD_tnt.ATHB24_col_a_m1	2	4.07E-03	8
ARF	ACTAATTA	4	1	4.26E-05	ZFHD_tnt.ATHB33_colamp_a_m1	0	4.07E-03	8
ARF	ACTAATTA	4	1	4.26E-05	ZFHD_tnt.ATHB34_col_a_m1	7	4.07E-03	8
ARF	ACTAATTA	4	1	4.26E-05	Homeobox_tnt.ATHB20_col_a_m1	2	4.07E-03	8
ARF	ACTAATTA	4	1	4.26E-05	ZFHD_tnt.ATHB23_col_b_m1	9	4.07E-03	8
ARF	ACTAATTA	4	1	4.26E-05	Homeobox_tnt.HAT5_col_a_m1	2	4.07E-03	8
ARF	ACTAATTA	4	1	4.26E-05	Homeobox_tnt.ATHB13_colamp_a_m1	2	4.07E-03	8
ARF	ACTAATTA	4	1	4.26E-05	HB_tnt.ATHB53_col_a_m1	2	4.10E-03	8
ARF	ACTAATTA	4	1	4.26E-05	Homeobox_tnt.ATHB13_col_a_m1	2	4.53E-03	8
ARF	ACTAATTA	4	1	4.26E-05	ZFHD_tnt.ATHB34_colamp_a_m1	0	4.83E-03	8
ARF	ACTAATTA	4	1	4.26E-05	Homeobox_tnt.HAT5_colamp_a_m1	3	5.03E-03	8
ARF	ACTAATTA	4	1	4.26E-05	Homeobox_tnt.HDG1_colamp_a_m1	2	5.03E-03	8
ARF	ACTAATTA	4	1	4.26E-05	Homeobox_tnt.ATHB20_colamp_a_m1	2	5.10E-03	8
ARF	ACTAATTA	4	1	4.26E-05	HB_tnt.ANL2_col_a_m1	2	5.10E-03	8
ARF	AAATGTYT	8	20	4.26E-05	MYBrelated_tnt.LCL1_col_m1	2	4.07E-03	8
ARF	AAATGTYT	8	20	4.26E-05	MYBrelated_tnt.LCL1_colamp_a_m1	3	4.07E-03	8

Table S9. (*Continued.*)

