OPINION

The genetics of convergent evolution: insights from plant photosynthesis

Karolina Heyduk, Jose J. Moreno-Villena, Ian S. Gilman, Pascal-Antoine Christin and Erika J. Edwards

Abstract | The tree of life is resplendent with examples of convergent evolution, whereby distinct species evolve the same trait independently. Many highly convergent adaptations are also complex, which makes their repeated emergence surprising. In plants, the evolutionary history of two carbon concentrating mechanisms (CCMs) — C_4 and crassulacean acid metabolism (CAM) photosynthesis — presents such a paradox. Both of these modifications of ancestral C_3 photosynthesis require the integration of multiple anatomical and biochemical components, yet together they have evolved more than one hundred times. The presence of CCM enzymes in all plants suggests that a rudimentary CCM might emerge via relatively few genetic changes in potentiated lineages. Here, we propose that many of the complexities often associated with C_4 and CAM photosynthesis may have evolved during a post-emergence optimization phase. The ongoing development of new model clades for young, emerging CCMs is enabling the comparative studies needed to test these ideas.

Convergent evolution has been pervasive throughout the history of life. Even very complicated adaptations, such as camera eyes in animals[1], sex determination systems in eukaryotes[2] and eusociality in insects[3], have evolved multiple times. In plants, convergence has led to repeated transitions in flower colour[4,5] and scent profiles[6] to attract pollinators, multiple origins of parasitic lifestyles[7] and a multitude of other traits. Many convergent traits are genetically simple or perform secondary functions that are not linked to core metabolism. A notable exception is the repeated modification of one of the most fundamental processes on Earth, photosynthesis. Although the machinery involved in the sequestration of light energy is mostly conserved from cyanobacteria to flowering plants, the biochemical pathways involved in the capture of atmospheric carbon vary quite widely across photosynthetic organisms. In C_3 photosynthesis, which is used by most plants, atmospheric CO_2 is directly fixed by the Calvin cycle via the enzyme Rubisco. However, Rubisco has affinities for both CO_2 and O_2 (REF.[8]), and the processing of fixed O_2 releases CO_2 and results in energetically wasteful reactions[9]. To circumvent this problem, land plants have repeatedly evolved carbon concentrating mechanisms (CCMs) known as C_4 photosynthesis and crassulacean acid metabolism (CAM), which consist of both anatomical and biochemical adaptations that internally concentrate CO_2 before its fixation by Rubisco (FIG. 1), thereby making photosynthesis more efficient[10,11].

Over the past 25 years, improvements in our understanding of plant phylogenetic relationships have been instrumental in assessing the number and timing of CCM origins[12,13], their relationship with environmental factors[14–17] and the history of associated phenotypic changes[18–21]. Both C_4 and CAM photosynthesis represent textbook examples of convergent evolution, each having likely evolved independently more than 60 times[22,23]. On one hand, it is remarkable that CCMs could have evolved so frequently because of their seeming complexity and the fact that some relevant mutations would affect the primary metabolism of a plant. In this context, it might be supposed that maladaptive mutations would be quite common and that the evolutionary trajectory of each CCM might be exceedingly narrow and rarely stumbled upon[24,25]. On the other hand, the sheer number of CCM origins argues precisely the opposite — these traits must be relatively simple to evolve and might be assembled via multiple and varied trajectories and, as such, are evolutionarily accessible (sensu Maynard Smith[26,27]).

In this Perspective, we highlight evidence that supports the view that CCMs are evolutionarily accessible, with a focus on the genetic elements that might give rise to their repeated origins. We first review shared key components of and differences between C_4 and CAM photosynthesis, both of which use similar biochemical pathways. We highlight recent comparative studies in well-sampled clades that suggest that a newly emerging, rudimentary CCM may be less complex than many of the highly optimized C_4 and CAM species that are commonly studied. We briefly review the roles of CCM enzymes in C_3 species to emphasize that, at the broadest scale, all plants routinely express the essential biochemical building blocks of a functional CCM pathway. In this context, we discuss new evidence that some lineages may be genetically predisposed to evolve CCMs. Throughout the article, we emphasize how the growing field of comparative genomics is already catalysing discovery and improving our understanding of CCM evolution.

C_4 and CAM photosynthesis

Both C_4 and CAM photosynthesis work to reduce photorespiration, which occurs when Rubisco binds to O_2 instead of CO_2. Plant photorespiration increases with temperature owing to the decreased solubility of CO_2 and the reduced specificity of Rubisco at higher temperatures[28]. Photorespiration also increases with abiotic stresses, such as drought, that force stomata to close and prevent CO_2 from entering the leaf and reaching Rubisco[29]. It is thought that photorespiratory stress is the major driver of C_4 and CAM evolution[25,30], and many plant species inhabiting hot and dry habitats use one of these CCMs. Both CCMs require that acquisition of atmospheric CO_2 is separated from CO_2 fixation via the Calvin cycle

