A roadmap for research on crassulacean acid metabolism (CAM) to enhance sustainable food and bioenergy production in a hotter, drier world

Summary

Crassulacean acid metabolism (CAM) is a specialized mode of photosynthesis that features nocturnal CO₂ uptake, facilitates increased water-use efficiency (WUE), and enables CAM plants to inhabit water-limited environments such as semi-arid deserts or seasonally dry forests. Human population growth and global climate change now present challenges for agricultural production systems to increase food, feed, forage, fiber, and fuel production. One approach to meet these challenges is to increase reliance on CAM crops, such as Agave and Opuntia, for biomass production on semi-arid, abandoned, marginal, or degraded agricultural lands. Major research efforts are now underway to assess the productivity of CAM crop species and to harness the WUE of CAM by engineering this pathway into existing food, feed, and bioenergy crops. An improved understanding of CAM has potential for high returns on research investment. To exploit the potential of CAM crops and CAM bioengineering, it will be necessary to elucidate the evolution, genomic features, and regulatory mechanisms of CAM. Field trials and predictive models will be required to assess the productivity of CAM crops, while new synthetic biology approaches need to be developed for CAM engineering. Infrastructure will be needed for CAM model systems, field trials, mutant collections, and data management.

I. Introduction

Two of the grand challenges facing our society in the twenty-first century are: the continuing rapid expansion of the world’s human population, now at 7.2 billion, which is expected to increase by 33–71% by 2100 (Gerland et al., 2014); and the potential increase in the frequency and intensity of drought, along with decreases in soil moisture, related to global climate change (Dai, 2013; Cook et al., 2014) (Fig. 1). These two externalities could seriously impact future food and energy security while increased competition for land and water resources between urban growth and agricultural production systems will intensify demands for limited freshwater resources. Fortunately, a viable solution to these challenges exists in crassulacean acid metabolism (CAM), a specialized type of photosynthesis that results in enhanced plant water-use efficiency (WUE). With an inverted day/night pattern of stomatal closure/opening relative to the more typical C₃ and C₄ crops, the WUE of CAM plants can be six-fold higher than that of C₃ plants and three-fold higher than that of C₄ plants under comparable conditions (Borland et al., 2009). Most present-day food crops (e.g. rice (Oryza sativa L.), corn (Zea mays L.),) and bioenergy crops (e.g. poplar (Populus spp.), switchgrass (Panicum virgatum L.), sugar-cane (Saccharum spp.)) use C₃ or C₄ photosynthesis, whereas CAM crops have yet to be extensively adopted and developed.

Two strategies could be used to explore the potential of CAM for food and biomass production: the development of CAM crops as new sources for food and biomass; and the transfer of CAM machinery into existing food and biomass crops (Fig. 1). Multiple CAM species are currently used as food sources that provide fruits, vegetables, and various natural products (Supporting Information Table S1). Some CAM plants (e.g. Agave spp.) have potential as biofuel crops due to their high theoretical biomass yield (Davis et al., 2014) and low recalcitrance for biofuels conversion (Li et al., 2014). The use of CAM species or the application of engineered CAM to improve plant WUE could curtail crop losses under catastrophic episodes of heat and drought and contribute to the expansion of crop production into abandoned or semi-arid lands. Here we outline a research roadmap that identifies some important scientific questions in CAM research, and provides direction for realizing the potential of CAM for human good in terms of food, feed, fiber, and fuel production. The infrastructure needs for further developing the CAM research community are discussed.

II. Research questions

1. How did CAM evolve from a C₃ ancestor?

CAM has evolved multiple times in diverse lineages of vascular plants and is found in over 400 distinct genera across 36 families (J. A. C. Smith et al., unpublished). However, our understanding of the evolutionary history of CAM is still rudimentary. There are several reasons for this, all of which present formidable, yet surmountable, challenges.

First, there is continuing debate about how exactly to define a CAM plant (Winter et al., 2015). Numerous surveys of succulent plants have provided evidence of a clear bimodal distribution of δ¹³C: ¹²C isotope ratios, with a minimum in the frequency distribution typically observed at a δ¹³C value of c. −20‰. Species with δ¹³C values less negative than −20‰ correspond to obligate CAM plants engaged in fixing the majority of their CO₂
at night. However, δ13C values between −20‰ and −27‰ might indicate a small but significant degree of nocturnal CO2 fixation (Winter & Holtum, 2002). The ability to carry out some dark CO2 fixation while still primarily engaged in C3 photosynthesis is very likely a key intermediate step along the C3-to-CAM evolutionary trajectory. However, very little is known about the prevalence and phylogenetic distribution of this low-level CAM activity, which can only be detected by direct measurements of CO2 exchange and acidity fluctuations on living accessions under conditions conducive to the expression of CAM.