OG[a]	Motif	Pos[a]	Neg[c]	*q*-value	Target ID[d]	Optimal offset	*q*-value	Overlap
ARF	AAATGTYT	8	20	4.26E-05	MYBrelated_tnt.LHY1_col_a_m1	2	4.07E-03	8
ARF	AAATGTYT	8	20	4.26E-05	ND_tnt.AT2G28920_col_a_m1	2	4.07E-03	8
ARF	AAATGTYT	8	20	4.26E-05	MYBrelated_tnt.EPR1_col_m1	2	4.07E-03	8
ARF	AAATGTYT	8	20	4.26E-05	MYBrelated_tnt.At5g52660_col_a_m1	3	4.07E-03	8
ARF	AAATGTYT	8	20	4.26E-05	MYBrelated_tnt.At4g01280_col_a_m1	3	4.07E-03	8
ARF	AAATGTYT	8	20	4.26E-05	MYBrelated_tnt.At5g52660_colamp_a_m1	4	4.07E-03	8
ARF	AAATGTYT	8	20	4.26E-05	MYBrelated_tnt.EPR1_colamp_a_m1	3	4.07E-03	8
ARF	AAATGTYT	8	20	4.26E-05	MYBrelated_tnt.At3g09600_col_a_m1	4	4.07E-03	8
ARF	AAATGTYT	8	20	4.26E-05	MYBrelated_tnt.At3g09600_colamp_a_m1	3	4.07E-03	8
ARF	AAATGTYT	8	20	4.26E-05	MYBrelated_tnt.At4g01280_colamp_a_m1	3	4.10E-03	8
ARF	AAATGTYT	8	20	4.26E-05	MYBrelated_tnt.LHY1_colamp_a_m1	3	4.83E-03	8
ARF	AAATGTYT	8	20	4.26E-05	MYBrelated_tnt.RVE1_col_a_m1	4	5.10E-03	8
ARF	AAATGTYT	8	20	4.26E-05	MYBrelated_tnt.AT3G10113_col_a_m1	5	5.48E-03	8
ARF	AAATGTYT	8	20	4.26E-05	REMB3_tnt.AT2G31460_col_a_m1	4	5.48E-03	8
ARF	AAATGTYT	8	20	4.26E-05	CPP_tnt.AT2G20110_col_a_m1	4	6.92E-03	8
ARF	ATWTGGAG	4	1	4.26E-05	MYBrelated_tnt.At5g47390_colamp_a_m1	0	8.53E-03	8
ARF	ATWTGGAG	4	1	4.26E-05	MYBrelated_tnt.AT5G61620_col_a_m1	1	9.18E-03	8
ARF	ATWTGGAG	4	1	4.26E-05	C2C2YABBY_tnt.CRC_colamp_a_m1	8	9.21E-03	8
ARF	ATWTGGAG	4	1	4.26E-05	MYBrelated_tnt.AT5G56840_col_a_m1	1	9.46E-03	8
ARF	ATWTGGAG	4	1	4.26E-05	E2FDP_tnt.E2FA_col_a_m1	0	9.93E-03	8
ARF	CCGATWG	4	1	4.26E-05	C3H_tnt.CDM1_colamp_a_m1	2	7.37E-03	7
ARF	CCGATWG	4	1	4.26E-05	LOBAS2_tnt.LBD2_col_a_m1	2	7.37E-03	7
ARF	CCGATWG	4	1	4.26E-05	MYBrelated_tnt.AT3G10580_colamp_a_m1	2	7.37E-03	7
ARF	CCGATWG	4	1	4.26E-05	C2C2gata_tnt.GATA20_col_a_m1	3	8.31E-03	5
ARF	CCGATWG	4	1	4.26E-05	LOBAS2_tnt.LBD2_colamp_a_m1	2	9.18E-03	7
ARF	TCGACAWA	4	1	4.26E-05	C3H_tnt.U2AF35B_col_a_m1	0	4.07E-03	7
ARF	TCGACAWA	4	1	4.26E-05	C3H_tnt.TZF9_col_a_m1	0	4.07E-03	6
ARF	TCGACAWA	4	1	4.26E-05	ARF_ecoli.MP_col_m1	1	4.07E-03	8
ARF	TCGACAWA	4	1	4.26E-05	AP2EREBP_tnt.DDF1_colamp_a_m1	6	4.07E-03	7
ARF	TCGACAWA	4	1	4.26E-05	C2H2_tnt.IDD2_col_a_m1	6	4.07E-03	8
ARF	TCGACAWA	4	1	4.26E-05	C2H2_tnt.IDD5_col_m1	6	4.07E-03	8
ARF	TCGACAWA	4	1	4.26E-05	C2H2_tnt.NUC_colamp_a_m1	6	4.07E-03	8
ARF	TCGACAWA	4	1	4.26E-05	C2H2_tnt.AtIDD11_colamp_a_m1	6	5.03E-03	8
ARF	TCGACAWA	4	1	4.26E-05	C2H2_tnt.MGP_colamp_a_m1	6	5.03E-03	8
ARF	TCGACAWA	4	1	4.26E-05	AP2EREBP_tnt.AT1G77200_colamp_a_m1	4	5.03E-03	8
ARF	TCGACAWA	4	1	4.26E-05	AP2EREBP_tnt.DDF2_col_a_m1	6	5.03E-03	8
ARF	TCGACAWA	4	1	4.26E-05	C2H2_tnt.At5g66730_colamp_a_m1	6	5.03E-03	8
ARF	TCGACAWA	4	1	4.26E-05	AP2EREBP_tnt.Rap210_colamp_a_m1	6	5.03E-03	8

Table S9. (*Continued.*)