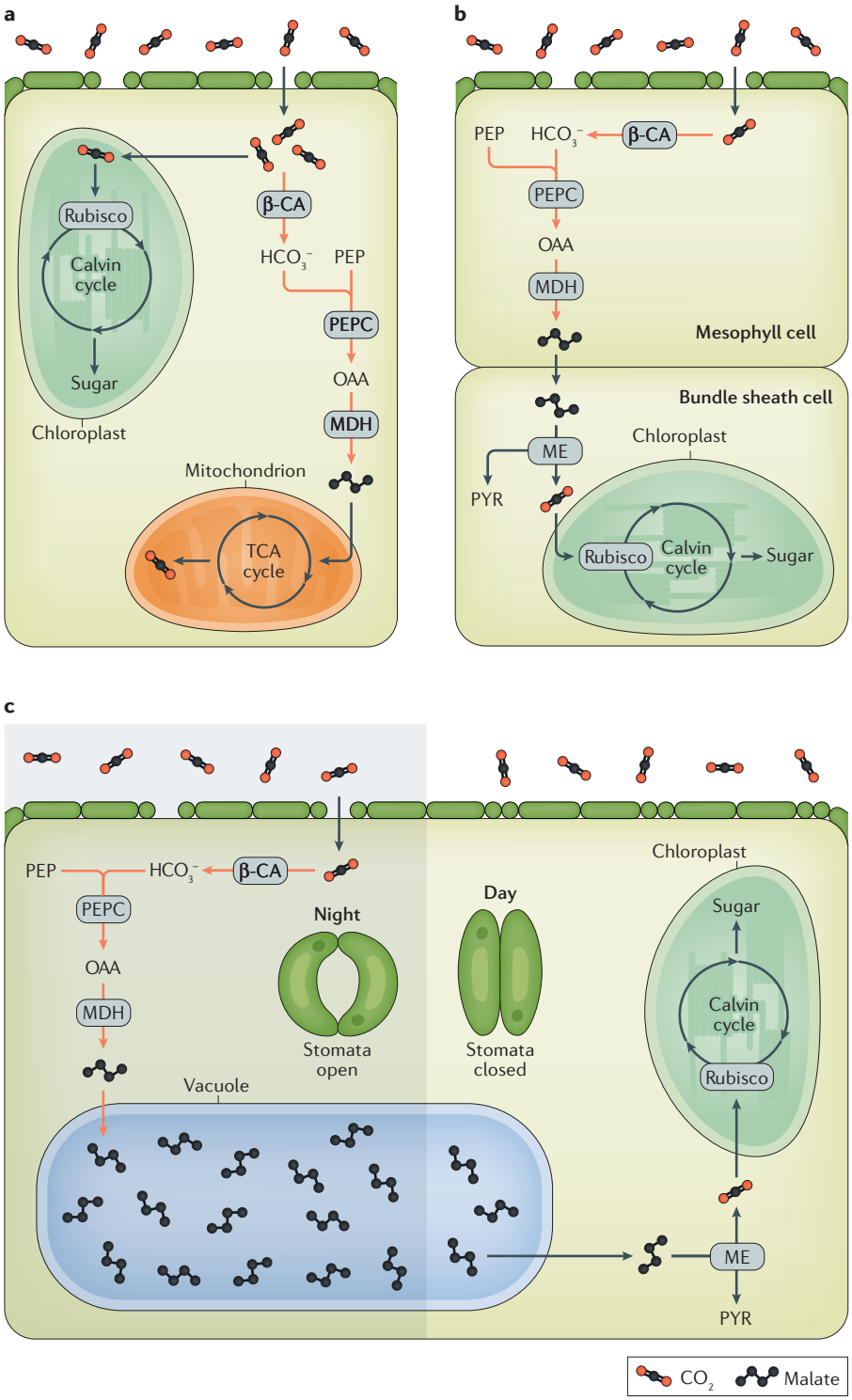


Fig. 1 | **Biochemistry of C_3 and CCM pathways.** **a** | The CO_2 that C_3 plants acquire from the atmosphere is largely fixed by Rubisco in the Calvin cycle. Small amounts of CO_2, mainly from respiration, can be used by anaplerotic pathways that replenish intermediates in the tricarboxylic acid (TCA) cycle. This process involves the conversion of CO_2 to bicarbonate (HCO_3^-) by a carbonic anhydrase (β-CA) and then subsequent carboxylation of phosphoenolpyruvate (PEP) by a carboxylase (PEPC) to a four-carbon acid, oxaloacetate (OAA). The OAA is further converted to malate by malate dehydrogenase (MDH); finally, malate is fed into the TCA cycle. **b** | The C_4 carbon concentrating mechanism (CCM) is largely composed of these same enzymes that are spatially separated into distinct compartments. In the most common type of C_4 photosynthesis, CO_2 diffuses into the mesophyll cells, where it is converted to malate via the same pathway (β-CA, PEPC and MDH). Malate is then transported into adjoining bundle sheath cells, where it is decarboxylated through one or two different malic enzymes (MEs), releasing CO_2 for efficient carboxylation via Rubisco. This process also releases pyruvate (PYR), which can be used to regenerate the PEP substrate or for carbohydrate production. Although not shown here, in some C_4 lineages, different intermediates are formed instead of malate (for example, aspartate), and other decarboxylases exist, which require slightly different biochemistry and transporters. **c** | Crassulacean acid metabolism (CAM) species are nearly identical to C_4 plants in their biochemical pathway of carbon acquisition and fixation, although separation of uptake and fixation occurs temporally rather than spatially. At night, CO_2 is taken up and converted to malate in the same manner as in C_4 plants but then stored in the vacuoles. During the day, the stomata close and malate is decarboxylated in the cytosol, again flooding Rubisco with high concentrations of CO_2. Decarboxylation can occur in a variety of locations within the cell, including the cytoplasm, mitochondria and chloroplasts.

and use the same set of enzymes to do so. However, C_4 and CAM plants fundamentally differ in how this separation is achieved (FIG. 1).

