Comparative genomics can provide new insights into the evolutionary mechanisms and gene function in CAM plants

Xiaohan Yang1,2,*, Degao Liu3, Timothy J. Tschaplinski1,2, and Gerald A. Tuskan1,2,

1 Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA
2 The Center for Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA
3 Department of Genetics, Cell Biology and Development and Center for Precision Plant Genomics, University of Minnesota, Saint Paul, MN 55108, USA

* Correspondence: yangx@ornl.gov

Received 18 April 2019; Editorial decision 28 August 2019; Accepted 29 August 2019

Editor: David Williams, University of Wyoming, USA

Abstract

Crassulacean acid metabolism (CAM) photosynthesis is an important biological innovation enabling plant adaptation to hot and dry environments. CAM plants feature high water-use efficiency, with potential for sustainable crop production under water-limited conditions. A deep understanding of CAM-related gene function and molecular evolution of CAM plants is critical for exploiting the potential of engineering CAM into C₃ crops to enhance crop production on semi-arid or marginal agricultural lands. With the newly emerging genomics resources for multiple CAM species, progress has been made in comparative genomics studies on the molecular basis and subsequently on the evolution of CAM. Here, recent advances in CAM comparative genomics research in constitutive and facultative CAM plants are reviewed, with a focus on the analyses of DNA/protein sequences and gene expression to provide new insights into the path and driving force of CAM evolution and to identify candidate genes involved in CAM-related biological processes. Potential applications of new computational and experimental technologies (e.g. CRISPR/Cas-mediated genome-editing technology) to the comparative and evolutionary genomics research on CAM plants are offered.

Keywords: Comparative genomics, crassulacean acid metabolism, drought stress, evolution, gene function, genome editing, photosynthesis

Introduction

Crassulacean acid metabolism (CAM) is a CO₂-concentrating mechanism in plants as an adaptation to hot and dry environments. CAM plants can be divided into two main categories: (i) facultative CAM plants, which exclusively or predominantly perform C₃ photosynthesis (or C₄ photosynthesis in some instances) under well-watered conditions, with CAM induced or up-regulated in a reversible manner in response to reduced water availability caused by drought or high salinity stress, and (ii) constitutive CAM plants, in which CAM is always expressed in mature photosynthetic tissues (Winter, 2019). In comparison with other types of photosynthesis, CAM provides a unique mechanism resulting in high water-use efficiency (WUE), which can be 3–20-fold and 2–10-fold that of C₃ and C₄ plants, respectively (Borland et al., 2009). The high WUE of CAM plants results from special temporal patterns of CO₂ fixation and stomatal movement with four temporally separated phases. Phase I is the period of opening of stomata for primary atmospheric CO₂ fixation at phosphoenolpyruvate carboxylase (PEPC) at night when temperature is lower. Phase II takes places at the transition from dark to dawn, featuring a shift
from C₄ carboxylation by PEPC to ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO)-mediated CO₂ assimilation. Phase III is the period of RuBisCO refixation of CO₂ released from malate decarboxylation, with stomata closing during the daytime when temperature is higher. Phase IV occurs toward the end of the light period when stomata reopen to allow net CO₂ uptake as a result of the exhausting of malate reserves (Osmond, 1978). Furthermore, during the daytime CAM concentrates CO₂ released from malate in the vicinity of RuBisCO and consequently reduces photorespiration mediated by RuBisCO, a process that can decrease photosynthesis by up to 40% in C₃ plants (Ehleringer and Monson, 1993), resulting in a theoretical biochemical efficiency of CAM higher than that of C₃ photosynthesis (Davis et al., 2014). Inspired by the potential for enhancing the sustainable production of food and biomass on semi-arid, abandoned, or marginal agricultural land (Borland et al., 2009, 2014; Yang et al., 2015), rich genomics resources have been created for multiple CAM species, including *Phalaenopsis equestris* (Cai et al., 2014), pineapple (Ming et al., 2015), *Agave* (Gross et al., 2013; Abraham et al., 2016), *Kalanchoë* (Yang et al., 2017), *Erycina* (Heyduk et al., 2019a), *Yucca* (Heyduk et al., 2019b), *Talinum* (Brilhau et al., 2016), *Portulaca* (Christin et al., 2014), and *Sedum* (Wai et al., 2019). Leveraging these genomics resources, the CAM research community has made progress in comparative genomics research determining gene function originating in CAM plants and investigating the molecular signatures of CAM evolution. In this review, we summarize the recent progress in comparative analysis of DNA/protein sequences and diel gene expression patterns to identify candidate genes involved in CAM-related biological processes (e.g. CO₂ fixation, stomatal movement, circadian rhythm). Also, based on the new insights gained from linking genes to the differences between C₃ and CAM physiology, we update our understanding of the molecular mechanisms underlying the evolution of CAM from C₃ plants.