Second, many CAM-evolving groups are also spectacularly diverse. Scoring the presence or absence of CAM has been accomplished for over 1000 species of orchids (Silvera et al., 2009, 2010), yet this covers only a fraction of all orchid species (> 25 000). A study of nearly 2000 species of bromeliads represents the single largest carbon-isotope survey to date (Crain et al., 2004, 2015), corresponding to almost two-thirds of the family; by contrast, phylogenies of the Bromeliaceae have so far contained fewer than 200 taxa, including both C3 and CAM species (Givnish et al., 2011, 2014; Silvestro et al., 2014). Other CAM groups are equally daunting: a recent attempt to reconstruct CAM evolution in the genus Euphorbia (c. 2000 species) has provided valuable insight into a previously understudied family (Horn et al., 2014), but only c. 10% of the genus was sampled. On the other hand, the suborder Portulacineae of Caryophyllales (Arakaki et al., 2011; Edwards & Ogburn, 2012), which currently has a relatively more complete phylogenetic sampling, lacks an equivalently fine-grained survey of CAM capability. To understand the evolutionary dynamics of CAM origins and losses, a complete sampling of both phylogeny and phenotype is required. Thus, there is a strong need for the strategic development of specific lineages as model systems (see Section III.1) coupled with genomics research (see Section II.2) to infer the evolutionary trajectory of CAM.
2. What are the genomic features of CAM plants?

The growth and development of CAM plants are controlled by functional elements encoded in the genome. Identification of the genomic features of CAM plants will benefit greatly from the construction of the pan-genomes (Hirsch et al., 2014) of CAM species from diverse lineages that encompass obligate CAM species and species with a facultative (inducible) component of CAM. Comparative genomics approaches will allow the identification of both the core genes that are shared by different CAM species and the genes that are specific to different biochemical CAM types (Holtum et al., 2005). Three important types of functional elements are embedded in genomic sequences: protein-coding genes, noncoding RNA (ncRNA) genes, and regulatory cis-elements. Protein-coding genes express mRNAs that encode translational information for protein synthesis. The ncRNA genes, which are not translated into proteins, produce transcripts that function directly as structural, catalytic, or regulatory RNAs (Eddy, 2001). The cis-elements play important roles in regulating the expression of protein-coding genes and ncRNA genes. The neofunctionalization and subfunctionalization of at least some of the genes required for CAM are likely to have occurred through the differentiation of cis-regulatory elements that control the magnitude and patterns of gene expression (Monson, 2003; Hibberd & Covshoff, 2010). Equally important are mutations or polymorphisms within the protein-coding regions that result in modified functional domains that might have been necessary to adjust the kinetic properties of enzymes and transporters required for CAM.

Ongoing and future genome projects should identify genes recruited specifically to CAM function and their associated cis-elements through analysis of conserved noncoding sequences among co-expressed genes within species, or orthologous genes shared between different CAM species. The cis-regulatory elements identified via computational analysis should be validated using reporter genes, such as GUS (for tissue-specific expression), GFP (for cell-specific expression), and LUC (for temporal expression). Special attention should be given to promoters responsible for drought-inducible CAM expression in facultative CAM plants such as Clusia pratensis Seem. and Mesembryanthemum crystallinum L., in addition to those controlling temporal and cell-specific gene expression. The comparison of diverse CAM genomes should explain the degree of evolutionary flexibility that allowed this complex trait to emerge. For example, have the same gene orthologs been recruited to a CAM function in different species, or have different orthologs been recruited (Christin et al., 2015)? Applying a commonly agreed-upon set of criteria for identifying such genomic features will be critical for answering this question. A comparative framework for identifying genomic elements relevant to CAM is illustrated in Fig. 2.

![Fig. 2. Genomic elements of crassulaceous acid metabolism (CAM) plants in a comparative framework. The X-axis represents the pan-genome, which is divided into homolog groups including the groups (i.e. ortholog groups) shared by two or more species/genomes and the groups (i.e. paralogs) representing species-specific gene expansion families. The Y-axis represents the plant species, which can be categorized as facultative CAM (fCAM), obligate CAM (oCAM), or C3 photosynthesis (C3). C3 species are used as comparators to identify CAM-related genes. The Z-axis represents the important information relevant to the genes in each individual homolog group (e.g. gene expression, DNA polymorphism, post-translational modification, cis-regulatory elements, epigenetic modification, three-dimensional (3D) protein structure). Y, the homolog group contains gene(s) from the corresponding species; —, the homolog group does not contain gene(s) from the corresponding species. The homolog groups can be divided into six categories: (1) common, homolog groups shared by C3, fCAM, and oCAM species; (2) C3-specific, homolog groups shared only among C3 species; (3) core CAM, homolog groups shared only by fCAM and oCAM species; (4) inducible CAM, homolog groups shared only by fCAM species; (5) obligate CAM, homolog groups shared only by oCAM species; and (6) species-specific, species-specific gene expansion families.](image-url)
3. What are the molecular mechanisms regulating CAM?

