Title

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

Journal

New Phytologist, 207(3)

ISSN

0028-646X

Authors

Yang, Xiaohan

Cushman, John C

Borland, Anne M

et al.

Publication Date

2015-08-01

DOI

10.1111/nph.13393

Peer reviewed

Viewpoints

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 CO_2 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 C_3 and C_4 crops, the WUE of CAM plants can be six-fold higher than that of C_3 plants and three-fold higher than that of C_4 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.), sugarcane (*Saccharum* spp.)) use C_3 or C_4 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 C_3 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 $^{13}C:{}^{12}C$ isotope ratios, with a minimum in the frequency distribution typically observed at a $\delta^{13}C$ value of *c.* $-20‰$. Species with $\delta^{13}C$ values less negative than $-20‰$ correspond to obligate CAM plants engaged in fixing the majority of their CO_2

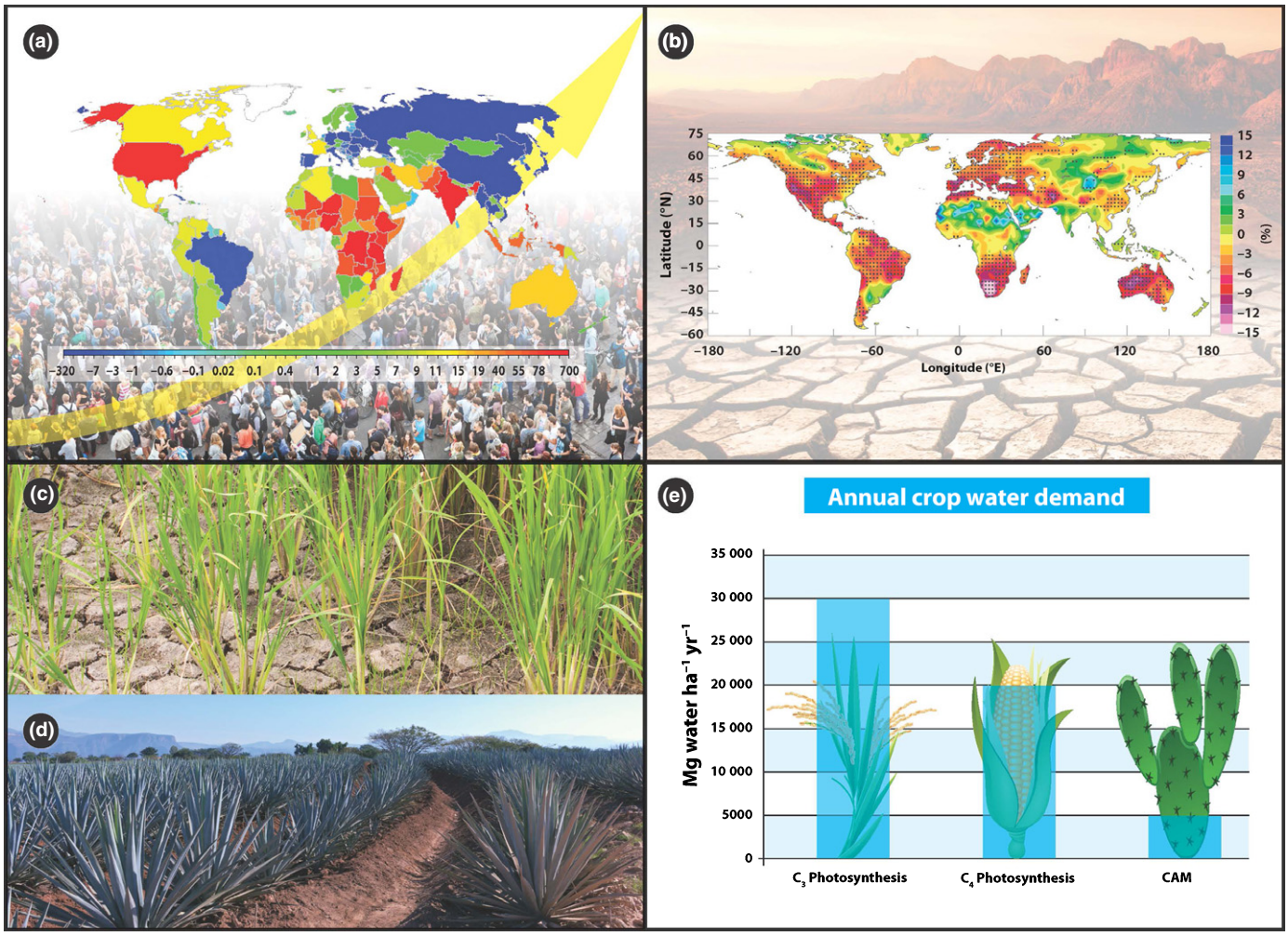

Fig. 1 Challenges and crassulacean acid metabolism (CAM) solution. (a) Expected geographical population changes in millions from 2013 to 2100 (A. E. Raftery, University of Washington, 2013, based on methodology described in Raftery *et al.* (2012, 2013, 2014)). (b) Percentage changes from 1980–1999 to 2080–2099 in the multi-model ensemble mean soil-moisture content in the top 10 cm layer (Dai, 2013). (c) The production of C_3 crops is negatively impacted by drought stress, rice image © 2013 Techin24. (d) Water-use efficient CAM crop (e.g. *Agave*) for biomass production on marginal land, *Agave* image © 2014 jferrer; and (e) CAM plants demand much less water than C_3 or C_4 plants (Borland *et al.*, 2009); thus, CAM crops and non-CAM crops engineered with CAM provide an excellent water-saving strategy to address the grand challenges caused by future increases in both world population and drought stress, rice image © 2012 KhartesevaTetiana, maize image © 2011 tilo, cactus image © 2008 andylin.

at night. However, $\delta^{13}C$ values between −20‰ and −27‰ might indicate a small but significant degree of nocturnal CO_2 fixation (Winter & Holtum, 2002). The ability to carry out some dark CO_2 fixation while still primarily engaged in C_3 photosynthesis is very likely a key intermediate step along the C_3-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 CO_2 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 (Crayn *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 C_3 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.