### The course of CAM evolution

It is widely accepted that CAM has evolved from C₃ photosynthesis (West-Eberhard et al., 2011). However, the path of C₃→CAM evolution remains unclear. From the metabolic perspective, it was recently argued that C₃→CAM evolution represents a continuum and the requirements for CAM (e.g. nocturnal CO₂ uptake, enzyme expression for CO₂ fixation, storage of organic acids, associated transport activities, carbon flow through glycolytic and gluconeogenic pathways) are all present in C₃ plants and only need enhancement for transition from C₃ to CAM (Bräutigam et al., 2017). This argument oversimplifies the complexity of the molecular mechanisms underlying the C₃→CAM transition. Recently, a genome-wide comparative genomics analysis of three constitutive CAM plant species (*Ananas comosus*, *Kalanchoë fedtschenkoi*, and *P. equestris*) in comparison with non-CAM species predicted that changes in at least 60 genes (i.e. rewiring of diel gene expression pattern in 54 genes and protein sequence mutations in another six genes) were potentially involved in CAM evolution (Yang et al., 2017). In comparison with C₃ species, changes in the expression of a large number of protein-encoding genes were also revealed in other comparative analyses of constitutive CAM plant species such as *Agave* (Abraham et al., 2016; Yin et al., 2018) and *Ananas comosus* (Ming et al., 2015). Additionally, many genes, including transcription factors, are potentially involved in CAM induction in facultative CAM plants such as *Mesembryanthemum crystallinum* (Amin et al., 2019). Besides protein-encoding genes, microRNAs (miRNAs) are likely involved in CAM gene regulation (Wai et al., 2017). Although analysis of phenotypical diversity suggests the hypothesis that there is a C₃–CAM continuum, the genotypic evidence does not support this hypothesis (Winter, 2019). Therefore, we propose two alternative models for CAM evolution: (i) a single-path model, in which the evolution of constitutive CAM was mediated by facultative CAM (Fig. 1A), or (ii) a multi-path model, in which facultative CAM and constitutive CAM evolved independently from C₃ (Fig. 1B). The path of C₃→CAM evolution is still controversial. Even though there are plant families and even single genera (e.g. *Chusa*) containing facultative CAM, weak CAM and strong CAM species, we cannot exclude the possibility that facultative CAM and constitutive CAM evolved independently from a common C₃ ancestor. Further experimental research needs to be conducted to test the hypotheses in Fig. 1 by addressing the following questions: Can any C₃ species be directly converted to facultative CAM, weak CAM, or strong CAM? Can any facultative

![Fig. 1](image-url)
CAM species be directly converted to weak CAM or strong CAM? Can any weak CAM species be directly converted to strong CAM? If yes, how many and what genes will need to be engineered through overexpression, knockout/knockdown, and/or rewired temporal expression?

It can be assumed that facultative CAM and weak CAM have a major difference in the regulation of gene expression, with the expression of CAM-related genes induced by water-deficit condition in facultative CAM plants whereas the expression of CAM-related genes is mainly under ontogenetic control in weak CAM plants. The conversion of facultative CAM to weak CAM would involve a major change in the regulatory mechanism, which is not trivial from the perspective of genetic engineering.

Besides the aforementioned changes in protein sequence and temporal/spatial gene expression required for C3-to-CAM evolution, it was recently hypothesized that anatomical modifications (e.g., an increase in leaf succulence) are the rate-limiting step in the evolution of strong CAM whereas the evolution of facultative CAM or weak CAM does not require major anatomical changes (Edwards, 2019; Heyduk et al., 2019b). It would be necessary to identify the genes responsible for the anatomical modifications required for strong CAM evolution.

The driving force of CAM evolution

The interplay of historical contingency (i.e., the outcomes of processes are unpredictable and sensitive to random events) and natural selection (i.e., a deterministic process reflecting systematic differences in the propensity of alternative genotypes to survive and reproduce, depending on their fitness in a given environment) is the driving force of evolution (Blount et al., 2018). Although historical contingency and natural selection have been demonstrated to be responsible for the outcomes of evolution in various organisms, their relative contribution to the rich diversity of life on our planet is still a hot topic of debate. This is true even within the context of CAM evolution. CAM has evolved from C3 photosynthesis many times independently from diverse plant lineages (Edwards and Ogburn, 2012), featuring phenotypical convergence in physiological and metabolic attributes, e.g., PEPC-mediated CO2 fixation at night and the inverted day/night pattern of stomatal movement (Yang et al., 2015). Since CAM occurs within various phylogenetic hierarchies, including entire orders (Edwards and Ogburn, 2012), entire families (Simpson, 2010), or entire genera (Crayn et al., 2004), we hypothesize that historical contingency, as defined above, has played a role in CAM evolution. This hypothesis is further supported by the following genomic information: (i) in comparison with non-CAM species, single-site convergent mutations were found in only two of the three distantly related CAM species studied by Yang et al. (2017) (A. comosus, K. fedtschenkoi, and P. equestris), and (ii) no significant expansion of gene families is shared among these three CAM species (data not shown). However, in comparison with a C3 species (Arabidopsis), convergent change in the diel expression patterns of more than 50 genes was found in two distantly related CAM species studied (K. fedtschenkoi and A. comosus) (Yang et al., 2017), suggesting that natural selection on gene expression has contributed to the regulatory mechanisms enabling the C3-to-CAM conversion. Caution should be taken in interpreting the results from this comparative analysis of limited number of species because it is possible that the number of genes showing convergent change in the diel expression patterns may be reduced when more plant species are added to the comparative analysis in the future. Recently, through comparative analysis of orthologous gene pairs between CAM species Agave americana and three non-CAM species, including two C3 species (Arabidopsis and Oryza sativa) and one C4 species (Zea mays), Yin et al. (2018) identified 64 genes that were shown to carry at least one positively selected site in the protein sequences, including phosphoenolpyruvate carboxylase kinase 1 (PPCK1), which plays an important role in PEPC-mediated CO2 fixation in CAM plants. Therefore, both protein sequence mutations and changes in diel gene expression patterns were under natural selection during CAM evolution. According to this information, we can assume that both historical contingency and natural selection are important factors driving CAM evolution.