CAM regulation has been studied from numerous perspectives including light–dark and circadian clock control over each 24-h cycle, as well as the developmental and abiotic stress-dependent regulation of the establishment of CAM (Cushman & Bohnert, 1999; Hartwell, 2006; Freschi & Mercier, 2012). CAM requires strict temporal control of the associated metabolism in order to prevent futile cycling between dark-period CO₂ fixation to malate and light-period malate decarboxylation. Furthermore, the signal transduction pathways that control inverse stomatal opening and closing in CAM plants must be deciphered in detail, as there is currently very little direct experimental data relating to stomatal guard cell signaling in CAM species. Also, redox control could be critical for synchronization of metabolism and transport over the diel CAM cycle, as suggested by the observation that some enzymes involved in the Calvin–Benson cycle in C₃ and C₄ plants, algae, and cyanobacteria are regulated by thioredoxins to achieve higher activities in reduced states than in oxidized states (Michel et al., 2013).

In a mature maize leaf, a C₃-to-C₄ developmental gradient exists between the leaf base (C₃) and the leaf tip (C₄) (Li et al., 2010). In Agave americana var. marginata Trel., young and mature leaves on the same plant use the C₃ and CAM photosynthetic pathways, respectively (X. Yang et al., unpublished data). It would therefore be very useful to compare the gene expression patterns between C₃ and CAM leaf tissue within the same CAM plant in order to understand the developmental regulation of CAM. To understand abiotic stress-dependent regulation of CAM, studies must be undertaken using truly facultative CAM species (e.g. M. crystallinum, C. pratensis) to identify the regulatory components required for the drought induction of CAM, and the subsequent return to C₃ under well-watered conditions.

Achieving a comprehensive understanding of CAM-associated regulatory mechanisms will benefit greatly from the application of functional genomics approaches that include both computational predictions based on omics data and experimental characterization using molecular and genetics tools (Fig. 3). Dissecting the complexity of CAM regulation will be facilitated by the construction of protein–protein interaction and gene regulatory networks (GRNs), which are the collection of interactions between transcription factors and their target genes. The integration of complete genome sequences with RNA-seq and chromatin immunoprecipitation sequencing (ChIP-seq) experiments offers an exciting platform to dissect and model GRNs using computational approaches such as Bayesian inference, Boolean modeling, linear and nonlinear regression methods, Granger causality-based inference, and cross-correlation analysis (Wallach et al., 2010; Marbach et al., 2012; Middleton et al., 2012; Krouk et al., 2013; Moghaddam & Van den Ende, 2013; Tam et al., 2013).

Despite recent and continuing advances with a range of omics projects ongoing in CAM species (Borland et al., 2014), a key area that remains lacking is the study of post-translational modifications associated with CAM regulation. Reversible phosphorylation of phosphoenolpyruvate carboxylase (PPC) by its specific, circadian clock-controlled protein kinase, PPCK, is one of the few CAM regulatory steps understood in any detail (Hartwell et al., 1999; Taybi et al., 2000; Boxall et al., 2005; Dever et al., 2015). In the coming years, research programs focused on achieving a comprehensive understanding of CAM regulation must determine the role of regulatory processes such as post-transcriptional modification of mRNA stability or translatability and post-translational modulation of protein activity (e.g. phosphorylation/dephosphorylation, ubiquitination, glycosylation).