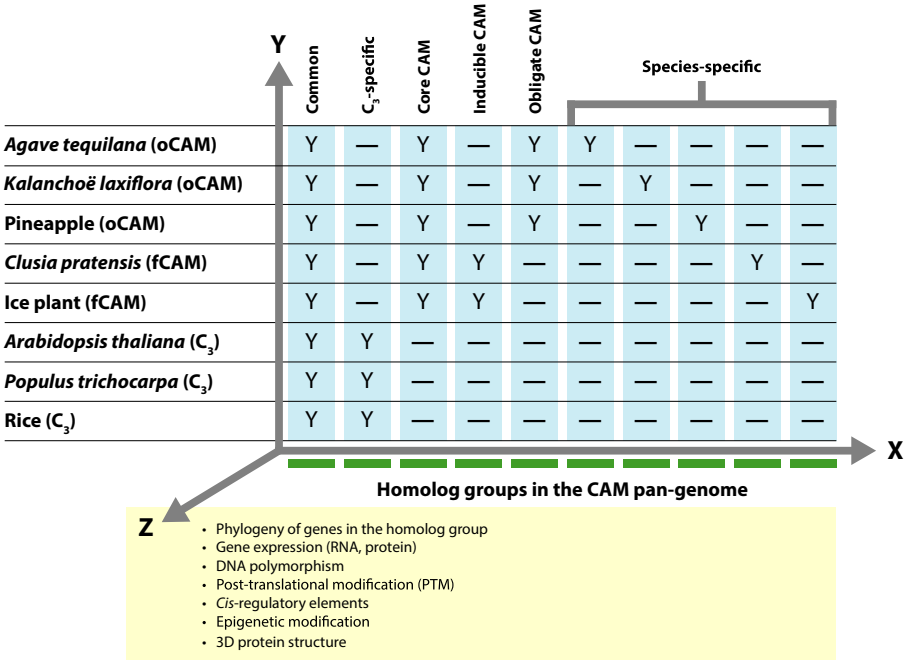

Fig. 2 Genomic elements of crassulacean acid metabolism (CAM) plants in a comparative framework. The *X*-axis represents the pan-genome, which is divided into homolog groups including the groups (i.e. ortholog groups) shared by two or more species/genomes and the groups (i.e. paralogs) representing species-specific gene expansion families. The *Y*-axis represents the plant species, which can be categorized as facultative CAM (fCAM), obligate CAM (oCAM), or C_3 photosynthesis (C_3). C_3 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 C_3, fCAM, and oCAM species; (2) C_3-specific, homolog groups shared only among C_3 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.

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 CO_2 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 C_3 and C_4 plants, algae, and cyanobacteria are regulated by thioredoxins to achieve higher activities in reduced states than in oxidized states (Michelet *et al.*, 2013).

In a mature maize leaf, a C_3-to-C_4 developmental gradient exists between the leaf base (C_3) and the leaf tip (C_4) (Li *et al.*, 2010). In *Agave americana* var. *marginata* Trel., young and mature leaves on the same plant use the C_3 and CAM photosynthetic pathways, respectively (X. Yang *et al.*, unpublished data). It would therefore be very useful to compare the gene expression patterns between C_3 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 C_3 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 phospho*enol*pyruvate 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 C_3 or C_4 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-C_3/C_4 engineering is realistic because: (1) CAM has evolved from diverse C_3 species via convergent or parallel evolution (see Section II.1); (2) the existence of facultative CAM species, in which CAM can be induced from C_3 or C_4 by drought or salt stress (see Section III.1), suggests that no incompatibilities exist between CAM and C_3/C_4 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 C_3 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 C_3 crops for CAM engineering, representing bioenergy and food crops, respectively. If the CAM-into-C_3 engineering effort is successful, the potential of CAM-into-C_4 engineering can be investigated as a means to further enhance the WUE of major C_4 crops such as corn and sorghum (*Sorghum bicolor* (L.) Moench). Engineering of CAM into C_3 crops will require a temporal reprogramming (e.g. diel-cycle shift) of the expression of genes shared between the C_3 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 C_3 species remains unclear; however, multiple CAM-related genes will need to be manipulated in a modular manner, including: (1) a carboxylation module for CO_2 fixation and nocturnal accumulation of malic acid in the vacuole; (2) a decarboxylation module for release of CO_2 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 C_3 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