Spatial separation of CO_2 uptake and fixation in C_4 species. In C_4 plants, CO_2 uptake and fixation occur in spatially distinct compartments within the leaf. In most C_4 species, uptake and fixation occur in distinct mesophyll (M) cells and bundle sheath (BS) cells, respectively (FIG. 1), although variations exist across independent origins of C_4, and in several cases CO_2 assimilation and reduction take place in distinct regions of a single cell[31]. Specific leaf properties are required to sustain synchronized C_4 reactions across two distinct cell types, including high leaf vein density, high BS:M cell volume ratio and the localization of most chloroplasts to BS cells[32,33]. In C_4 plants, spatial separation of reactions results in a tenfold increase in CO_2 in the BS cells relative to M cells[34,35] and enables efficient carbon fixation in conditions where CO_2 is limiting, such as warm, open habitats in the low-CO_2 atmosphere that has prevailed since the Oligocene[12,36,37]. Typically occupying high-light habitats, C_4 plants have the highest measured photosynthetic rates of all plants owing to the extremely high carboxylation rates of the Calvin cycle when it is not limited by either CO_2 or light reaction products[38]. Examples of C_4 plant species include maize (*Zea mays*), sugarcane (*Saccharum officinarum*) and sorghum (*Sorghum bicolor*) in the grass family (Poaceae), as well as members of the sunflower family (*Flaveria*, Asteraceae) and amaranths (*Amaranthus*, Amaranthaceae).

Temporal separation of CO_2 uptake and fixation in CAM species. In CAM species, the carbon uptake and fixation reactions are temporally distinct. CO_2 is primarily acquired at night when transpiration rates are low and is converted to malate and stored in vacuoles as malic acid. During the day, stomata close and malic acid is decarboxylated, resulting in increased CO_2 concentrations around Rubisco[39]. The large amount of malic acid stored in CAM plants demands bigger vacuoles and cells, and therefore CAM is generally associated with a succulent leaf phenotype. The temporal configuration of reactions, the inverted stomatal behaviour and the low transpirational demand at night help CAM plants achieve the highest water use efficiencies in the plant world[40]. As a result, many desert species are CAM plants, including cacti (Cactaceae), agaves (Asparagaceae), and euphorbs (Euphorbiaceae). Less intuitively, many tropical forest plants such as orchids (Orchidaceae) and bromeliads (Bromeliaceae) also use CAM; these are mostly epiphytes (plants that grow on other plants) and thus can also experience substantial drought stress.

Alternative and variant CCMs. C_4 and CAM are quite distinct from C_3 photosynthesis, but a number of intermediate photosynthetic forms also exist. A third type of CCM, sometimes called C_2, concentrates CO_2 in BS cells by restricting elements of the photorespiratory cycle to those cells; C_2 has long been supposed to be a precursor to C_4 (REFS[30,41]), although it is theoretically not necessary for C_4 evolution[42]. A variety of other phenotypes that are intermediate between C_3 and C_4, termed C_3+C_4, have been identified, which have a limited C_4 photosynthetic cycle, sometimes in addition to a C_2 cycle[41]. CAM photosynthesis, unlike C_4, is exceptionally phenotypically plastic because all photosynthetic cells in CAM plants still have a fully functional C_3 cycle. As a result, many CAM species can perform varying degrees of CAM or C_3 carbon fixation depending on their developmental stage or on environmental conditions[43,44]. For example, there are many so-called C_3+CAM species, which predominantly rely on C_3 photosynthesis but fix a small amount of carbon via the CAM pathway nocturnally[45–47]. Some plant species can use both C_4 and CAM in a single individual[48] (FIG. 2a), although these instances are very rare. In general, intermediate C_3+CAM, like C_2 or C_3+C_4, is thought to be an evolutionary precursor to a full CAM metabolism[23,46].