4. How might CAM be engineered into C₃ or C₄ plants?

As discussed in Section I, CAM-engineering is a viable strategy to improve WUE in existing non-CAM crops for food and biomass production in dryland areas. In principle, CAM-into-C₃/C₄ engineering is realistic because: (1) CAM has evolved from diverse C₃ species via convergent or parallel evolution (see Section II.1); (2) the existence of facultative CAM species, in which CAM can be induced from C₃ or C₄ by drought or salt stress (see Section III.1), suggests that no incompatibilities exist between CAM and C₃/C₄ at the organismal level; and (3) CAM is a single-cell carbon concentrating mechanism that does not require differentiated mesophyll and bundle sheath cell types, each with their own specialized metabolic adaptations. Ideally the C₃ target species for CAM engineering should meet the following criteria: (1) a genome that has been fully sequenced and well annotated; (2) an easily transformed species with a well-established stable transformation protocol; (3) a large impact on food or bioenergy production; and (4) a crop that is currently not well suited for production on dryland. Poplar and rice are examples of such candidate target C₃ crops for CAM engineering, representing bioenergy and food crops, respectively. If the CAM-into-C₃ engineering effort is successful, the potential of CAM-into-C₃ engineering can be investigated as a means to further enhance the WUE of major C₄ crops such as corn and sorghum (Sorghum bicolor (L.) Moench). Engineering of CAM into C₃ crops will require a temporal reprogramming (e.g. diel-cycle shift) of the expression of genes shared between the C₃ and CAM pathways, transferring CAM-specific genes, possibly modifying endogenous genes (i.e. silencing or knockout) in the host, engineering of leaf anatomical traits (e.g. succulence, cell size, intercellular air space), and most likely an inducible system that would initiate CAM when desired (e.g. under drought stress). Currently, the exact number of genes needed to introduce CAM into a C₃ species remains unclear; however, multiple CAM-related genes will need to be manipulated in a modular manner, including: (1) a carboxylation module for CO₂ fixation and nocturnal accumulation of malic acid in the vacuole; (2) a decarboxylation module for release of CO₂ from malate; (3) a stomatal control module for nocturnal stomatal opening and stomatal closure during the daytime; and (4) an anatomical module for increasing leaf succulence (Borland et al., 2014). Furthermore, these four CAM modules need to be integrated to establish CAM as an efficient system in C₃ plants. Hypothesized minimal gene sets for the carboxylation and decarboxylation CAM modules are listed in Table S2. Elucidation of the equivalent gene lists for stomatal control and succulence must await detailed studies of these processes in CAM species, as the required data are currently available.
Fig. 3 An integrative functional genomics approach for crassulacean acid metabolism (CAM) plants. (a) Omics data (i.e. genomics, transcriptomics, proteomics, metabolomics) are generated for CAM plants, funnel image © 2013 TimArbaev, Opuntia image © andylin, Agave image © SSSCC, pineapple image © 2012 julichka. (b) The omics data are analyzed to predict CAM-related genes and putative gene (regulatory) networks; and (c) various experimental approaches are used to characterize the CAM genes predicted by approaches in (b). \([\text{CH}_2\text{O}]_n\), carbohydrates; OAA, oxaloacetate; PEP, phosphoenolpyruvate; RuBP, ribulose 1,5-bisphosphate; triose-P, triose phosphate; TF, transcription factor; Y1H, yeast one-hybrid; Y2H, yeast two-hybrid; BiFC, bimolecular fluorescence complementation.
lacking. Potential crosstalk between the CAM modules needs to be considered. For example, a reduction in the partial pressure of CO$_2$ inside the leaf (\(p_i\)) due to PPC activity in the dark following the successful introduction of a functioning carboxylation module could result in nocturnal stomatal opening whereas an increase in \(p_i\) due to CO$_2$ released from malate by the decarboxylation module during the daytime could induce stomatal closure. Thus, insertion of additional genes for a stomatal control module might be obviated by the successful installation of diel malate turnover in the leaf ground mesophyll. CAM engineering is well beyond the capacity of traditional plant biotechnology that is limited to transferring and controlling only a few genes. Synthetic biology offers the potential to address the challenge of CAM engineering via new concepts and toolboxes (DePaoli et al., 2014). The application of synthetic biology to CAM engineering involves five steps: (1) establishment of a parts library (e.g. genes, promoters, terminators, genetic insulators); (2) circuit design; (3) assembly of multi-gene constructs; (4) transfer (i.e. in planta gene stacking); and (5) evaluation of engineered plants (Fig. 4). Multiple iterations of steps 2–5 will be required to achieve optimized performance of engineered CAM. The parts information needs to be derived from knowledge of the core CAM genes and regulatory mechanisms (see Sections II.2–II.3) and informed by phylogenetic analyses (see Section II.1) that identify independently evolving modules of traits. To prevent influence by inappropriate or competing signals emanating from their surrounding genomic environment, it is necessary to protect transgenes with genetic insulators, which are a class of DNA sequence elements with the ability to block the action of a distal enhancer on a promoter or act as barriers to prevent the advance of nearby condensed chromatin that might otherwise silence expression (West et al., 2002; She et al., 2010). Finding suitable insulators for target C$_3$ species is an important task for CAM-into-C$_3$ engineering.

To streamline the downstream processes and facilitate collaboration in the CAM research community, a standard for the construction of the gene parts and circuits should be established. Circuit design can adopt a modular approach wherein the parts are first assembled into carboxylation, decarboxylation, stomatal control, and anatomical modules; these modules are then connected into a CAM system. Various methods have been developed for assembling multi-gene constructs (DePaoli et al., 2014); however, none of them allow flexible, clean, and efficient assembly of parts from a single universal library. New high-throughput methods for in vitro assembly of multi-gene constructs, such as those described recently for mammalian systems (Guye et al., 2013; Torella et al., 2014), are needed. A significant challenge for transferring assembled CAM gene modules will be to insert these large multi-gene constructs into the plant genome while maintaining the structural and functional stability of the modules. Methodologies that are site-specific, functional for multiple rounds of targeted in vivo insertions, and compatible with multiple methods of plant transformation still need to be developed. Progress on both in vitro assembly of DNA parts and in planta gene stacking for iterative insertion of marker-free DNA modules is underway and merits further development (H. C. DePaoli et al., unpublished). The transgenic plants generated during CAM engineering should be evaluated using omics approaches (e.g. transcriptomics, proteomics, metabolomics, phenomics). These
omics data can be used for system dynamics modeling (Borland & Yang, 2013; Owen & Griffiths, 2013) and diel flux balance analysis (Cheung et al., 2014) to inform metabolic and regulatory refinements that will improve the performance of engineered CAM. Moreover, the diel flux balance model could aid the optimal design of the carboxylation, decarboxylation, and stomatal control modules before they are engineered into a C3 species.