Gene duplication (at the levels of the individual gene, chromosomal segment, or entire genome) is a major force for generation of evolutionary novelties leading to adaptation to the environment (Kanazawa et al., 2009). Although no recent genome-wide duplication has been detected in the A. comosus genome (Ming et al., 2015), recent whole-genome duplication events were found in several other CAM lineages, including Agave (Ming et al., 2015) and K. fedtschenkoi (Yang et al., 2017). With such massive gene duplication events functional diversification is expected to follow (Yang et al., 2006). In K. fedtschenkoi, diversification in diel expression profiles of duplicated genes were revealed in multiple CAM-related gene families, including the tonoplast aluminum-activated malate transporter, malate dehydrogenase, PEPC, and malic enzyme gene families (Yang et al., 2017). Diversification of gene expression post-duplication suggests that natural selection is operating to create neo- or sub-functionalization.

Determination of gene function required for CAM using comparative genomics research

Although multiple CAM plant genomes have been sequenced, the bona fide function of the majority of these genes remains largely unknown and/or potentially misannotated. Comparative genomics effectively predicts genomic function and provides evolutionary insights across the phenomenal diversity of living systems (Butler et al., 2017; Rogers, 2018). A rational strategy for elucidating gene function involved in CAM is to identify the changes in DNA/protein sequences and gene expression associated with the ontogenetic C3-to-CAM transition in constitutive CAM species, environmental induction/up-regulation of CAM expression in facultative CAM species, and differences between CAM and non-CAM species, as illustrated in Fig. 2. Here, several examples are presented to show how this strategy has been used to identify candidate genes related to CAM.
In general, constitutive CAM is under ontogenetic control (Winter, 2019). In A. americana ‘Marginata’, the young leaf performs C₃ photosynthesis and the mature leaf performs CAM photosynthesis (Abraham et al., 2016). Recently, Yin et al. (2018) performed comparative analysis of gene expression between mature and young leaf tissues of A. americana ‘Marginata’ and identified a co-expression cluster containing 1509 genes, which showed up-regulated expression in the mature leaves at night in comparison with mature leaves during the daytime and the young leaves during both night-time and daytime. This co-expression cluster contains several known CAM genes, including PPCK₁, and it was therefore proposed that the genes in this module are related to the temporal reprogramming of metabolism underpinning CAM (Yin et al., 2018). Also, Ping et al. (2018) showed that the expression of a PPCK gene in Phalaenopsis aphrodite subsp. formosana was positively correlated with the progression from C₃ to CAM at different developmental stages. Similarly, differential expression between non-photosynthetic (white base) and photosynthetic (green tip) tissue of the CAM plant A. comosus was revealed for genes encoding proteins (Ming et al., 2015) as well as microRNAs (Wai et al., 2017), suggesting that both protein-encoding genes and non-coding RNAs are involved in the developmental shift from C₃ to CAM in this species.

It is agreed that facultative CAM is under environmental control (Winter, 2019). Recently, Brilhaus et al. (2016) and Maleckova et al. (2019) identified candidate transcription factors potentially involved in the regulation of the C₃-to-CAM transition in the facultative CAM species Talinum triangulare, which is capable of reversible transitioning from C₃ photosynthesis to weak CAM in response to drought stress. More recently, eight transcription factors were identified as potential candidates regulating the induction of CAM by water-deficit stress in another facultative CAM species, Mesembryanthemum crystallinum (Amin et al., 2019). These comparative genomics analyses of stress-induced CAM versus C₃ photosynthesis provide insights into the course of CAM evolution. However, additional effort will be needed to separate true CAM-related genes and to establish their function, from the non-CAM genes that are concomitantly responsive to water-deficit stress.