OG[a]	Motif	Pos[a]	Neg[c]	*q*-value	Target ID[d]	Optimal offset	*q*-value	Overlap
ARF	TCGACAWA	4	1	4.26E-05	C2H2_tnt.IDD7_col_a_m1	6	5.03E-03	8
ARF	TCGACAWA	4	1	4.26E-05	AP2EREBP_tnt.AT3G16280_colamp_a_m1	3	5.10E-03	8
ARF	TCGACAWA	4	1	4.26E-05	C2H2_tnt.NUC_col_a_m1	6	5.10E-03	8
ARF	TCGACAWA	4	1	4.26E-05	AP2EREBP_tnt.DEAR5_colamp_a_m1	3	5.10E-03	8
ARF	TCGACAWA	4	1	4.26E-05	AP2EREBP_tnt.DDF1_col_a_m1	6	5.10E-03	8
ARF	TCGACAWA	4	1	4.26E-05	C2H2_tnt.At5g66730_col_m1	6	5.10E-03	8
ARF	TCGACAWA	4	1	4.26E-05	AP2EREBP_tnt.DEAR3_colamp_a_m1	5	5.10E-03	8
ARF	TCGACAWA	4	1	4.26E-05	C2H2_tnt.At1g14580_col_a_m1	6	5.43E-03	8
ARF	TCGACAWA	4	1	4.26E-05	AP2EREBP_tnt.AT3G60490_col_a_m1	3	5.48E-03	8

Table S10. Orthologs of *Talinum triangulare* and at least one other facultative CAM species (OrthoFinder analysis) and their assignment to clusters with altered temporal patterns upon ABA treatment.

Imms cluster	Number of genes	Number of ORFs
1	72	120
2	11	15
3	6	11
4	20	27
5	26	30
6	22	27
7	97	147
Total DEGs	254	377
Not a DEG	1347	2062
Grand total	1550	2439

Table S11. Statistics for genome assembly obtained with SMARTdenovo

Length parameters	Length [bp]
Total length	621,661,000
N50	15,160,000
Other parametrs	
L50	13
genome in scaffolds > 50 kb	99.96%

Acknowledgements

I would like to express my sincere gratitude to **Prof. Dr. Andreas Weber** for giving me the opportunity and trust to work on a project which became both a challenge and fun. I appreciate his consistent support in form of advice and financial means making lots of sequencing as well as my attendance at numerous conferences possible.

Very special thanks go to **Dr. Dominik Brilhaus**, who supported me with the initial bureaucracy, introduced me to the laboratory, the world of CAM and large data, who absolutely unselfishly shared his experience with me, who spent long hours discussing the project with me and was supportive when my stress levels became too high.

Furthermore, I would like to thank the remaining members of my thesis committee: **Prof. Dr. Matias Zurbriggen** and **Dr. Shizue Matsubara, PD** for their insightful comments.

I thank **apl. Prof. Dr. Peter Jahns** for allowing me to use his PAM equipment whenever and for how long needed.

I am grateful for the opportunities provided by the **iGRAD-Plant** program: annual retreats, seminars with other PhD students and support by attendance of workshops. I would like to thank **Dr. Petra Fackendahl** for being in touch with me starting with the admission process, for helping me during the first weeks and months in Germany as well as throughout the whole time afterwards.

I would like to thank **Anja Nöcker** for her patience and support in the field of bureaucracy.

I appreciate the huge contribution of **Raja Stracke** to the genome sequencing. I would like to thank **Dipl. Ing. Katharina Raba** (UKD) for performing the flow cytometry analyses. I thank **Samantha Flachbart** for her assistance throughout the whole time.

I would like to thank the **team of gardeners**, who grew tens and maybe even hundreds of plants, which I could spray and grind afterwards.

I am glad I could work with great colleagues the whole time. **Maria Graf, Elisabeth Klemp** and **Katrin Weber** from the MS Laboratory were patient with me and extremely helpful during metabolite measurements. I appreciate the enjoyable cooperation with **Thomas Wrobel** while preparing a manuscript together. I thank **Nils Koppers** for the introduction to the Linux world and helping me out with setting up new software. I appreciate discussions and input from **Meng-Ying Lin** and **Alisandra Denton**.

I thank **Franziska Kuhnert, Anastasija Plett** and **Marc-Sven Röll** as well as several graduated colleagues for introducing me to the laboratory routines and always having a piece of advice for me. On a more personal note, I appreciate the patience of all my colleagues when I first started speaking German, being such great teachers – may it be scientific or less academic vocabulary – and introducing me to the German culture.

Last but not least, I thank **Deutscher Akademischer Austauschdienst (DAAD)** for the support in form of a graduate fellowship and beyond.

I thank my family and **C. R**. for his support and very special friendship.