5. How can sustainable CAM crop production systems be established?

CAM crops such as Agave, Opuntia, and pineapple (Ananas comosus (L.) Merr.) have potential as biofuel and food crops on abandoned, marginal, and degraded land in light of published reports on their high productivities (Table S1). Agave and Opuntia species have been used traditionally in a wide range of foods, beverages, food products, forage, fodder, and also dietary supplements, pharmaceuticals, and cosmetics. The many industrial uses of Agave fibers include cordage, textiles, construction materials, and solid fuels (Cushman et al., 2015). However, more extensive field trials are required to provide data for the food, feed, fiber, and bioenergy uses of CAM species to integrate into a framework that considers sustainable yields given externalities of land availability, management inputs, economics, and market demand (Davis et al., 2011; Nunez et al., 2011; Yan et al., 2011; Lewis et al., 2015). Davis et al. (2011) estimated that substantial abandoned agricultural land exists globally and suggested that this land could be reclaimed and repurposed for bioenergy production. Furthermore, in areas where CAM is a viable option, biomass production should be evaluated against sustainability metrics that include water quality, water use, fertilizer inputs, potential herbicide and pesticide applications, and biodiversity (McBride et al., 2011). Such information would be useful for resource assessment models and for evaluating environmental consequences of CAM plantations.

System-level analysis of agricultural production has been applied to many agricultural production settings and should be similarly developed for CAM plants (Davis et al., 2015). To date, there has been only one detailed life-cycle analysis (LCA) that addresses a CAM crop (Yan et al., 2011). That study concludes that an Agave-based bioenergy system would have greater energy returns per unit of energy input than a bioenergy system based on maize. Moreover, greenhouse gas emissions per unit of energy produced from Agave would be much lower than those from a maize grain system (Yan et al., 2011). Physiological models of CAM also require the development of tools comparable to those used for C3 and C4 crops (Davis et al., 2015). To assess the relative benefits of CAM cropping systems, studies should undertake comparative physiological modeling and LCA for obligate CAM crops and other bioenergy crops including Jatropha (C3), poplar (C3), willow (Salix spp.; C3), Miscanthus (C4), sugarcane (C4), and switchgrass (C4).

In addition, productivity models could be valuable tools for identifying management scenarios suitable for sustainable crop production. Such models exist for only a few bioenergy crops (Nair et al., 2012) and have proven useful in efforts to tailor crops and cropping systems to various environments (Miguez et al., 2012).

The most widely used model for CAM plants is the environmental productivity index (EPI), which estimates potential yield based on temperature, soil water, and solar radiation (Nobel, 1984; Garcia de Cortazar & Nobel, 1990). Owen & Griffiths (2014) have developed a geospatial model based on the EPI approach to predict bioethanol yield potential for Agave and Opuntia species in Australia, and used that model to predict crop production on low-grade and marginal lands under current and future climate conditions (Figs 5, S1). Simulations highlight that the same WUE features of the CAM pathway which distinguish it from C3 and C4 bioenergy candidates also offer resilience to predicted climate change.

Although the EPI has proven useful, it lacks mechanistic details. Owen & Griffiths (2013) have thus developed a systems dynamics model of CAM that integrates biochemical and physiological constraints to predict leaf-level gas exchange and titratable acidity fluctuations. While this model does not allow physiological predictions at the canopy scale, key regulatory components of this model could be manipulated to simulate CAM expression across contrasting succulent life forms. The model was able to identify parameters that limit carbon uptake over the diel cycle and thus may prove useful as a tool to help target synthetic biology approaches to improve crop production (see Section II.4). Opportunities exist to incorporate gene regulatory and metabolic networks into this model and to link CAM expression to whole-plant traits such as net assimilation rate, relative growth rate, and the allocation and partitioning of carbon among plant components.

III. Infrastructure for the CAM community

1. Model systems for CAM research

Model systems are key elements of integrative research programs. The development of three types of model systems is suggested: phylogenetic lineages that include both C3 and CAM (and potentially also C4) species for studying CAM evolution; CAM species with a small genome, short-life cycle, and well-established genetic transformation system for functional genomics research; and CAM species with potential for food, feed, and biomass production as model crops.