Comparative analysis of drought-responsive genes between facultative CAM species and closely related C₃ species has the potential of narrowing down the list of candidate genes relevant to CAM in facultative CAM species. It is possible that the drought-induced genes in a C₃ species closely related to a facultative CAM species have been co-opted to induce the CAM genes in the facultative CAM species. For example, existing regulatory proteins induced by drought in the C₃ close relative may bind to promoters of the CAM genes in the facultative CAM species and bring about the induction of these genes just as they control other drought-responsive genes and pathways in the ancestral C₃ species. Thus, some of the drought-induced genes in a C₃ species may be the same genes that induce CAM in a closely related facultative CAM species. Furthermore, the candidate genes need to be tested using experimental approaches, such as loss-of-function analysis via knocking-out or knocking-down of the candidate genes in the facultative CAM species and gain-of-function analysis via overexpression of the candidate genes in the closely related C₃ species, followed by evaluation of CAM-related traits (e.g., temporal pattern of CO₂ uptake, malic acid accumulation) in the loss-of-function and gain-of-function mutant lines under drought stress. Stable genetic transformation is a bottleneck in experimental determination of gene function in plant species, for which robust transformation protocols have not been established or are difficult to establish. Therefore, it is critical to establish diverse CAM model species that are readily amenable to stable genetic transformation in order to be able to test the function of candidate CAM genes in a diverse range of CAM levels and phylogenetic origins (Hartwell et al., 2016). For example, transcriptome-sequencing analysis of facultative the CAM species Clusia pratensis (Winter and Holtum, 2014) in comparison with its closely related C₃ species C. multiflora, C. toucheensis, or C. grandiflora (Zambrano et al., 2014) under normal and drought conditions would help identify candidate genes responsible for the C₃-to-CAM transition in C. pratensis. However, plant genetic transformation systems have not been reported for C. pratensis and its closely related C₃ species yet, limiting our ability to experimentally validate computationally predicted CAM-related genes in C. pratensis. Therefore, there is an urgent need to establish efficient plant transformation systems for facultative CAM species and their closely related C₃ species to facilitate the discovery of true CAM-related genes in facultative CAM species.

According to the relative contribution of dark CO₂ fixation to the total 24 h net atmospheric CO₂ fixation, constitutive CAM plants can be divided into two categories: strong constitutive CAM plants with dark CO₂ fixation greatly exceeding CO₂ fixation in the light under well-watered condition (K. Winter, personal communication) and weak constitutive CAM plants in which dark CO₂ fixation accounts for less than ~5% of total carbon gain (Winter,
analysis of gene expression between closely related C₃ and CAM species, such as CAM species in the genus Clusia (e.g. C. alata, C. hilariana, or C. rosa) versus closely related C₃ species (e.g. C. multiflora, C. toechensis or C. grandiflora) (Zambrano et al., 2014), CAM species in the genus Erycina (e.g. E. pusilla) versus the C₃ species E. crista-galli (Heyduk et al., 2019a), and CAM species in the genus Yucca (Y. aloifolia) versus the C₃ species Y. filamentosa (Heyduk et al., 2019a), can help elucidate gene function associated with the transition from C₃ to CAM in a specific lineage while minimizing the false-positives caused by phylogenetic divergence. Recently, a comparative transcriptomic analysis of two closely related C₃ and CAM species in the genus Erycina revealed candidate genes displaying significant differences in network connectivity, which were enriched with genes involved in light sensing and abscisic acid (ABA) signaling, including lov kelch protein 2 (involved in blue light sensing through regulation of light-induced protein degradation), phytochrome B/D (involved in red/far-red light sensing), time for coffee (responsible for the amplitude of the circadian clock but not thought to be directly involved in light signaling), protein phosphatase 2A (implicated in light signaling via dephosphorylation of phototropin 2), and seed longevity 1 (encoding RSL1, which acts as a master negative regulator of the ABA signaling pathway by targeting pyrabactin resistance 1 (PYR1) and PYR–like (PYL) ABA receptors for degradation) (Heyduk et al., 2019a). Second, comparative genomics analysis of multiple (i.e. more than one) distantly related CAM species, along with non-CAM species, can help uncover the gene function associated with phenotypic convergence among CAM plants. For example, it is widely accepted that there is one CAM-specific PEPC isoform (called PEPC1 herein), which is highly abundant and responsible for the CO₂ fixation during the night-time (Abraham et al., 2016; Boxall et al., 2017). PEPC1 is sensitive to feedback inhibition by malate (Boxall et al., 2017). Through genome-wide comparative analysis, Yang et al. (2017) identified a new PEPC isoform (called PEPC2 herein) in two distantly related constitutive CAM plant species (K. fedtschenkoi and P. equestris), which share a single amino acid mutation (from a basic amino acid, R/K/H, to an acidic amino acid, D) in comparison with non-CAM plant species. This convergent single amino acid mutation appears to alleviate the feedback inhibition by malate. This PEPC2 gene was also shown to have a much higher transcript abundance during the dark period, suggesting the regulation of PEPC2 might be different from that of PEPC1 in K. fedtschenkoi and P. equestris (Yang et al., 2017). Since the PEPC2 transcript was shown to have much lower abundance compared with the PEPC1 transcript in the whole leaf tissue of Kalanchoe (Yang et al., 2017), it would be interesting to determine whether the PEPC2 transcripts are exclusively (or preferentially) expressed in other cell types of the leaf rather than the mesophyll cells. For example, PEPC is known to play important roles in the guard cells of plant leaves (Outlaw, 1990; Tarczyński and Outlaw, 1990; Outlaw et al., 2002; Lawson, 2009; O’Leary et al., 2011), so it is feasible that a second PEPC gene (PEPC2) detected in Kalanchoe leaves would be far less abundant than the CAM-specific PEPC gene (PEPC1). Future experiments, such as analysis of loss-of-function mutants and protein subcellular
localization, will be needed to elucidate the function of the Kalanchoë PEPC2 gene.