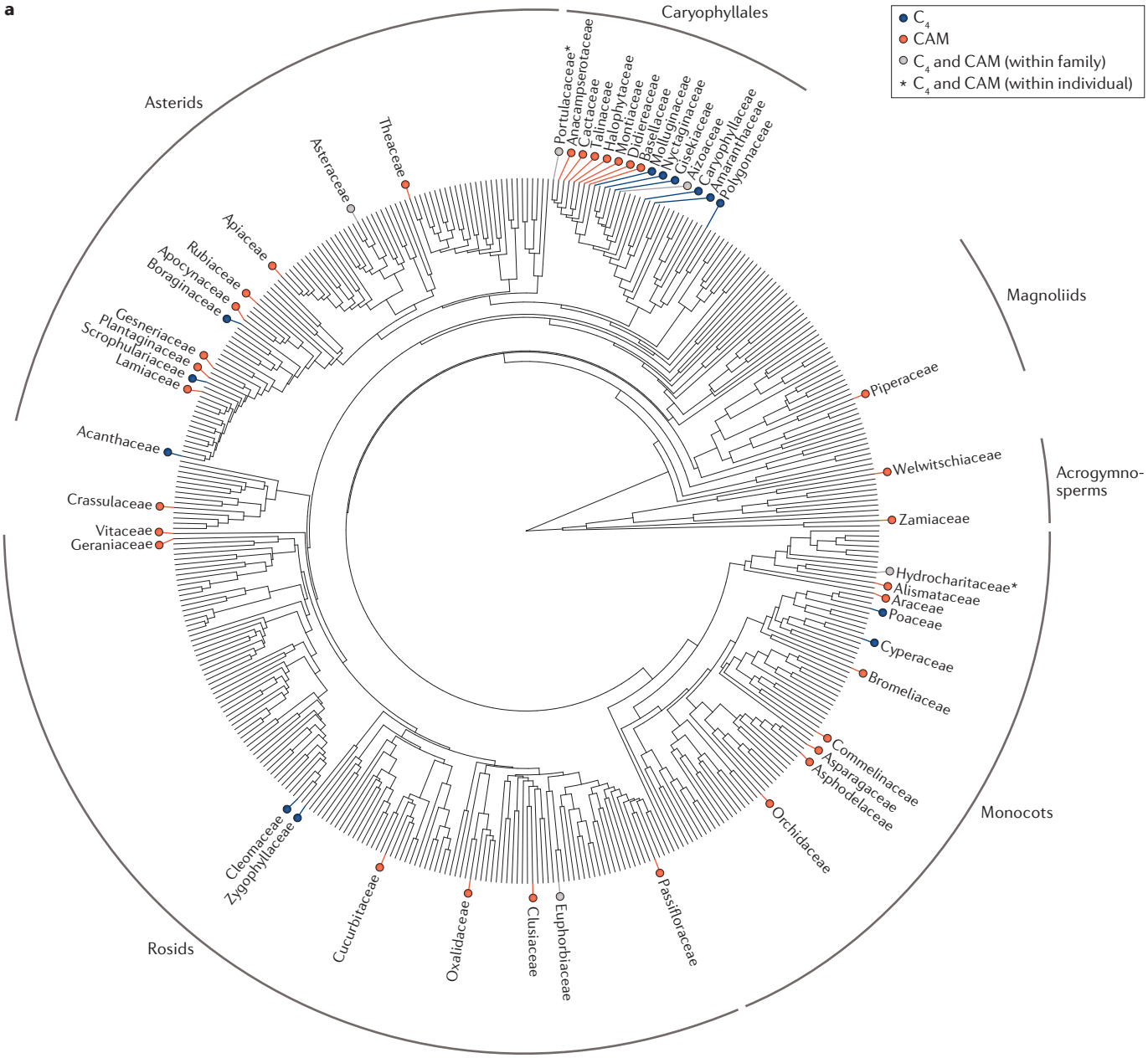
Insights from comparative genomics

The regulation of cellular and temporal compartmentalization of CCM biochemistry is complicated, and, unsurprisingly, the genomics era has facilitated a new surge of research into the genetics of these complex traits. The past few years have seen the rapid production of genomic resources for an increasing number of species. The first genome of a C_4 plant (*S. bicolor*) was assembled in 2009 (REF.[49]), and the first CAM genome (*Phalaenopsis equestris*, an orchid) was assembled in 2015 (REF.[50]), and genomes and transcriptomes are now being sequenced for a variety of plants spanning CCM — including *Z. mays* (maize) and *Ananas comosus* (pineapple)[50,51] — and non-CCM species[52]. While fully sequenced reference genomes are still relatively cost-prohibitive for plants with large genomes, sequencing expressed genes via transcriptomics can be done at substantially lower cost and is much easier to perform in non-model systems. Although transcriptomes are unable to uncover regulatory regions and other non-coding sequences, they provide information on relevant changes in gene expression as well as any changes to gene sequences and possible selection across the genome. The growing taxonomic diversity of genomic resources will permit well-designed comparative analyses to better identify the genetic changes underlying the evolution of new phenotypes, such as a CCM.

Broad comparisons likely overestimate CCM complexity. Examination of only extant diversity, which is often an unavoidable constraint, will likely overestimate the number of changes required for an initial transition from C_3 photosynthesis to a CCM. In general, comparisons between groups chosen because they differ in one particular aspect of their phenotype will also capture many unrelated differences, and the number of these irrelevant differences will increase with the evolutionary distance separating the groups being compared. Therefore, when traits are measured that differ between CCM and non-CCM lineages, the captured changes are not necessarily required for CCM evolution, as they may also include those that have occurred before or after the CCM origination (BOX 1; FIG. 2b). Although the changes that occur after the origin of the CCM are important for understanding CCM diversity, function and especially optimization, they may complicate an evaluation of the relative ease or difficulty with which CCMs have evolved.

Historically, many genomic studies of CCM evolution have relied on comparisons between distantly related C_3 and CCM taxa[53,54], largely because genomic resources were only available for model systems. Such broad evolutionary comparisons between C_3 and CCM species have often highlighted hundreds to thousands of differentially expressed genes[55–57]. Although such studies have been foundational for developing comparative techniques and hypotheses, more recent work has focused on closely related C_3 and CCM taxa for finer-scale resolution of the genetic underpinnings of CCMs[30,45,58,59]. In *Flaveria* (Asteraceae), a lineage comprising closely related C_3, C_4, C_2 and C_3+C_4 species, comparative studies found hundreds to over a thousand differentially expressed transcripts between different photosynthetic types[51]. Similarly, roughly 600 transcripts were differentially expressed in closely related C_3 and C_4 members of Cleomaceae[58]. Even by reducing evolutionary distance, comparative transcriptomic studies between closely related species are still likely inflating the estimates of genetic changes required for CCM origins.

Emerging model systems to study CCM evolution. Model clades such as *Flaveria* will continue to be highly relevant to CCM evolution, but the problem of overestimating complexity can be reduced further. Ideally, a spectrum of photosynthetic phenotypes, all within a single species, would be studied to capture the emergence of a rudimentary CCM. Such a system is currently being developed in the grass species *Alloteropsis semialata*[60]. The existence of C_3 and C_4 genotypes of *A. semialata* has been known for some time, but recent work has demonstrated a diverse spectrum of physiology that spans C_3 and C_4 in important ways. Phenotypically intermediate populations of *A. semialata* that perform a weak C_4 cycle have been identified[61], and genome-wide analyses have provided evidence of genetic exchange between different photosynthetic types[62]. C_4 populations of *A. semialata* show incomplete segregation of Rubisco into BS cells[63] and have a photosynthetic efficiency below that of other C_4 grasses[61] and can, therefore, be considered as having a rudimentary CCM. Comparative work has indicated that the initial emergence of C_4 photosynthesis in the group involved upregulation of only a handful of genes[64]. Subsequent refinement of the CCM in *Alloteropsis* is supported by the presence of genes better suited for the C_4 context in only some of the C_4 populations of *A. semialata* and by evidence of sustained positive selection within the C_4 group[64,65].