Potential model lineages with species showing C3, CAM and C4 Ideally, model lineages would include species from multiple CAM origins and variations in the operation of CAM. The neotropical genus Clusia is the only genus of woody eudicotyledonous trees reported to use CAM (Lützge, 2006, 2008). This genus includes obligate CAM and C3 species as well as species that show facultative CAM and reversible shifts between C3 and CAM (Winter & Holtum, 2014). A unique opportunity exists with Portulaca, a lineage that includes the only known examples of C4 species in which CAM can be induced by drought stress (Koch & Kennedy, 1980, 1982; Christin et al., 2014). Understanding the molecular, anatomical, and metabolic mechanisms that allow for the co-existence of C4 and CAM could facilitate engineering of both pathways into a single plant. Another good model lineage is...
Erycina, a group of orchids containing members that perform C3 or CAM photosynthesis (Silvera et al., 2010).

Model species for functional genomics Two Kalanchoë species, K. laxiflora Baker and K. fedtschenkoi R.-Hamet & Perrier, have been established as obligate CAM model systems due to their relatively small genome sizes, short life cycle, and amenability to stable transformation (Aida & Shibata, 1996; Garces et al., 2007; Garcia-Sogo et al., 2010; Dever et al., 2015). Their genomes are currently being sequenced, assembled, and annotated (Table S3). In addition, the common ice plant (M. crystallinum) is a well-studied, facultative CAM model, in which CAM can be induced by salinity or water-deficit stress (Winter & Holtum, 2005, 2007, 2014). M. crystallinum has played a seminal role in our understanding of the function and subcellular localization of enzymes involved in CAM function (Holtum & Winter, 1982; Winter et al., 1982), as well as defining many of the corresponding genes for these enzymes (Cushman et al., 2008b) and intracellular transporters (Kore-eda et al., 2013). The ice plant transcriptome and genome are currently being sequenced, assembled, and annotated (Table S3). Another emerging model is Sedum telephium L. (= Hylotelephium telephium subsp. telephium L.), a C3–CAM intermediate (Groenhof et al., 1990; Borland, 1996). Genomic resources for the genus Sedum have been developed (Chao et al., 2010; Gao et al., 2013), and sequencing of the genome of S. telephium is in progress (Table S3).

Model CAM crops Currently, only a limited number of CAM crops (e.g. Agave spp., Manfreda spp., Polianthes spp., Prochnyanthes mexicana (Zucc.) Rose, Aloe vera (L.) Burm.f., A. comosus, Hylocereus spp., Opuntia ficus-indica (L.) Mill., Stenocereus spp., Vanilla planifolia Jacks. ex Andrews) have been used for the production of food, bioenergy, fiber, and animal feed (Table S1). The production scale of existing CAM crop is currently much smaller than that of major C3 or C4 crops, although the benefits of these CAM plants as cash crops are of considerable importance to the economies of many nations in the tropics and subtropics. Due to the potential increase in the frequency and
intensity of drought (Fig. 1), CAM crops could play an increasing role in meeting our future needs for food and bioenergy. Therefore, we propose more widespread planting of CAM crops, with an initial focus on three major CAM crops as models: *Agave*, *Opuntia*, and pineapple. *Agave* spp. are economically important CAM crop species, holding great potential for production of biofuel, fiber, food, and animal feed in water-limited areas (Li et al., 2014; Nava-Cruz et al., 2014). *Agave* was recently added to the list of potential dedicated biomass crops in the United States (US DOE, 2014). Given the importance of *A. tequilana* F.A.C. Weber for commercial alcohol production and as a representative for all agaves, its genome is currently being sequenced, assembled, and annotated (Table S3).

*Opuntia* spp. (e.g. *O. ficus-indica*) have been introduced as forage and fodder crops in many semi-arid regions of the world (Russell & Felker, 1987; Nobel, 1994; Le Houérou, 1996). The young cladodes and fruits are not only consumed as food directly or as diverse processed food items, but are also used in a wide range of other products such as sweeteners, food coloring, dietary supplements, cosmetics, and medicines (Feugang et al., 2006; Mösshammer et al., 2006). Large-scale transcriptome and genome sequencing of this CAM species is in progress (Table S3).

Pineapple, the third most important tropical fruit after banana and citrus, is cultivated in over 80 countries in tropical and subtropical regions worldwide. In *vitro* plantlets of pineapple perform C₄ photosynthesis, while adult plants perform CAM photosynthesis constitutively (Freschi et al., 2010). Both its genome and transcriptome have been sequenced (Table S3).

While the development of sustainable production systems is focused on the earlier three model CAM crops, the potential of other CAM crops for food production should be exploited. In addition to the crops listed in Table S1, it is necessary to evaluate the food quality and yield of other CAM species in order to develop new crops for the sustainable production of food and other high-value products.