**New technologies for CAM comparative and evolutionary genomics research**

With more genomics resources being created for CAM plants, it is becoming increasingly important to draw a deeper understanding of CAM evolution and gene function using computational and experimental approaches (Liu et al., 2018). This deeper understanding can be facilitated through the use of new technologies developed over the past few years, including new computational tools and genome-editing technologies. Homology searches have been widely used in comparative genomics research. Recently, Price and Arkin (2017) developed an online tool called PaperBLAST that integrates sequence homology search and text mining to quickly find similar proteins with functional information buried in the scientific literature. Using a PaperBLAST search, more than 200 new candidate genes with potential roles in stomatal movement were identified in *K. fedtschenkoi* (Moseley et al., 2019).

Protein sequence alignments have been widely used for CAM comparative genomics research, for phylogenetic analysis of gene families, and for the identification of protein sequence mutations in CAM plants in comparison with their orthologs in non-CAM species. Studying the structural changes caused by the amino acid mutations is helpful for understanding the impact of the protein sequence mutations on CAM evolution (Yang et al., 2017). However, comparative analysis of three-dimensional protein structure is technically challenging compared with comparative analysis of one-dimensional protein sequences. To address this technical challenge, Babbitt et al. (2018) recently developed a GUI-based pipeline, called DROIDS (Detecting Relative Outlier Impacts in Dynamic Simulations), for observing the biophysical effects of amino acid mutations in proteins and successfully applied this computational pipeline to the identification of disease-related functional changes in molecular dynamics caused by genetic mutation. This same computational tool has potential for investigating the role of protein sequence mutations in CAM evolution.

RNA interference (RNAi) has great potential for loss-of-function analysis in CAM plants (Dever et al., 2015; Boxall et al., 2017). For example, it was recently reported that RNAi-mediated down-regulation of two candidate genes involved in malate decarboxylation, encoding NAD-malic enzyme and pyruvate orthophosphate dikinase, respectively, in the constitutive CAM plant *K. fedtschenkoi* dramatically altered the temporal pattern of CAM CO$_2$ fixation (Dever et al., 2015), confirming that these two genes play important roles in CAM physiology in this species. More recently, Boxall et al. (2017) demonstrated that RNAi-mediated reduction of PPCK1 transcript abundance in *K. fedtschenkoi* caused a dramatic decrease in dark phosphorylation of PEPC and consequently reduced dark period CO$_2$ fixation, along with a loss of rhythmic expression pattern of multiple circadian clock genes, providing new insight into the function of PPCK1 in CAM plants. Precise genome engineering with CRISPR–Cas systems has shown great potential for functional genomics research in C$_3$ and C$_4$ plants, which has been summarized in several recent reviews (Liu et al., 2016; Scheben and Edwards, 2018; Zhu et al., 2018; Chen et al., 2019). CRISPR/Cas-mediated gene knockout technology has several advantages over the RNAi-mediated gene knockdown approach: (i) it allows for complete shutdown of gene function; (ii) it is useful for functional characterization of intronic and intergenic regions; (iii) it has relatively low off-targeting activities; and (iv) it is feasible for characterization of non-coding RNAs functioning in the nucleus (Boettcher and McManus, 2015; Liu et al., 2016; Rani et al., 2016; Liu et al., 2017). On the other hand, in comparison with CRISPR/Cas-mediated gene knockout technology, RNAi offers two important benefits: (i) it can generate multiple mutant lines with different degrees of knockdown, facilitating the quantitative analysis of gene expression in association with corresponding phenotypes (Liu et al., 2016); and (ii) it is useful for studying the function of essential genes, for which the knockout mutations are lethal. Here, we focus on the utilization of CRISPR–Cas systems to understand CAM evolution and gene function.

(i) CRISPR/Cas-mediated knock-out can be applied to loss-of-function analysis of the candidate CAM-related genes predicted by CAM comparative and evolutionary genomics analysis (e.g. the genes involved in CO$_2$ fixation, stomatal movement/development, leaf succulence, vacuole development, and the circadian clock). Recently, we established an efficient CRISPR/Cas9 genome editing system as an amenable approach for generating gene knockout mutants in the CAM model species *K. fedtschenkoi* (Liu et al., 2019), demonstrating that CRISPR/Cas-mediated genome editing has the potential for comparative and evolutionary genomics research in CAM plants.