a
C4
CAM
C4 and CAM (within family)
C4 and CAM (within individual)
Caryophyllales
Asterids
Magnoliids
Acrogymno-sperms
Monocots
Rosids
Portulacaceae*
Anacampserotaceae
Cactaceae
Talinaceae
Halophytaceae
Montiaceae
Didiereaceae
Basellaceae
Molluginaceae
Nyctaginaceae
Cisekiaceae
Aizoaceae
Caryophyllaceae
Amaranthaceae
Polygonaceae
Theaceae
Asteraceae
Apiaceae
Rubiaceae
Apocynaceae
Boraginaceae
Gesneriaceae
Plantaginaceae
Scrophulariaceae
Lamiaceae
Acanthaceae
Crassulaceae
Vitaceae
Geraniaceae
Piperaceae
Welwitschiaceae
Zamiaceae
Hydrocharitaceae*
Alismataceae
Araceae
Poaceae
Cyperaceae
Bromeliaceae
Commelinaceae
Asparagaceae
Asphodelaceae
Orchidaceae
Passifloraceae
Euphorbiaceae
Clusiaceae
Oxalidaceae
Cucurbitaceae
Cleomaceae
Zygophyllaceae
b
True history
A
B
C†
D
E
F†
G†
H
I
J†
K
L
M
N
O†
P
Q†
R
S
T†
U
Reconstruction of phylogeny and character evolution
Complete taxon sampling
A
B
D
E
H
I
K
L
M
N
P
R
S
U
Limited taxon sampling
A
B
S
U

◀ Fig. 2 | **Evolutionary patterns of CCMs. a** | Lineages with carbon concentrating mechanisms (CCMs) are distributed across the phylogeny of vascular plants. In most cases, C_4 and crassulacean acid metabolism (CAM) are found in distinct lineages (blue and orange family labels, respectively); however, families in which there are examples of both C_4 and CAM photosynthesis exist (grey labels). Indeed, some individuals within some of these families (Portulacaceae and Hydrocharitaceae, indicated by asterisks) can use both C_4 and CAM photosynthesis. Only CCM lineages are labelled. Figure created using phylogeny from REF.[123]. **b** | Even with phylogenetic information, comparative studies can still inflate the number of changes required for a CCM to evolve. For example, in a theoretical tree of extant and extinct (†) taxa labelled A to U, a single origin of a CCM occurred at the blue dot. All blue branches indicate portions of the tree where the CCM exists. Black dots represent changes in ancestral C_3 species that pre-date the CCM origin, and yellow dots are changes that occurred after the CCM origin that may or may not be related to the CCM itself. If all extant taxa could be sampled, the number of changes inferred on the branch where the CCM occurred is relatively low; in a more realistic case, where only a subset of extant species are sampled, a greater number of changes are assigned to the branch with the CCM origin, even though many are unrelated to CCM evolution.

Although distinct C_3 and C_4 populations have not yet been reported in species other than *A. semialata*, we predict that new efforts to broadly screen wild populations of young CCM species will reveal subtle but important variation. For example, it is already known that the strength of the C_4 cycle varies between accessions of the C_3+C_4 intermediate *Mollugo verticillata*[66], potentially between populations of *Salsola divaricata*[59] and in the C_3+CAM intermediate hybrid *Yucca gloriosa*[67]. In addition, variation between accessions of the C_4 species *Gynandropsis gynandra* has been proposed as a system to identify the genetic determinants of C_4 photosynthesis[68]. The development of model systems in a greater number of species, particularly those with intermediate or variable CCM usage, will help determine whether patterns seen in *Alloteropsis* and *Flaveria* are representative of CCM evolution more generally. The generation of genomic resources for new and existing model species will enable genome-wide association studies and other approaches to detect the determinants of CCMs and potentially identify the precise genetic changes required for the initial emergence of a CCM and its subsequent adaptation.

All plants have key CCM elements

The known components of the CCM biochemical pathway are present in all vascular plants (FIG. 1), and the co-option of existing enzymes is a straightforward mechanism by which CCMs could initially develop. The core mechanism of carbon concentration — the fixation of CO_2 by the coupled action of the enzymes carbonic anhydrase and phosphoenolpyruvate carboxylase (PEPC) — is part of an anaplerotic pathway that supplements malate to the tricarboxylic acid (TCA) cycle and exists in the cytosol of all plants[69]. In a non-CCM plant, a rise in the concentration of cytosolic bicarbonate (HCO_3^-), for example, from an increase in respired CO_2, could result in increased malate production. If HCO_3^- levels are high enough, PEPC may generate more malate than the maintenance pathways can accommodate. In this scenario, the high malate concentration would typically feed back to inhibit PEPC[70]. However, if excess malate instead became sequestered in the vacuole or diffused to adjacent cells along a concentration gradient, this feedback would not occur, and a rudimentary CCM cycle could be established. Thus, the upregulation of a few key enzymes might be sufficient to initiate CCM evolution.