2. Field trials

Establishment of replicated field trials is critical for quantifying yields under contrasting environmental conditions. Such efforts would provide data essential to the development of empirical models that use environmental conditions as inputs and give probabilistic estimates of yield (unlike EPI) for CAM species. Such models, based on field observations, could be used to: (1) validate existing EPI-based models; (2) identify locations where sustainable and profitable yields are possible; and (3) more accurately simulate yield in a hotter, drier world under projected Intergovernmental Panel on Climate Change (IPCC) climate scenarios. Field trials for *Agave* are being conducted in Australia (Holtum & Chambers, 2010) and the United States (Davis, 2013), and should be replicated in Mexico with local species, where >200 varieties have been developed for either fiber or alcohol production (Colunga-GarcíaMarín & Zizumbo-Villarreal, 2007). A replicated field study by the Nevada Agricultural Experiment Station is currently underway to evaluate the productivity of three *Opuntia* species.

Beyond field trials, a network of common gardens for each major CAM crop should be established, with collections composed of multiple genotypes from natural populations and planted in clonal replicates in 3–4 alternate environmental conditions (e.g. soil extremes, temperature minima and maxima, water availability). Omics data (e.g. genomics, transcriptomics, metabolomics, phenomics) can be collected for individual plants in the common garden, which will serve as a foundation for unraveling the association between genomic elements and trait phenotypes. Results from such studies would be very useful to inform application of genomics for sustainable improvement in CAM crops.

3. Genetic mutant collections

Mutant collections should be created for model CAM species to facilitate functional genomics research noted earlier. Such a collection has been created for *M. crystallinum* using fast-neutron bombardment, which generates small deletions or rearrangements within the genome, and has been used for the isolation and characterization of CAM-deficient mutants (Cushman et al., 2008a). Targeted loss-of-function *Kalanchoë* mutants generated by RNAi-mediated gene silencing have been created for a wide selection of candidate genes with functions in CAM (Dever et al., 2015). Whole-genome sequencing of the genotypes within a network of common gardens (see Section III.2) can also help to discover naturally occurring loss- or gain-of-function mutants. Moreover, loss- or gain-of-function mutants can also be generated using emerging genome-editing technologies (e.g. CRISPR/Cas9) (Voytas & Gao, 2014).

4. Data management and analyses

A data management and computational platform for the integration of large data sets to support predictive biology of complex systems is necessary to advance CAM research. The US Department of Energy Systems Biology Knowledgebase (KBase, http://kbase.us) provides a computational environment that supports private and shared research, and is scalable to larger research communities, such that it could be leveraged by the CAM community to provide a uniform, narrative-based, computational platform and reproducible analytical workflows. The centralized cloud-based platform would be complemented with locally networked computational resources including a shared LIMS system, data archiving and retrieval systems, local QA/QC routines, and analytical workflows. A conceptual design of the fundamentals needed for development of a CAM computational platform based on KBase is illustrated in Fig. S2.

IV. Conclusions

The grand challenges caused by ever-increasing human population and predicted global warming will require scientific innovations to guarantee a secure and sustainable supply of food, feed, fiber, and fuel. As a proven mechanism for increasing WUE in plants, CAM offers great potential for enhancing the sustainable production of food and biomass on semi-arid, abandoned, or marginal agricultural lands. Thus, CAM research is poised to become a prominent research area in the plant sciences. The important research
questions and infrastructure needs discussed earlier could serve as a reference to prioritize the efforts of the CAM research community. These critical needs and future opportunities are summarized as a research roadmap for the CAM community (Fig. 6). The study of CAM evolution (Fig. 6a) will help elucidate whether the same gene orthologs have been recruited to a CAM function in different species, providing guidance for CAM gene discovery using a systems biology approach (Fig. 6b). For example, if the same gene orthologs have been recruited to a CAM function, the discovery of essential CAM genes and cis-regulatory changes required for CAM can be expedited through comparative analysis of omics data obtained from multiple diverse CAM species with C3 species as comparators (Figs 2, 3, 6c). This comparative analysis could identify four types of CAM genes: (1) CAM genes that have functionally equivalent C3 gene orthologs without significant differences in developmental, temporal, or stress-responsive expression patterns between CAM and C3 species; (2) CAM genes that have functionally equivalent C3 gene orthologs with significant differences in developmental, temporal, or stress-responsive expression patterns between CAM and C3 species; (3) CAM genes that have orthologs in C3 species but have gained new function; and (4) CAM-specific genes that have no orthologs in C3 species. Knowledge about these four types of CAM-related genes could inform the best strategy for CAM-into-C3 engineering (Figs 4, 6d) to enable enhanced food and bioenergy production on drylands using existing C3 crops (Fig. 6e). Specifically, CAM engineering will likely require the transfer of the type 2, 3 and 4 CAM genes described earlier, along with the cis-regulatory changes required to ensure CAM-like gene expression, into C3 species, and their integration with the C3 gene orthologs of the type 1 CAM genes, to form a complete CAM system. However, reiterative rounds of engineering the introduced genes will likely be necessary
to optimize the performance of the CAM system, particularly in the case of an inducible CAM pathway analogous to those present in facultative CAM species. Furthermore, the silencing of endogenous C3 gene orthologs of the type 2 and 3 CAM genes might be useful to avoid potential conflicts between CAM and C4 in engineered plants; however, the existence of facultative CAM species suggests that such conflicts are unlikely or can be overcome readily. In addition, deep understanding of the molecular basis of adaptive evolution of CAM plants could provide knowledge for informing genetic improvement in CAM crops for food, feed, and bioenergy production (Fig. 6f). For example, comparative genomics analysis of various Agave species in the CAM germplasm collection (Fig. 6g) could provide molecular information for increasing cold tolerance in cold-sensitive Agave crop species such as A. tequilana. Similarly, comparative genomics analysis of various pineapple and cactus species in the CAM germplasm collection (Fig. 6g) could provide molecular information for genetic improvement of food and feed quality and biomass yield in these two CAM crop species through genomic selection and breeding.