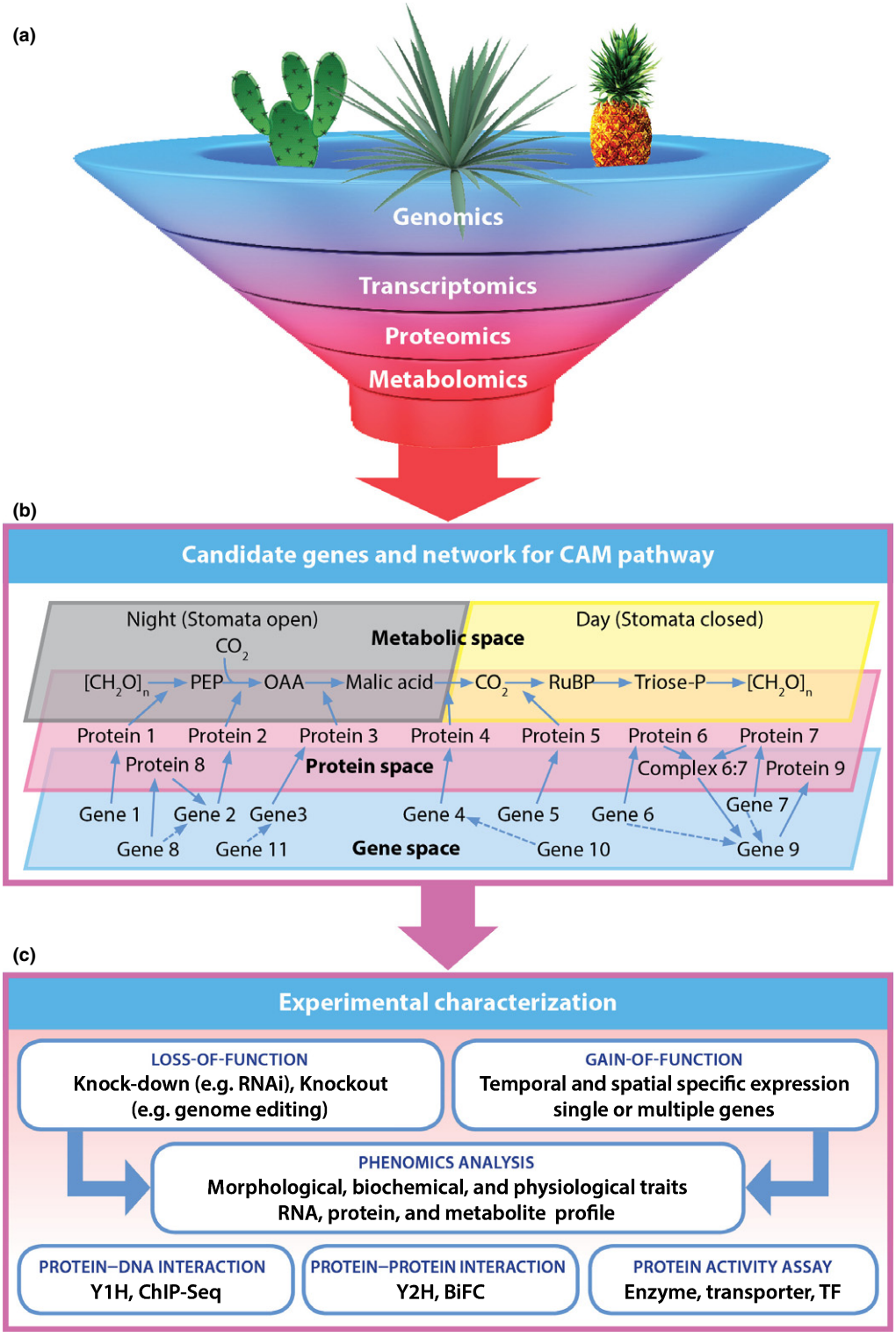

Fig. 3 An integrative functional genomics approach for crassulacean acid metabolism (CAM) plants. (a) Omics data (i.e. genomics, transcriptomics, proteomics, metabolomics) are generated for CAM plants, funnel image © 2013 TimArbaev, *Opuntia* image © andylin, *Agave* image © SSSCCC, pineapple image © 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). $[CH_2O]_n$, carbohydrates; OAA, oxaloacetate; PEP, phospho*enol*pyruvate; RuBP, ribulose 1,5-bisphosphate; triose-P, triose phosphate; TF, transcription factor; Y1H, yeast one-hybrid; Y2H, yeast two-hybrid; BiFC, bimolecular fluorescence complementation.

New Phytologist

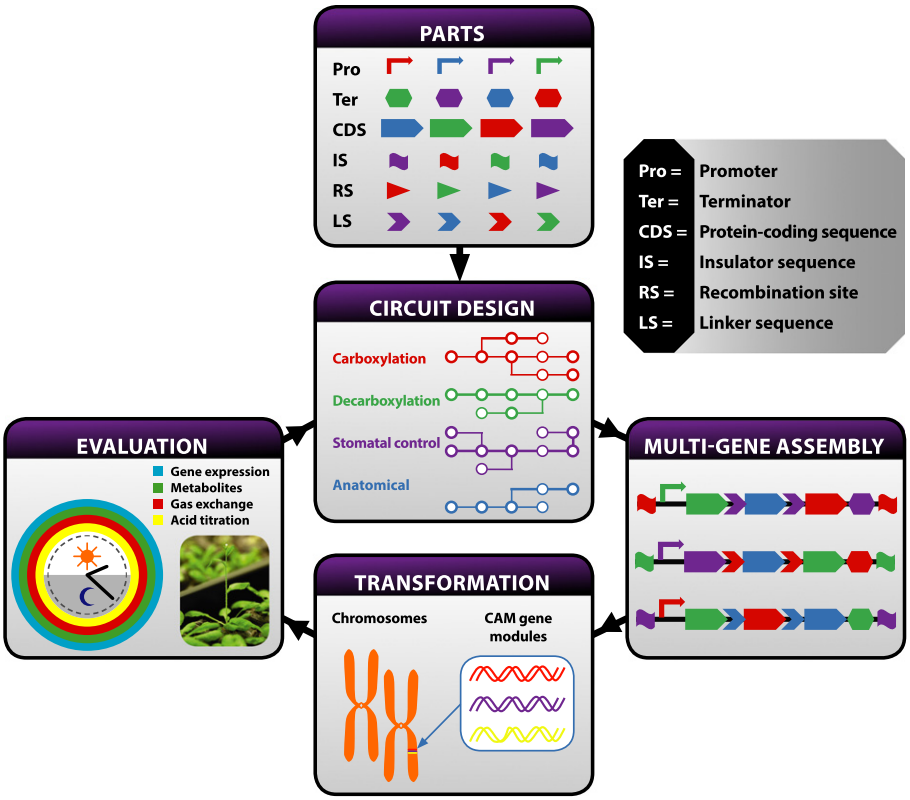

Fig. 4 Crassulacean acid metabolism (CAM) engineering using synthetic biology approaches.

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 C_3 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 C_3 and C_4 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* (C_3), poplar (C_3), willow (*Salix* spp.; C_3), *Miscanthus* (C_4), sugarcane (C_4), and switchgrass (C_4).

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 Cortázar & 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 C_3 and C_4 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 C_3 and CAM (and potentially also C_4) 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 C_3, CAM and C_4 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 (Lüttge, 2006, 2008). This genus includes obligate CAM and C_3 species as well as species that show facultative CAM and reversible shifts between C_3 and CAM (Winter & Holtum, 2014). A unique opportunity exists with *Portulaca,* a lineage that includes the only known examples of C_4 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 C_4 and CAM could facilitate engineering of both pathways into a single plant. Another good model lineage is