The non-photosynthetic roles of some CCM enzymes in C_3 species may also support the view that CCMs are evolutionarily accessible. For example, non-photosynthetic CCM pathways that act in a tissue-specific manner have been identified in several C_3 species: low-level nocturnal CO_2 fixation without malic acid accumulation occurs in tobacco leaves[71] and in cotton ovules[72], and a C_4-like mechanism is found in cells adjacent to

Box 1 | Refinement of a CCM pathway

Although all plant lineages possess the enzymatic machinery for a carbon concentrating mechanism (CCM), and some lineages appear to be potentiated for CCM evolution, phylogenetic analyses indicate that a subset of changes have occurred after the origin of the CCM pathway and are, therefore, likely to be involved in its refinement[68].

Refinement of C_4 pathways

Much more is known about when key post-origin CCM traits evolved in C_4 lineages than for crassulacean acid metabolism (CAM) lineages owing largely to more robust phylogenetic work in important C_4 lineages, predominantly the grasses.

Anatomical specialization. C_4 species arising from a single origin of the CCM show variation in anatomical traits important for C_4 function, such as vein spacing and bundle sheath cell size[19,106–108], which suggests that different lineages have different post-origin anatomical improvements for CCM activity.

Gene evolution. Although the core set of CCM enzymes is present in all plants, analyses of their amino acid sequence in C_4 lineages suggest they undergo improvement after the origin of the CCM. For example, some core CCM enzymes are known to vary in their amino acid sequence within C_4 lineages[98], molecular evolution studies indicate that certain residues are under positive selection within C_4 species[47,109,110], and convergent substitutions have been identified that are restricted to some C_4 species within each group[111–113]. Additionally, phylogenetic analyses indicate that a number of gene duplication events that allowed for enzyme adaptation occurred after a CCM origin[114].

Refinement of CAM pathways

Post-origin changes are thought to be required for the refinement of constitutive CAM activity, but such changes are not necessarily required in C_3+CAM species or in taxa that have recently evolved a CCM. However, the timing of these changes is largely unknown owing to a lack of systematic comparisons between C_3, C_3+CAM and taxa with strong CAM activity. Candidates for modifications that might occur after the origin of a CAM pathway include the following.

Stomatal regulation. Although some C_3 species are known to open their stomata at night[115,116], the inverse stomatal behaviour observed in CAM species (FIG. 1) likely requires large transcriptional reconfiguration. In particular, CAM species likely modify pathways involving abscisic acid and blue light sensing, two of the main pathways for stomatal signalling. Blue light in particular is a major signal for stomatal aperture, and some experiments on blue light sensitivity and transcriptomic analyses have suggested that guard cells in CAM species are less sensitive to blue light than guard cells in C_3 species[117–120].

Circadian regulation. Some CAM enzymes show clear circadian behaviour[121], and, indeed, integration of carbon fixation and metabolism with an endogenous clock is thought to be critical for efficient CAM function[121,122]. Comparative genomics between pineapple (*Ananas comosus*, Bromeliaceae) and unrelated C_3 and C_4 species showed an increase in clock-regulated regulatory motifs upstream of canonical CAM genes (FIG. 1c), indicating circadian control of gene expression[87]. Although it is likely that fine-scaled circadian control of CAM gene expression occurs only with or after the origin of CAM, further work is needed to show circadian regulation as a refinement of the CAM cycle.

the vasculature of tobacco stems, whereby photosynthetic cells use CO_2 that has been respired by roots and transported through the xylem[73]. Genes that give identity to root endodermal cells have also been shown to pleiotropically affect differentiation of BS cells in the leaf, which are important for C_4: SCARECROW and SHORTROOT mutants have deformed endodermal cell layers in the roots and a proliferation of BS cells in the leaves[74,75]. Perhaps most notably, *Camellia oleifera* (Theaceae, Ericales), a mesic forest tree, has been shown to respond to leaf fungal infection by strongly inducing tissue succulence and a fully functional CAM cycle[76]. The upregulation of a complete CAM cycle in a species so distantly related from any known CCM origin emphasizes both the ubiquitous presence of functional CCM enzymes in all plants and their frequent co-option into a variety of cellular roles.

Potentiation for CCM evolution

CCM origins seem to be phylogenetically clustered, with large regions of the land plant phylogeny completely lacking any known C_4 or CAM plants[23] (FIG. 2a). Thus, although all plants have the biochemical components for a CCM, some lineages may be more likely to evolve CCMs because of differential exposure to environmental pressures or the evolution of characteristics — either anatomical or genomic — that later facilitate the co-option of these enzymes into photosynthetic metabolic pathways.

Anatomical potentiation. The idea that some lineages may be potentiated (sensu Blount[77]) for CCM evolution has often been discussed[28], but primarily in the context of leaf anatomy. The spatial separation of carbon assimilation and fixation via Rubisco in C_4 plants is generally realized across distinct cell types[32,33] (FIG. 1). The typical C_4 anatomical traits, such as large BS:M cell volume ratio and small distance between veins, have evolved before the C_4 pathway in a number of lineages including grasses[19], *Flaveria*[18] and Cleomaceae[78], suggesting that C_4 origins are clustered in certain regions of the plant phylogeny[79] (FIG. 2a) because some lineages happened to first evolve a C_4-like anatomy. Anatomical potentiation could have arisen either from environmental pressures or stochastic processes; regardless, the establishment of a C_4-like anatomy then allowed for frequent transitions to C_4 photosynthesis within these lineages[64].

The phylogenetic distribution of CAM photosynthesis appears to be much less clustered, and the evidence of anatomical enabling is mixed[46,80] (FIG. 2a). Many known C_3+CAM species are only mildly succulent, whereas species with strong CAM activity typically have large, tightly packed cells and thick photosynthetic tissues[80,81]. In the Agavoideae, tissue succulence seemingly evolved before the emergence of strong CAM photosynthesis[82], but there are few studies that combine anatomy, CAM activity and a well-resolved and sampled phylogeny. Although more studies are needed, the lack of identifiable anatomical specialization in C_3+CAM plants suggests that leaf anatomical changes are likely necessary for the development of strong CAM activity but not for the evolution of a weak CAM cycle.