Acknowledgements

This review is based on work supported by the Department of Energy (DOE), Office of Science, Genomic Science Program under award number DE-SC0008834, and the National Science Foundation under award number DEB-1252901. The contents of this review are solely the responsibility of the authors and do not necessarily represent the official views of the DOE. Additional support is provided by grants from the Nevada Agricultural Experiment Station (project numbers NAES 000377 and 000380), and the UK Biotechnology and Biological Sciences Research Council (grant no. BB/F009313/1). The authors wish to thank Mary Ann Cushman for critical review and clarifying comments on the manuscript and Lori Kunder (Kunder Design Studio) for assistance with figure preparation. Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the US DOE under Contract Number DE-AC05-00OR22725.

Xiaohan Yang1, John C. Cushman2, Anne M. Borland1,3, Erika J. Edwards4, Stan D. Wullschleger5, Gerald A. Tuskan1, Nick A. Owen6, Howard Griffths6, J. Andrew C. Smith7, Henrique C. De Paoli1, David J. Weston1, Robert Cottingham1, James Hartwell6, Sarah C. Davis9, Katia Silvera10, Ray Ming11,12, Karen Schlauch13, Paul Abraham14, J. Ryan Stewart15, Hao-Bo Guo16, Rebecca Albion17, Jungmin Ha18, Sung Don Lim2, Bernard W. M. Wone2, Won Cheol Yim2, Travis Garcia2, Jesse A. Mayer2, Juli Peterit2, Sujithkumar S. Nair2, Erin Casey3, Robert L. Hettich4, Jerry Jenkins2,23, Jeremy Schmutz23,24, and Joseph A. M. Holtum25

1Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6407, USA;
2Department of Biochemistry and Molecular Biology, University of Nevada, MS330, Reno, NV 89557-0330, USA;
3School of Biology, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK;
4Department of Ecology and Evolutionary Biology, Brown University, Box G-W, Providence, RI 02912, USA;
5Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6301, USA;
6Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge, CB2 3EA, UK;
7Department of Plant Sciences, University of Oxford, South Parks Road, Oxford, OX1 3RB, UK;
8Department of Plant Sciences, Institute of Integrative Biology, University of Liverpool, Liverpool, L69 7ZB, UK;
9Voinovich School of Leadership and Public Affairs and Department of Environmental and Plant Biology, Ohio University, Athens, OH 45701, USA;
10Smithsonian Tropical Research Institute, PO Box 0843-03092, Balboa, Ancon, Republic of Panama;
11Department of Plant Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA;
12FAFU and UIUC-SIB Joint Center for Genomics and Biotechnology, Fujian Agriculture and Forestry University, Fuzhou 350002, China;
13Nevada Center for Bioinformatics, University of Nevada, MS330, Reno, NV 89557-0330, USA;
14Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA;
15Department of Plant and Wildlife Sciences, Brigham Young University, 4105 Life Sciences Building, Provo, UT 84602, USA;
16Department of Biochemistry and Cellular and Molecular Biology, University of Tennessee, Knoxville, TN 37996, USA;
17Department of PBS, Faculty of Engineering Technology, TC Bioengineering Technology, KU Leuven, Campus Geel, Kleinhovestraat 4, B-2440, Geel, Belgium;
18Key Laboratory of Forest Genetics and Breeding, Research Institute of Subtropical Forestry, Chinese Academy of Forestry, Fuyang, 311400, China;
19Centro de Investigación Científica de Yucatán, Calle 43 No. 130, Colonia Chuburná de Hidalgo, CP 97200, Mérida, México;
20Department of Botany, University of São Paulo, São Paulo 05508-090, Brazil;
21Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, ON M5S3B2, Canada;
22Department of Food Research, School of Chemistry, Universidad Autónoma de Coahuila, Saltillo, México;
23HudsonAlpha Institute for Biotechnology, 601 Genome Way, Huntsville, AL 35801, USA;
24US Department of Energy Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA;
25College of Marine and Environmental Sciences, James Cook University, Townsville, 4811, QLD Australia (*Author for correspondence: tel +1 865 241 6895; email yangx@ornl.gov)
References