(ii) Functional characterization of the changes in DNA/protein sequences associated with CAM evolution can be achieved by the cytidine deaminase editor (Komor et al., 2016) and adenine base editor (Gaudelli et al., 2017) built on the CRISPR/Cas9 system. Recently, a CRISPR-guided DNA polymerase tool was engineered to continuously diversify all nucleotides within a tunable window length at user-defined loci in *Escherichia coli* (Halperin et al., 2018), which potentially provides a new strategy for investigating the function of the candidate DNA/protein sequences shared in CAM species.

(iii) To study the impact of convergent shifts in diel patterns of gene expression in CAM species in comparison with C$_3$ species, the CRISPR/Cas–mediated knock-in approach can be used to replace the promoters of CAM genes with the promoters of their orthologs in C$_3$ plants. However, it is difficult to generate CRISPR/Cas-mediated knock-in mutants in plants. Several new strategies, such as DNA replicons (Baltes et al., 2014), sequential transformation (Miki et al., 2018), and RNA transcript-templated homologous recombination (Li et al., 2019), could be tested in CAM plants to improve the efficiency of CRISPR/Cas–mediated knock-in.
(iv) Synthetic transcriptional activators and repressors, which are built on deactivated versions of Cas9 (dCas9) fused to transcriptional activation domains and repression domains, respectively, were recently used to control gene expression in non-CAM plants (Zalatan et al., 2015; Papikian et al., 2019). This dCas9-mediated gene regulation technology has great potential for testing the impact of gene expression changes on CAM evolution.

(v) Comparing and contrasting the CAM gene function revealed by CRISPR-Cas mediated approaches and RNAi could provide complementary and comprehensive insights into CAM evolution.

Conclusion and perspectives

Over the past 5 years, comparative analyses of protein sequences and gene expression data have predicted hundreds of candidate genes that are potentially associated with metabolic processes, stomatal movement, and circadian rhythms in constitutive and facultative CAM plant species. Are these candidates, or just a subset of them, really involved in CAM-related biological processes? In other words, what is the minimum number of genes that are necessary and sufficient for converting from C3 to CAM photosynthesis? To determine whether the candidate genes are necessary for CAM photosynthesis, knockout mutants for each of these candidates need to be generated, followed by evaluating the impact of the mutants on CAM physiology. It would be technically challenging to generate knockout mutants for such a large number of candidate genes using traditional experimental approaches. In the foreseeable future, high-throughput CRISPR-based knockout systems need to be established to knock out the majority of the CAM candidate genes predicted in model CAM species (e.g., K. fedtschenkoi). To determine whether a set of candidate genes are sufficient for C3-to-CAM photosynthesis transition, various combinations of these genes should be overexpressed in model C3 species (e.g., Arabidopsis, tobacco) with appropriate promoters to ensure that the diel expression patterns of the transgenes recapitulate those in CAM plants. For example, the candidate genes involved in carboxylation, decarboxylation, stomatal movement, leaf succulence, vacuole development, and circadian rhythm in constitutive CAM plants as well as abiotic stress-responsive transcription factors in facultative CAM plants are targets for engineering CAM into C3 model species (Borland et al., 2014; Ming et al., 2015; Yang et al., 2015, 2017; Bräutigam et al., 2017; Liu et al., 2018; Moseley et al., 2018, 2019; Amin et al., 2019).

Ecophysiological studies on various constitutive and facultative CAM plant species have generated several alternative hypotheses about the evolutionary path of C3-to-CAM progression. With more CAM-related genes being identified and characterized, the existing hypotheses for CAM evolution will be tested at the molecular level through comparative analysis of CAM-related genes in natural variants and mutant lines. With the rapid advance in genome and transcriptome sequencing of CAM plants, it is expected that comparative genomics research will provide new insights into the evolution of CAM, and the causal linkage between genes and CAM traits will be unraveled in the years to come. Application of new technologies and approaches may accelerate the progress in evaluating the feasibility of the two alternative models for CAM evolution: (i) the single-path model, in which the evolution of constitutive CAM was mediated by facultative CAM or (ii) the multi-path model, in which facultative CAM and constitutive CAM evolved independently from C3 photosynthesis (Fig. 1). These two models are not necessarily mutually exclusive. It is possible that both models are feasible depending on the phylogenetic position of a specific lineage and its evolutionary predispositions to follow a specific route to CAM during evolution.

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.

Amin AB, Rathnayake KN, Yim WC, Garcia TM, Wone B, Cushman JC, Wone BWM. 2019. Crassulacean acid metabolism abiotic stress-responsive transcription factors: a potential genetic engineering approach for improving crop tolerance to abiotic stress. Frontiers in Plant Science 10, 129.

Babbitt GA, Mortensen JS, Coppola EE, Adams LE, Liao JK. 2018. DROIDS 1.20: A GUI-based pipeline for GPU-accelerated comparative protein dynamics. Biophysical Journal 114, 1009–1017.

Baltes NJ, Gil-Humanes J, Cermak T, Atkins PA, Voytas DF. 2014. DNA replicons for plant genome engineering. The Plant Cell 26, 151–163.