Genetic potentiation. Similar to environment and anatomy, genetic characteristics could also predispose certain lineages to evolve a CCM. In C_4 and CAM lineages, CCM enzymes need to be expressed at specific locations[83–85], times[86,87] and levels[88,89]. Non-CCM lineages that have regulatory pathways in place that could confer these properties on CCM enzymes might be predisposed to evolving CCMs. Indeed, genome-wide comparisons and transcription factor binding assays between C_3 and C_4 species have revealed shared regulatory motifs[90,91]. For example, a simple 59 bp deletion upstream of a gene encoding the P-subunit of glycine decarboxylase in the C_3 species *Arabidopsis thaliana* is sufficient to confer BS cell-specific expression, which is required for C_2 biochemistry[92]. Another *A. thaliana* gene, *NAD-me*, which encodes a CCM decarboxylating enzyme that also has roles in mitochondrial and chloroplast housekeeping, contains motifs within its coding sequence called duons that have regulatory functions. These sequences are also found in the C_4 species *G. gynandra*, in which they are necessary and sufficient for BS cell-specific expression[93]. Thus, these sequences confer C_4 activity through changes to gene regulation rather than via alterations to the coding sequence of the enzyme. Orthologues of C_4 genes in C_3 species have also been shown to be transcriptionally induced by light[84,94] and regulated by chloroplast-to-nucleus signalling[95], two characteristics of many enzymes in the C_4 pathway. The prevalence of shared regulatory DNA sequences and transcriptional cascades in C_3 and CCM species suggests that pre-existing regulatory mechanisms may facilitate the repeated evolution of C_4 photosynthesis in certain lineages.

Genetic potentiation can also occur in certain groups as a result of genome-wide patterns and processes. In some cases, the ability of plant genomes to undergo dramatic reconfiguration may have facilitated the evolution of a CCM. For example, in C_4 maize, hundreds of DNA motifs associated with C_4 have apparently been moved around the genome by transposons[96]. Lateral gene transfer has even played a role in CCM evolution in certain cases; genes optimized for C_4 function were laterally transferred to *A.* semialata from distantly related C_4 grasses and are preferentially expressed relative to the native gene copies[64]. Hybridization between C_3 and CAM species in *Yucca* gave rise to a new C_3+CAM species[67], and population gene flow between

Glossary

Anaplerotic
Chemical reactions that provide intermediates to various metabolic pathways, including the tricarboxylic acid (TCA) cycle.

Carboxylation
The addition of a carboxyl group to a substrate, often via a carboxylase enzyme.

Co-option
The recruitment of a gene, enzyme or other trait for an alternative function.

Decarboxylated
Pertaining to a molecule from which a carboxyl group has been removed by a decarboxylase enzyme in a process that releases CO_2.

Duons
Portions of the genome that both code for amino acids and provide motifs that can regulate gene expression.

Gene flow
Movement of genetic information between populations.

Genome-wide association studies
Analyses that correlate genetic markers from across the genome with a phenotype of interest in order to find loci underlying traits.

Lateral gene transfer
Movement of genes between individuals by mechanisms other than sexual reproduction.

Photorespiration
Fixation of oxygen by Rubisco, resulting in a loss of energy and a release of CO_2 but no net gain in carbohydrates.

Transpiration
The passive movement of water via stomata from the leaf intercellular airspace to the atmosphere along the water concentration gradient.

C_4 and non-C_4 genotypes in *A. semialata* facilitated movement of C_4-adapted alleles across the range of the species[62]. Additionally, in some CCM lineages, duplicated genes are associated with co-option into CCM pathways[97–100] and subsequent positive selection[101,102]. By contrast, other studies have indicated that recruitment of genes into a CCM was not immediately preceded by a gene duplication event but instead relied on differentiation of *cis*-regulatory elements in ancient paralogues[50,85,87]. The evidence from such genomic comparative studies suggests that re-wiring of transcriptional cascades — whether through gene duplications, genomic rearrangements or molecular evolution — is critical for CCM evolution.

Future perspectives

As sequencing costs decrease and protocols become more streamlined, gene expression data for non-model species are becoming increasingly available, which allows relative gene expression to be compared across hundreds or thousands of species in a single analysis. As a simple illustration of the types of studies we imagine will become increasingly common and insightful, we analysed the relative levels of expression of major PEPC gene copies in leaves of C_3 species across flowering plants (using 1KP data[103,104], Supplementary Methods, Supplementary Figure 1 and Supplementary Table 1). PEPC is the best studied enzyme in the CCM pathway and is an exceptional system to explore the mechanics of gene recruitment into a novel function. Most flowering plants have at least three main copies of genes encoding PEPC, with two of those copies arising from independent duplication events in eudicots and monocots[48,105] (Supplementary Figure 1). Studies in grasses and orchids have suggested that the copy that is most highly expressed in C_3 relatives is the one recruited into the CCM[88,89]. We specifically focused on C_3 species in flowering plant lineages that have evolved CCMs, in which the copy recruited into the CCM is known from empirical studies. Because we relied on transcriptomic data rather than genomic data, we summed the expression of multiple assembled transcripts per PEPC paralogue per species. Even so, in 12 of the 14 CCM-evolving lineages, the copy recruited into the CCM is more highly expressed than other PEPC copies in the C_3 taxa (FIG. 3). While the recruitment bias is likely driven by expression levels of the two copies of PEPC, the differences in the expression patterns could result from a variety of processes. Perhaps these copies are more highly expressed because of a dosage effect caused by local gene duplications (we did not quantify lineage-specific copy numbers per PEPC paralogue), which could then allow for subsequent specialization of paralogues[100]. Alternatively, the increased transcript abundance of one copy over the other might have been selected for because of a non-photosynthetic role of PEPC, which simultaneously made it amenable to co-option by a CCM. Although our analysis is based on a broadly sampled transcriptome data set that was not designed for this sort of question, the emergence of such a striking pattern reveals the great potential of comparative genomics to identify the genetic potentiation of CCM evolution. Although the use of existing data sets such as 1KP for analysis of CCM evolution offers exciting possibilities, better phylogenetic resolution of CCM-evolving lineages will continue to expand the availability of model systems that are particularly powerful for assessing initial genetic changes associated with CCM evolution.