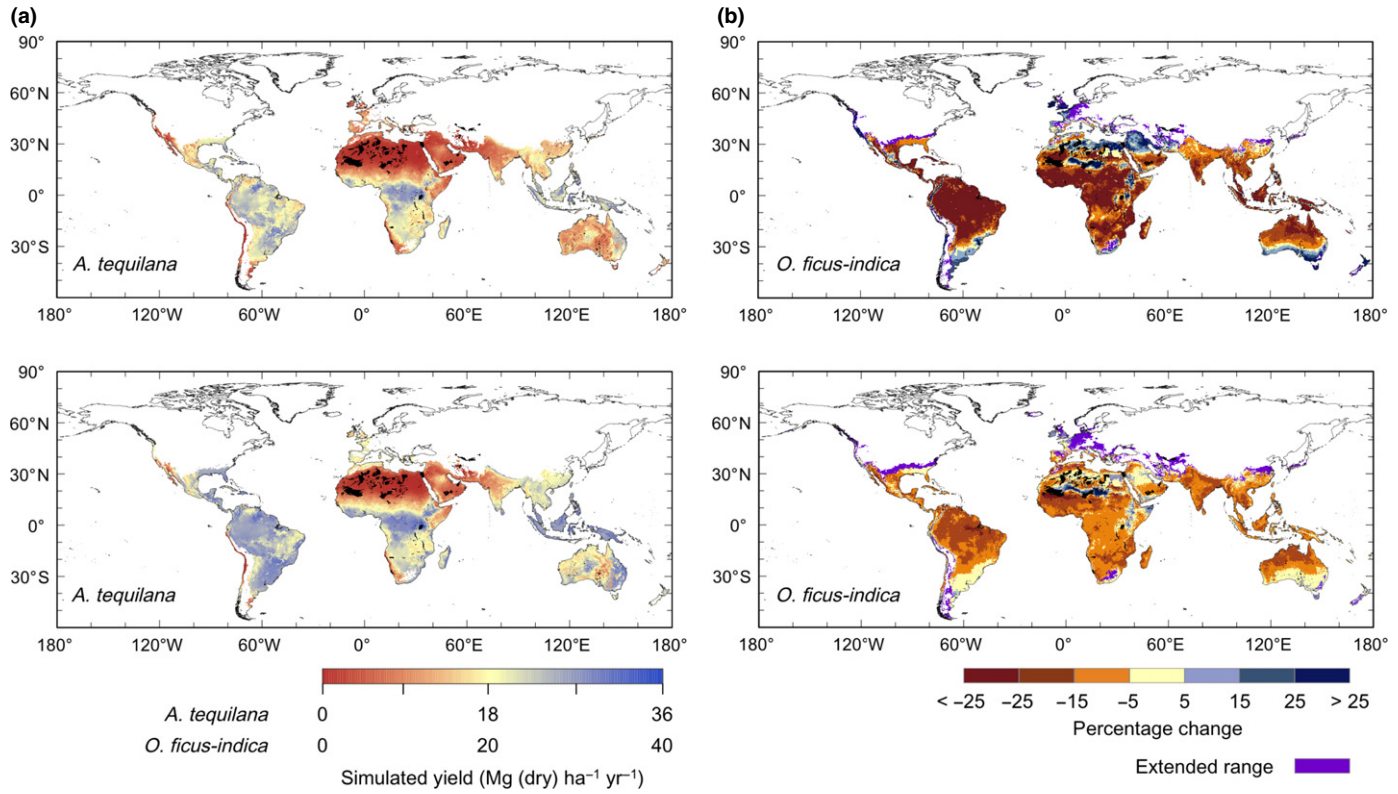

Fig. 5 Predicted productivity of *Agave tequilana* and *Opuntia ficus-indica* under current and future climate conditions. (a) Simulations under current climate conditions show that the geographical distribution of highly productive areas (environmental productivity index (EPI) > 0.5) is more restricted for *A. tequilana* than for *O. ficus-indica* because *A. tequilana* has a comparatively higher sensitivity to nocturnal temperature and lower capacity to buffer against periods of low soil water potential. The higher saturation point for carbon uptake response to photosynthetically active radiation (PAR) of *O. ficus-indica* compared to *A. tequilana* (35 vs 29 mol m^{-2} d^{-1}, respectively) has a negative impact on yields at latitudes > 30°S or 30°N. The productivity distribution of both species is restricted to areas where t_{min} > 0°C; simulations used environmental inputs averaged over the period 1950–2000; EPI was scaled with a value for maximum productivity (P_m) that could occur under irrigation and optimal planting density (44 and 46 Mg (dry) ha^{-1} yr^{-1} for *A. tequilana* and *O. ficus-indica*, respectively). Productivity simulations may be linearly re-scaled for different values of P_m. (b) Simulated productivity under future climate conditions show the percentage change in productivity between the present and worst-case climate scenario in the year 2070 (AR5 representative concentration pathway 8.5 W m^{-2}, 70RCP8.5). Inside the range of latitudes from 30°S to 30°N, *A. tequilana* simulations suggest that climate change will have a greater negative impact on productivity compared to *O. ficus-indica*. Outside this range, climate change has a beneficial impact on *A. tequilana* productivity. In general, *O. ficus-indica* displays stronger resilience to climate impacts.

Erycina, a group of orchids containing members that perform C_3 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 C_3–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 C_3 or C_4 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; Moßhammer *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 C_3 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 genomic selection methods for genetic 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