Blount ZD, Lenski RE, Losos JB. 2018. Contingency and determinism in evolution: Replaying life’s tape. Science 362, eaam5979.

Boettcher M, McManus MT. 2015. Choosing the right tool for the job: RNAi, TALEN, or CRISPR. Molecular Cell 58, 575–585.

Borland AM, Griffiths H, Hartwell J, Smith JA. 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, 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 LV, Kněrová J, Gould PD, Hartwell J. 2017. Phosphorylation of phosphoenolpyruvate carboxylase is essential for maximal and sustained dark CO2 fixation and core circadian clock operation in the obligate crassulacean acid metabolism species Kalanchoë fedtschenkoi. The Plant Cell 29, 2519–2536.

Bräutigam A, Schlüter U, Eisenhut M, 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 AP. 2016. Reversible burst of transcriptional changes during induction of crassulacean acid metabolism in Talinum triangulare. Plant Physiology 170, 102–122.

Acknowledgements

This research is supported by the Department of Energy (DOE), Office of Science, Genomic Science Program under Award Number DE-SC0008834. The writing of this manuscript was supported by the Center for Bioenergy Innovation (CBI), a U.S. Department of Energy (DOE) Bioenergy Research Center supported by the Office of Science, Office of Biological and Environmental Research (OBER). This manuscript has been authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under Contract Number DE-AC05-00OR22725.
Butler JB, Vaillancourt RE, Potts BM, Lee DJ, King GJ, Baten A, Shepherd M, Freeman JS. 2017. Comparative genomics of Eucalyptus and Corymbia reveals low rates of genome structural rearrangement. BMC Genomics 18, 397.

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

Ceusters J, Borland AM, Taybi T, Frans M, Gods C, De Proft MP. 2014. Light quality modulates metabolic synchronization over the diel phases of crassulacean acid metabolism. Journal of Experimental Botany 65, 3705–3714.

Chen K, Wang Y, Zhang R, Zhang H, Gao C. 2019. CRISPR/Cas genome editing and precision plant breeding in agriculture. Annual Review of Plant Biology 70, 667–697.

Chin MT, Winter K, Leebens-Mack J. 2010. Light quality modulates metabolic synchronization over the diel phases of crassulacean acid metabolism. Journal of Experimental Botany 61, 1587.

Davis SC, LeBauer DS, Long SP. 2014. Light to liquid fuel: theoretical and realized energy conversion efficiency of plants using crassulacean acid metabolism (CAM) in arid conditions. Journal of Experimental Botany 65, 3471–3478.

Dever LV, 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.

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

Edwards EJ, Ogburn RM. 2012. Angiosperm responses to a low-CO₂ world: CAM and C₄ photosynthesis as parallel evolutionary trajectories. International Journal of Plant Sciences 173, 724–733.

ELeringer JR, Monson RK. 1993. Evolutionary and ecological aspects of photosynthetic pathway variation. Annual Review of Ecology and Systematics 24, 411–439.

Gaudelli NM, Komor AC, Rees HA, Packer MS, Badran AH, Bryson DI, Liu DR. 2017. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. Nature 551, 464–471.

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, 563.

Halperin SO, Tou CJ, Wong EB, Modavi C, Schaffer DV, Dueber JE. 2018. CRISPR-guided DNA polymerases enable diversification of all nucleotides in a tunable window. Nature 560, 248–252.

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.

Herppich W, Herppich M, Von Willert DJ. 1992. The irreversible C₄ to CAM shift in well-watered and salt-stressed plants of Mesembryanthemum crystallinum is under strict ontogenetic control. Botanica Acta 105, 34–40.

Heyduk K, Hwang M, Albert V, Silvera K, Lan T, Farr K, Chang TH, Chan MT, Winter K, Leebens-Mack J. 2019a. 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, Moreno-Villena JJ, Gilman IS, Christin PA, Edwards EJ. 2019b. 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.

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

Kanazawa A, Liu B, Kong F, Arase S, Abe J. 2009. Adaptive evolution involving gene duplication and insertion of a novel Tyl/copia-like retrotransposon in soybean. Journal of Molecular Evolution 69, 164–175.

Kinosita T, Doi M, Suetsugu N, Kagawa T, Wada M, Shimazaki K. 2001. Phot1 and phot2 mediate blue light regulation of stomatal opening. Nature 414, 656–660.

Komor AC, Kim YB, Packer MS, Zunis JA, Liu DR. 2016. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature 533, 440–442.

Lawson T. 2009. Guard cell photosynthesis and stomatal function. New Phytologist 181, 13–34.

Li S, Li J, He Y, Xu M, Zhang J, Du W, Zhao Y, Xia L. 2019. Precise gene replacement in rice by RNA transcript-templated homologous recombination. Nature Biotechnology, 37, 445–450.

Li 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.

Li D, Hu R, Palla KJ, Tuskan GA, Yang X. 2016. Advances and perspectives on the use of CRISPR/Cas9 systems in plant genomics research. Current Opinion in Plant Biology 30, 70–77.