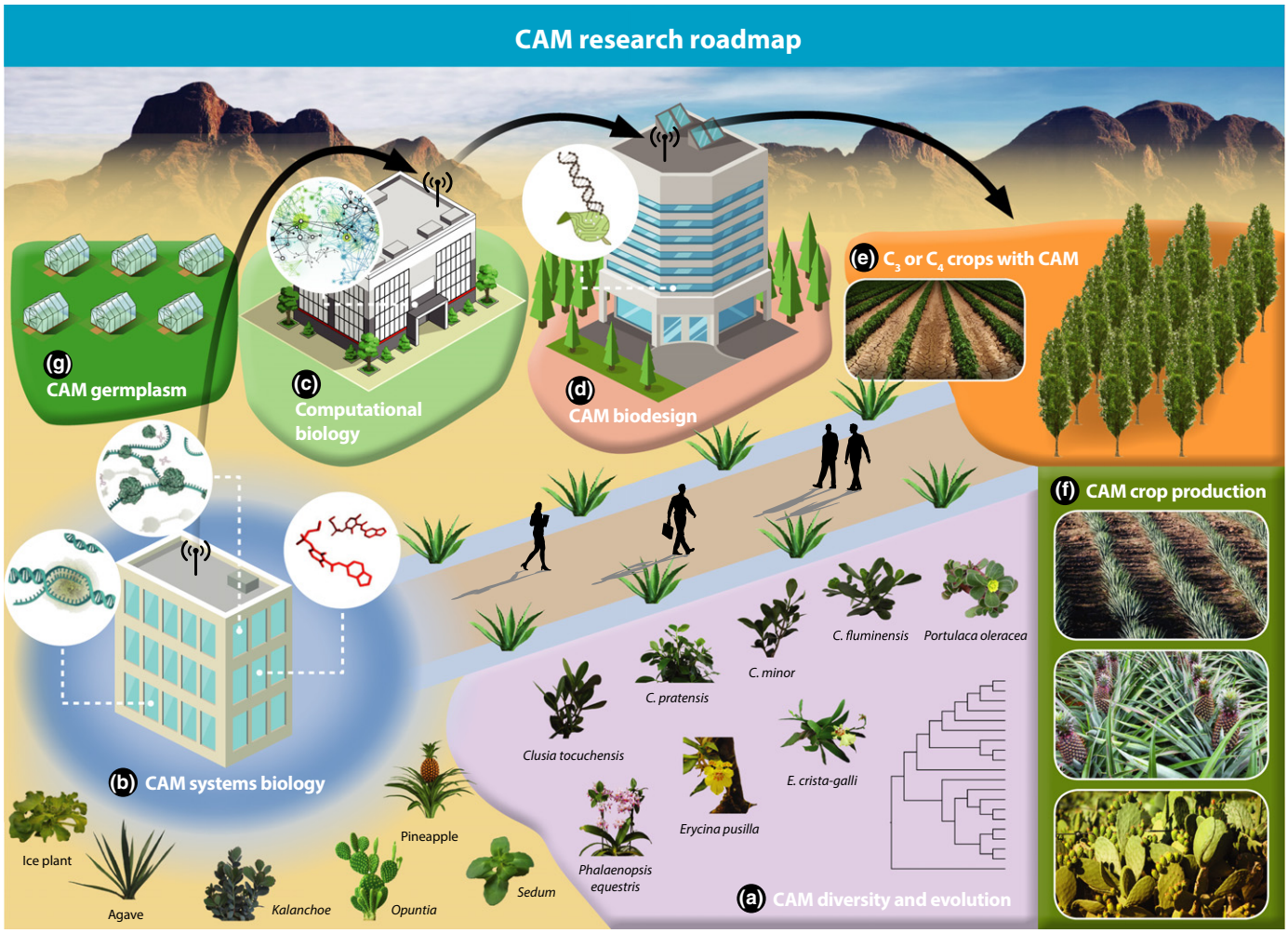

Fig. 6 Crassulacean acid metabolism (CAM) research roadmap. (a) Genetic diversity and evolution of CAM. (b) CAM systems biology with a focus on genomics, proteomics, and metabolomics, building image © Mayrum, antenna image © 2008 Angelhell, molecule images courtesy of the European Bioinformatics Institute at http://www.ebi.ac.uk/training/online/course/introduction-metabolomics/what-metabolomics, *Opuntia* image © 2011 Li Jingwang, pineapple image © 2014 Denyshuter. (c) Data management and analysis, building image © 2011 Chuvipro, network image © 2014 aleksandarvelasevic. (d) CAM engineering, building image © 2014 Bilgic. (e) C_3 or C_4 crops engineered with CAM for improvement in water-use efficiency and drought tolerance, cotton image © 2010 EIFlacodelNorte, poplar image © 2008 DNT59. (f) CAM crop production, *Agave* image © Noradoa, pineapple image © 2011 sybil, *Opuntia* image © 2012 Pgiam; and (g) CAM plant germplasm collection, glasshouse image © 2011 VLADGRIN, people images © Rawpixel, *Agave* images © 2009 Magnolja.

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 C_3 species as comparators (Figs 2, 3, 6c). This comparative analysis could identify four types of CAM genes: (1) CAM genes that have functionally equivalent C_3 gene orthologs without significant differences in developmental, temporal, or stress-responsive expression patterns between CAM and C_3 species; (2) CAM genes that have functionally equivalent C_3 gene orthologs with significant differences in developmental, temporal, or stress-responsive expression patterns between CAM and C_3 species; (3) CAM genes that have orthologs in C_3 species but have gained new function; and (4) CAM-specific genes that have no orthologs in C_3 species. Knowledge about these four types of CAM-related genes could inform the best strategy for CAM-into-C_3 engineering (Figs 4, 6d) to enable enhanced food and bioenergy production on drylands using existing C_3 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 C_3 species, and their integration with the C_3 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 C_3 gene orthologs of the type 2 and 3 CAM genes might be useful to avoid potential conflicts between CAM and C_3 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 Yang[1]*, John C. Cushman[2], Anne M. Borland[1,3], Erika J. Edwards[4], Stan D. Wullschleger[5], Gerald A. Tuskan[1], Nick A. Owen[6], Howard Griffiths[6], J. Andrew C. Smith[7], Henrique C. De Paoli[1], David J. Weston[1], Robert Cottingham[1], James Hartwell[8], Sarah C. Davis[9], Katia Silvera[10], Ray Ming[11,12], Karen Schlauch[13], Paul Abraham[14], J. Ryan Stewart[15], Hao-Bo Guo[16], Rebecca Albion[2], Jungmin Ha[2], Sung Don Lim[2], Bernard W. M. Wone[2], Won Cheol Yim[2], Travis Garcia[2], Jesse A. Mayer[2], Juli Petereit[13], Sujithkumar S. Nair[5], Erin Casey[3], Robert L. Hettich[14], Johan Ceusters[17], Priya Ranjan[1], Kaitlin J. Palla[1], Hengfu Yin[18], Casandra Reyes-García[19], José Luis Andrade[19], Luciano Freschi[20], Juan D. Beltrán[7], Louisa V. Dever[8], Susanna F. Boxall[8], Jade Waller[8], Jack Davies[8], Phaitun Bupphada[8], Nirja Kadu[8], Klaus Winter[10], Rowan F. Sage[21], Cristobal N. Aguilar[22], Jeremy Schmutz[23,24], Jerry Jenkins[23] and Joseph A. M. Holtum[25]

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6407, USA;
[2]Department of Biochemistry and Molecular Biology, University of Nevada, MS330, Reno, NV 89557-0330, USA;
[3]School of Biology, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK;
[4]Department of Ecology and Evolutionary Biology, Brown University, Box G-W, Providence, RI 02912, USA;
[5]Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6301, USA;
[6]Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge, CB2 3EA, UK;
[7]Department of Plant Sciences, University of Oxford, South Parks Road, Oxford, OX1 3RB, UK;
[8]Department of Plant Sciences, Institute of Integrative Biology, University of Liverpool, Liverpool, L69 7ZB, UK;
[9]Voinovich School of Leadership and Public Affairs and Department of Environmental and Plant Biology, Ohio University, Athens, OH 45701, USA;
[10]Smithsonian Tropical Research Institute, PO Box 0843-03092, Balboa, Ancon, Republic of Panama;
[11]Department of Plant Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA;
[12]FAFU and UIUC-SIB Joint Center for Genomics and Biotechnology, Fujian Agriculture and Forestry University, Fuzhou 350002, China;
[13]Nevada Center for Bioinformatics, University of Nevada, MS330, Reno, NV 89557-0330, USA;
[14]Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA;
[15]Department of Plant and Wildlife Sciences, Brigham Young University, 4105 Life Sciences Building, Provo, UT 84602, USA;
[16]Department of Biochemistry and Cellular and Molecular Biology, University of Tennessee, Knoxville, TN 37996, USA;
[17]Department of M²S, Faculty of Engineering Technology, TC Bioengineering Technology, KU Leuven, Campus Geel, Kleinhoefstraat 4, B-2440, Geel, Belgium;
[18]Key Laboratory of Forest Genetics and Breeding, Research Institute of Subtropical Forestry, Chinese Academy of Forestry, Fuyang, 311400, China;
[19]Centro de Investigación Científica de Yucatán, Calle 43 No. 130, Colonia Chuburná de Hidalgo, CP 97200, Mérida, México;
[20]Department of Botany, University of São Paulo, São Paulo 05508-090, Brazil;
[21]Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, ON M5S3B2, Canada;
[22]Department of Food Research, School of Chemistry, Universidad Autónoma de Coahuila, Saltillo, México;
[23]HudsonAlpha Institute for Biotechnology, 601 Genome Way, Huntsville, AL 35801, USA;
[24]US Department of Energy Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA;
[25]College 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

Aida R, Shibata M. 1996. Transformation of *Kalanchoe blossfeldiana* mediated by *Agrobacterium tumefaciens* and transgene silencing. *Plant Science* **121**: 175–185.

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

Borland AM. 1996. A model for the partitioning of photosynthetically fixed carbon during the C_3–CAM transition in *Sedum telephium*. *New Phytologist* **134**: 433–444.

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

Borland AM, Hartwell J, Weston D, Schlauch K, Tschaplinski T, Tuskan G, Yang X, Cushman JC. 2014. Engineering crassulacean acid metabolism to improve water-use efficiency. *Trends in Plant Science* **19**: 327–338.

Borland AM, Yang X. 2013. Informing the improvement and biodesign of crassulacean acid metabolism via system dynamics modelling. *New Phytologist* **200**: 946–949.

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

Chao Y-E, Zhang M, Feng Y, Yang X-E, Islam E. 2010. cDNA-AFLP analysis of inducible gene expression in zinc hyperaccumulator *Sedum alfredii* Hance under zinc induction. *Environmental and Experimental Botany* **68**: 107–112.

Cheung CM, Poolman MG, Fell DA, Ratcliffe RG, Sweetlove LJ. 2014. A diel flux balance model captures interactions between light and dark metabolism during day–night cycles in C_3 and crassulacean acid metabolism leaves. *Plant Physiology* **165**: 917–929.

Christin P-A, Arakaki M, Osborne CP, Bräutigam A, Sage RF, Hibberd JM, Kelly S, Covshoff S, Wong GK-S, Hancock L *et al.* 2014. Shared origins of a key enzyme during the evolution of C_4 and CAM metabolism. *Journal of Experimental Botany* **65**: 3609–3621.

Christin P-A, Arakaki M, Osborne CP, Edwards EJ. 2015. Genetic enablers underlying the clustered origins of C_4 photosynthesis in angiosperms. *Molecular Biology and Evolution*. doi: 10.1093/molbev/msu410.

Colunga-GarcíaMarín P, Zizumbo-Villarreal D. 2007. Tequila and other *Agave* spirits from west-central Mexico: current germplasm diversity, conservation and origin. *Biodiversity and Conservation* **16**: 1653–1667.

Cook B, Smerdon J, Seager R, Coats S. 2014. Global warming and 21st century drying. *Climate Dynamics* **43**: 1–21.

Crayn DM, Winter K, Schulte K, Smith JAC. 2015. Photosynthetic pathways in Bromeliaceae: phylogenetic and ecological significance of CAM and C_3 based on carbon-isotope ratios for 1893 species. *Botanical Journal of the Linnean Society*. doi: 10.1111/boj.12275.

Crayn DM, Winter K, Smith JAC. 2004. Multiple origins of crassulacean acid metabolism and the epiphytic habit in the Neotropical family Bromeliaceae. *Proceedings of the National Academy of Sciences, USA* **101**: 3703–3708.

Cushman JC, Agarie S, Albion R, Elliot S, Taybi T, Borland A. 2008a. Isolation and characterization of mutants of common ice plant deficient in crassulacean acid metabolism. *Plant Physiology* **147**: 228–238.

Cushman JC, Bohnert HJ. 1999. Crassulacean acid metabolism: molecular genetics. *Annual Review of Plant Physiology and Plant Molecular Biology* **50**: 305–332.

Cushman JC, Davis SC, Yang X, Borland AM. 2015. Development and use of bioenergy feedstocks for semi-arid and arid lands. *Journal of Experimental Botany*. doi: 10.1093/jxb/erv087.

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

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

Davis SC. 2013. *Agave*: a potential bioenergy feedstock? *C4 + CAM Plant Biology Meeting*. 6–9 August 2013, Urbana, IL, USA. [WWW document] URL http://conferences.igb.illinois.edu/c4cam/ [accessed 23 March 2015].

Davis SC, Dohleman F, Long S. 2011. The global potential for *Agave* as a biofuel feedstock. *GCB Bioenergy* **3**: 68–78.

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.

Davis SC, Ming R, LeBauer DS, Long S. 2015. Toward system-level analysis of agricultural production from crassulacean acid metabolism (CAM): scaling from cell to full commercial production. *New Phytologist*. doi: 10.1111/nph.13522.

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

Dever 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.

Eddy SR. 2001. Non-coding RNA genes and the modern RNA world. *Nature Reviews Genetics* **2**: 919–929.

Edwards EJ, Ogburn RM. 2012. Angiosperm responses to a low-CO_2 world: CAM and C_4 photosynthesis as parallel evolutionary trajectories. *International Journal of Plant Sciences* **173**: 724–733.

Feugang J, Konarski P, Zou D, Stintzing F, Zou C. 2006. Nutritional and medicinal use of Cactus pear (*Opuntia* spp.) cladodes and fruits. *Frontiers in Bioscience* **11**: 2574–2589.

Freschi L, Mercier H. 2012. Connecting environmental stimuli and crassulacean acid metabolism expression: phytohormones and other signaling molecules. *Progress in Botany* **73**: 231–255.

Freschi L, Rodrigues MA, Domingues DS, Purgatto E, Van Sluys M-A, Magalhaes JR, Kaiser WM, Mercier H. 2010. Nitric oxide mediates the hormonal control of crassulacean acid metabolism expression in young pineapple plants. *Plant Physiology* **152**: 1971–1985.

Gao J, Sun L, Yang X, Liu J-X. 2013. Transcriptomic analysis of cadmium stress response in the heavy metal hyperaccumulator *Sedum alfredii* Hance. *PLoS One* **8**: e64643.

Garces HMP, Champagne CEM, Townsley BT, Park S, Malho R, Pedroso MC, Harada JJ, Sinha NR. 2007. Evolution of asexual reproduction in leaves of the genus *Kalanchoe*. *Proceedings of the National Academy of Sciences, USA* **104**: 15578–15583.

Garcia de Cortázar V, Nobel PS. 1990. Worldwide environmental productivity indices and yield predictions for a CAM plant, *Opuntia ficus-indica*, including effects of doubled CO_2 levels. *Agricultural and Forest Meteorology* **49**: 261–279.

Garcia-Sogo B, Pineda B, Castelblanque L, Anton T, Medina M, Roque E, Torresi C, Beltran JP, Moreno V, Canas LA. 2010. Efficient transformation of *Kalanchoe blossfeldiana* and production of male-sterile plants by engineered anther ablation. *Plant Cell Reports* **29**: 61–77.