Li D, Mewalal R, Hu R, Tuskan GA, Yang X. 2017. New technologies accelerate the exploration of non-coding RNAs in horticultural plants. Horticulture Research 4, 17031.

Li 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.

Maleckova E, Bridhais D, Wrobel TJ, et al. 2019. Transcript and metabolite changes during the early phase of abscisic acid-mediated induction of crassulacean acid metabolism in Tanninum triangulare. Journal of Experimental Botany 70, 6581–6598.

Miki D, Zhang W, Zeng W, Feng Z, Zhu JK. 2018. CRISPR/Cas9-mediated gene targeting in Arabidopsis using sequential transformation. Nature Communications 9, 1967.

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

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 Kalanchoe fedtschenkoi and a model C₄ photosynthesis plant Arabidopsis thaliana. Frontiers in Plant Science 9, 1757.

Moseley RC, Tuskan GA, Yang X. 2019. Comparative genomics analysis provides new insight into molecular basis of stomatal movement in Kalanchoe fedtschenkoi. Frontiers in Plant Science 10, 292.

O’Leary B, Park J, Plaxton WC. 2011. The remarkable diversity of plant PEP/CPP (phosphoenolpyruvate carboxylase): recent insights into the physiological functions and post-translational controls of non-photosynthetic PEP/CPP. The Biochemical Journal 436, 15–34.

Osmond CB. 1978. Crassulacean acid metabolism: a curiosity in context. Annual Review of Plant Physiology 29, 379–414.

Outlaw WH Jr. 1990. Kinetic properties of guard-cell phosphoenolpyruvate carboxylase. Biochimie und Physiologie der Pflanzen 166, 317–325.

Outlaw WH Jr, Du Z, Xia Meng F, Aghoram K, Riddle KA, Chollet R. 2002. Requirements for activation of the signal-transduction network that leads to regulatory phosphorylation of leaf guard-cell phosphoenolpyruvate carboxylase during fusicoccin-stimulated stomatal opening. Archives of Biochemistry and Biophysics 407, 69–71.

Papian K, Liu W, Gallego-Bartolomé J, Jacobsen SE. 2019. Site-specific manipulation of Arabidopsis loci using CRISPR-Cas9 SunTag systems. Nature Communications 10, 729.

Ping CY, Chen FC, Cheng TC, Lin HL, Lin TS, Yang WJ, Lee YL. 2018. Expression profiles of phosphoenolpyruvate carboxylase and phosphoenolpyruvate carboxylase kinase genes in Phalaenopsis, implications for regulating the performance of crassulacean acid metabolism. Frontiers in Plant Science 9, 1587.

Price MN, Arkin AP. 2017. PaperBLAST: text mining papers for information about homologs. mSystems 2, e00039-17.

Rani R, Yadav P, Barbadikar KM, Bialiyn N, Malhotra EV, Singh BK, Kumar A, Singh D. 2016. CRISPR/Cas9: a promising way to exploit genetic variation in plants. Biotechnology Letters 38, 1991–2006.
Rogers J. 2018. Adding resolution and dimensionality to comparative genomics: moving from reference genomes to clade genomics. Genome Biology 19, 115.

Scheben A, Edwards D. 2018. Towards a more predictable plant breeding pipeline with CRISPR/Cas-induced allelic series to optimize quantitative and qualitative traits. Current Opinion in Plant Biology 45, 218–225.

Simpson MG. 2010. Plant systematics, 2nd ed. San Diego, CA, USA: Elsevier Inc.

Tarczynski MC, Outlaw WH Jr. 1990. Partial characterization of guard-cell phosphoenolpyruvate carboxylase: kinetic datum collection in real time from single-cell activities. Archives of Biochemistry and Biophysics 280, 153–158.

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

Wai CM, Weise SE, Ozersky P, Mockler TC, Michael TP, VanBuren R. 2019. Time of day and network reprogramming during drought induced CAM photosynthesis in Sedum album. PLoS Genetics 15, e1008209.

West-Eberhard MJ, Smith JA, Winter K. 2011. Photosynthesis, reorganized. Science 332, 311–312.

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

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, Kalanchoe, 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.

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 Phytopologist 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, Tuskan GA, Cheng MZ. 2006. Divergence of the Dof gene families in poplar, Arabidopsis, and rice suggests multiple modes of gene evolution after duplication. Plant Physiology 142, 820–830.

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

Zalatan JG, Lee ME, Almeida R, et al. 2015. Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. Cell 160, 339–350.

Zambrano VAB, Lawson T, Olmos E, Fernández-García N, Borland AM. 2014. Leaf anatomical traits which accommodate the facultative engagement of crassulacean acid metabolism in tropical trees of the genus Clusia. Journal of Experimental Botany 65, 3513–3523.

Zhu K, Liu H, Chen X, Cheng Q, Cheng ZM. 2018. The kinome of pine-apple: catalog and insights into functions in crassulacean acid metabolism plants. BMC Plant Biology 18, 199.