Gerland P, Raftery A, Sevcikova H, Li N, Gu D, Spoorenberg T, Alkema L, Fosdick B, Chunn J, Lalic N *et al.* 2014. World population stabilization unlikely this century. *Science* **346**: 234–237.

Givnish TJ, Barfuss MHJ, Ee BV, Riina R, Schulte K, Horres R, Gonsiska PA, Jabaily RS, Crayn DM, Smith JAC *et al.* 2011. Phylogeny, adaptive radiation, and historical biogeography in Bromeliaceae: insights from an eight-locus plastid phylogeny. *American Journal of Botany* **98**: 872–895.

Givnish TJ, Barfuss MHJ, Van Ee B, Riina R, Schulte K, Horres R, Gonsiska PA, Jabaily RS, Crayn DM, Smith JAC *et al.* 2014. Adaptive radiation, correlated and contingent evolution, and net species diversification in Bromeliaceae. *Molecular Phylogenetics and Evolution* **71**: 55–78.

Groenhof AC, Smirnoff N, Bryant JA. 1990. The appearance of a new molecular species of phosphoenolpyruvate carboxylase (PEPC) and the rapid induction of CAM in *Sedum telephium* L. *Plant, Cell & Environment* **13**: 437–445.

Guye P, Li Y, Wroblewska L, Duportet X, Weiss R. 2013. Rapid, modular and reliable construction of complex mammalian gene circuits. *Nucleic Acid Research* **41**: e156.

Hartwell J. 2006. The circadian clock in CAM plants. In: Hall AJW, McWatters H, eds. *Annual plant reviews: endogenous plant rhythms*. Oxford, UK: Blackwell Publishing, 211–236.

Hartwell J, Nimmo G, Wilkins M, Jenkins G, Nimmo H. 1999. Phospho*enol*pyruvate carboxylase kinase is a novel protein kinase regulated at the level of expression. *Plant Journal* **20**: 333–342.

Hibberd JM, Covshoff S. 2010. The regulation of gene expression required for C_4 photosynthesis. *Annual Review of Plant Biology* **61**: 181–207.

Hirsch CN, Foerster JM, Johnson JM, Sekhon RS, Muttoni G, Vaillancourt B, Peñagaricano F, Lindquist E, Pedraza MA, Barry K. 2014. Insights into the maize pan-genome and pan-transcriptome. *Plant Cell* **26**: 121–135.

Holtum JAM, Chambers D. 2010. *Feasibility of Agave as a feedstock for biofuel production in Australia*. Canberra, ACT, Australia: Rural Industry Research and Development Corporation.

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

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

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

Koch KE, Kennedy RA. 1980. Characteristics of crassulacean acid metabolism in the succulent C_4 dicot, *Portulaca oleracea* L. *Plant Physiology* **65**: 193–197.

Koch KE, Kennedy RA. 1982. Crassulacean acid metabolism in the succulent C_4 dicot, *Portulaca oleracea* L under natural environmental conditions. *Plant Physiology* **69**: 757–761.

Kore-eda S, Nozawa A, Okada Y, Takashi K, Azad M, Ohnishi J, Nishiyama Y, Tozawa Y. 2013. Characterization of the plastidic phosphate translocators in the inducible crassulacean acid metabolism plant *Mesembryanthemum crystallinum*. *Bioscience Biotechnology and Biochemistry* **77**: 1511–1516.

Krouk G, Lingeman J, Colon AM, Coruzzi G, Shasha D. 2013. Gene regulatory networks in plants: learning causality from time and perturbation. *Genome Biology* **14**: 123.

Le Houérou H. 1996. The role of cacti (*Opuntia* spp.) in erosion control, land reclamation, rehabilitation and agricultural development in the Mediterranean Basin. *Journal of Arid Environments* **33**: 135–159.

Lewis SM, Gross S, Visel A, Kelly M, Morrow W. 2015. Fuzzy GIS-based multi-criteria evaluation for US *Agave* production as a bioenergy feedstock. *GCB Bioenergy* **7**: 84–99.

Li HJ, Pattathil S, Foston MB, Ding SY, Kumar R, Gao XD, Mittal A, Yarbrough JM, Himmel ME, Ragauskas AJ *et al.* 2014. *Agave* proves to be a low recalcitrant lignocellulosic feedstock for biofuels production on semi-arid lands. *Biotechnology for Biofuels* **7**: 50.

Li P, Ponnala L, Gandotra N, Wang L, Si Y, Tausta SL, Kebrom TH, Provart N, Patel R, Myers CR *et al.* 2010. The developmental dynamics of the maize leaf transcriptome. *Nature Genetics* **42**: 1060–1067.

Lüttge U. 2006. Photosynthetic flexibility and ecophysiological plasticity: questions and lessons from *Clusia*, the only CAM tree, in the neotropics. *New Phytologist* **171**: 7–25.

Lüttge U. 2008. *Clusia*: Holy Grail and enigma. *Journal of Experimental Botany* **59**: 1503–1514.

Marbach D, Costello JC, Küffner R, Vega NM, Prill RJ, Camacho DM, Allison KR, Kellis M, Collins JJ, Stolovitzky G. 2012. Wisdom of crowds for robust gene network inference. *Nature Methods* **9**: 796–804.

McBride AC, Dale VH, Baskaran LM, Downing ME, Eaton LM, Efroymson RA, Garten CT Jr, Kline KL, Jager HI, Mulholland PJ. 2011. Indicators to support environmental sustainability of bioenergy systems. *Ecological Indicators* **11**: 1277–1289.

Michelet L, Zaffagnini M, Morisse S, Sparla F, Pérez-Pérez ME, Francia F, Danon A, Marchand CH, Fermani S, Trost P. 2013. Redox regulation of the Calvin–Benson cycle: something old, something new. *Frontiers in Plant Science* **4**: 470.

Middleton AM, Farcot E, Owen MR, Vernoux T. 2012. Modeling regulatory networks to understand plant development: small is beautiful. *Plant Cell* **24**: 3876–3891.

Miguez FE, Maughan M, Bollero GA, Long SP. 2012. Modeling spatial and dynamic variation in growth, yield, and yield stability of the bioenergy crops *Miscanthus* × *giganteus* and *Panicum virgatum* across the conterminous United States. *GCB Bioenergy* **4**: 509–520.

Moghaddam MRB, Van den Ende W. 2013. Sweet immunity in the plant circadian regulatory network. *Journal of Experimental Botany* **64**: 1439–1449.

Monson RK. 2003. Gene duplication, neofunctionalization, and the evolution of C_4 photosynthesis. *International Journal of Plant Sciences* **164**: S43–S54.

Moßhammer M, Stintzing F, Carle R. 2006. Cactus pear fruits (*Opuntia* spp.): a review of processing technologies and current uses. *Journal of the Professional Association for Cactus Development* **8**: 1–25.

Nair SS, Kang S, Zhang X, Miguez FE, Izaurralde RC, Post WM, Dietze MC, Lynd LR, Wullschleger SD. 2012. Bioenergy crop models: descriptions, data requirements, and future challenges. *GCB Bioenergy* **4**: 620–633.

Nava-Cruz N, Medina-Morales M, Martinez J, Rodriguez R, Aguilar C. 2014. *Agave* biotechnology: an overview. *Critical Reviews in Biotechnology* **24**: 1–14.

Nobel PS. 1984. Productivity of *Agave deserti*: measurement by dry weight and monthly prediction using physiological responses to environmental parameters. *Oecologia* **64**: 1–7.

Nobel PS. 1994. *Remarkable Agaves and cacti*. New York, NY, USA: Oxford University Press.

Nunez HM, Rodriguez LF, Khanna M. 2011. *Agave* for tequila and biofuels: an economic assessment and potential opportunities. *GCB Bioenergy* **3**: 43–57.

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

Owen NA, Griffiths H. 2014. Marginal land bioethanol yield potential of four crassulacean acid metabolism candidates (*Agave fourcroydes*, *Agave salmiana*, *Agave tequilana* and *Opuntia ficus-indica*) in Australia. *GCB Bioenergy* **6**: 687–703.

Raftery AE, Alkema L, Gerland P. 2014. Bayesian population projections for the United Nations. *Statistical Science* **29**: 58–68.

Raftery AE, Chunn JL, Gerland P, Ševčíková H. 2013. Bayesian probabilistic projections of life expectancy for all countries. *Demography* **50**: 777–801.

Raftery AE, Li N, Sevcikova H, Gerland P, Heilig GK. 2012. Bayesian probabilistic population projections for all countries. *Proceedings of the National Academy of Sciences, USA* **109**: 13915–13921.

Russell C, Felker P. 1987. The prickly-pears (*Opuntia* spp., Cactaceae): a source of human and animal food in semiarid regions. *Economic Botany* **41**: 433–445.

She W, Lin W, Zhu Y, Chen Y, Jin W, Yang Y, Han N, Bian H, Zhu M, Wang J. 2010. The *gypsy* insulator of *Drosophila melanogaster*, together with its binding protein suppressor of hairy-wing, facilitate high and precise expression of transgenes in *Arabidopsis thaliana*. *Genetics* **185**: 1141–1150.

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

Silvera K, Santiago LS, Cushman JC, Winter K. 2010. The incidence of crassulacean acid metabolism in Orchidaceae derived from carbon isotope ratios: a checklist of the flora of Panama and Costa Rica. *Botanical Journal of the Linnean Society* **163**: 194–222.

Silvestro D, Zizka G, Schulte K. 2014. Disentangling the effects of key innovations on the diversification of Bromelioideae (Bromeliaceae). *Evolution* **68**: 163–175.

Tam GHF, Chang C, Hung YS. 2013. Gene regulatory network discovery using pairwise Granger causality. *IET Systems Biology* **7**: 195–204.

Taybi T, Patil S, Chollet R, Cushman J. 2000. A minimal Ser/Thr protein kinase circadianly regulates phospho*enol*pyruvate carboxylase activity in CAM-induced leaves of *Mesembryanthemum crystallinum*. *Plant Physiology* **123**: 1471–1482.

Torella JP, Boehm CR, Lienert F, Chen J-H, Wau KC, Silver PA. 2014. Rapid construction of insulated genetic circuits via synthetic sequence-guided isothermal assembly. *Nucleic Acid Research* **42**: 681–689.

US DOE. 2014. *Lignocellulosic biomass for advanced biofuels and bioproducts: Workshop report*. US Department of Energy Office of Science. [WWW document] URL genomicscience.energy.gov/biofuels/lignocellulose/ [accessed 23 March 2015].

Voytas DF, Gao C. 2014. Precision genome engineering and agriculture: opportunities and regulatory challenges. *PLoS Biology* **12**: e1001877.

Wallach R, Da-Costa N, Raviv M, Moshelion M. 2010. Development of synchronized, autonomous, and self-regulated oscillations in transpiration rate of a whole tomato plant under water stress. *Journal of Experimental Botany* **61**: 3439–3449.

West AG, Gaszner M, Felsenfeld G. 2002. Insulators: many functions, many mechanisms. *Genes & Development* **16**: 271–288.

Winter K, Foster JG, Edwards GE, Holtum JAM. 1982. Intracellular localization of enzymes of carbon metabolism in *Mesembryanthemum crystallinum* exhibiting C_3 photosynthetic characteristics or performing crassulacean acid metabolism. *Plant Physiology* **69**: 300–307.

Winter K, Holtum JAM. 2002. How closely do the $\delta^{13}C$ values of crassulacean acid metabolism plants reflect the proportion of CO_2 fixed during day and night? *Plant Physiology* **129**: 1843–1851.

Winter K, Holtum JAM. 2005. The effects of salinity, crassulacean acid metabolism and plant age on the carbon isotope composition of *Mesembryanthemum crystallinum* L., a halophytic C_3–CAM species. *Planta* **222**: 201–209.

Winter K, Holtum JAM. 2007. Environment or development? Lifetime net CO_2 exchange and control of the expression of crassulacean acid metabolism in *Mesembryanthemum crystallinum*. *Plant Physiology* **143**: 98–107.

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

Winter K, Holtum JAM, Smith JAC. 2015. CAM photosynthesis: a continuous or discrete trait? *New Phytologist*. doi: 10.1111/nph.13446.

Yan X, Tan DKY, Inderwildi OR, Smith JAC, King DA. 2011. Life cycle energy and greenhouse gas analysis for *Agave*-derived bioethanol. *Energy & Environmental Science* **4**: 3110–3121.

Supporting Information

Additional supporting information may be found in the online version of this article.

Fig. S1 Predicted productivity of *Agave tequilana* and *Opuntia ficus-indica* in North America under current and future climate conditions.

Fig. S2 Data management and analyses for crassulacean acid metabolism (CAM) research based on the Department of Energy (DOE) KBase.

Table S1 A list of selected crassulacean acid metabolism (CAM) crops and information related to their productivity, economic value, and natural products

Table S2 The hypothesized minimal gene sets for the different crassulacean acid metabolism (CAM) modules

Table S3 Transcriptomic and genomic information about major crassulacean acid metabolism (CAM) model species

Please note: Wiley Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.

Key words: bioenergy, crassulacean acid metabolism (CAM), drought, genomics, photosynthesis, roadmap, synthetic biology, water-use efficiency (WUE).