Conclusions

In this Perspective, we have outlined evidence supporting the idea that relatively few modifications are required to reach a rudimentary CCM phenotype, at least in groups that are potentiated. If correct, this

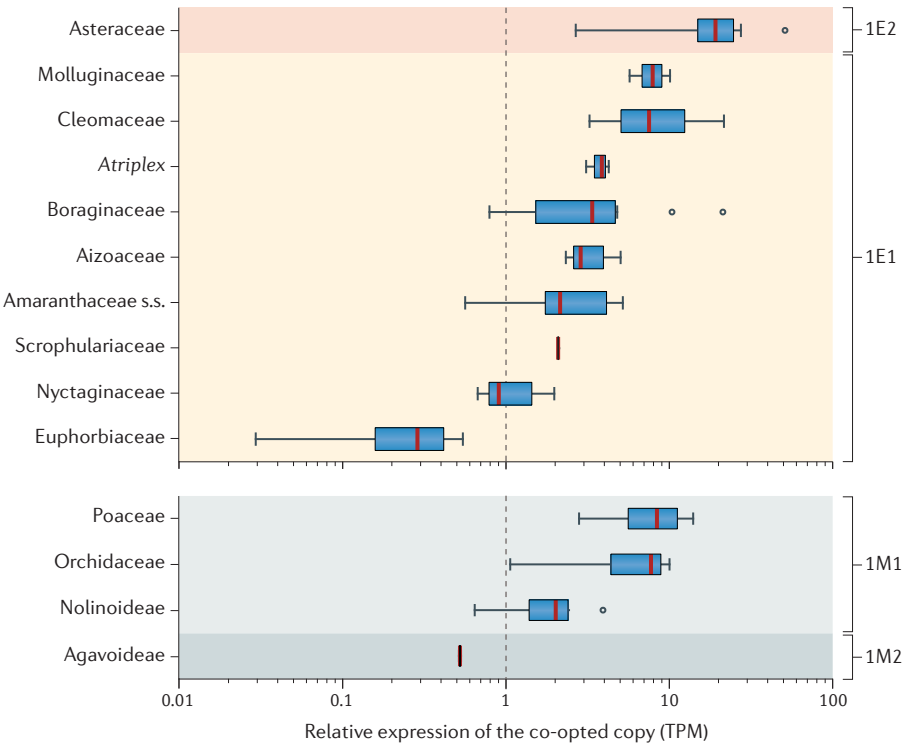


Fig. 3 | **Expression of PEPC in the leaves of C_3 angiosperms belonging to CCM-evolving lineages.** We used the 1KP[103] database to assess whether phosphoenolpyruvate carboxylase (PEPC) copies upregulated in C_3 species predict which copy is co-opted in C_4 or crassulacean acid metabolism (CAM) origins within the same lineage. Gene expression for PEPC was assessed per species for the major clades of the gene family: PPC-1E1 and PPC-1E2 in eudicots (top panel) and PPC-1M1 and PPC-1M2 in monocots (bottom panel) (Supplementary Figure 1). Expression was summed across all contigs from the transcriptomes of species that, based on phylogenetic analysis, were within each of the major clades (Supplementary Methods and Supplementary Table 1). Expression of PPC-2, a separate paralogue of PEPC, was not considered here, as it is not commonly recruited to a carbon concentrating mechanism (CCM). The background colour indicates which copy is used by the C_4 or CAM species within a given lineage, and the box plots represent the log of the ratio of expression (in transcript per million (TPM)) of the co-opted copy to the other copy for all species within that lineage. Clades for which the box plot falls to the right of the dashed line co-opt the more highly expressed copy of PEPC, whereas box plots to the left of the dashed line indicate clades that have co-opted the lesser expressed copy of PEPC. For example, C_4 Asteraceae species use PPC-1E2 in the CCM pathway, and their C_3 relatives have higher expression of PPC-1E2 relative to PPC-1E1. In the majority of lineages assessed, the co-opted copy is more highly expressed than the other copy in C_3 species, suggesting that this quality is important for recruitment into a CCM and may actually facilitate CCM evolution in certain lineages.

hypothesis would help to explain how such a seemingly complex trait has evolved so frequently. Once a rudimentary CCM is operating, selection to improve its efficiency will be strong because it relates to the primary metabolism of plants. Over time, this selection will lead to the accumulation of additional traits that further optimize the new pathway and that strongly contrast with those of their non-CCM relatives. Although these specializations are important for understanding the nature of extant CCMs, they lead to an overestimation of the initial changes required to evolve a CCM and, therefore, to an underestimation of its evolutionary accessibility.

We envision that future comparative genomic studies will continue to refine our understanding of CCM evolution in two important ways. First, the accumulation of genomic data from an increasingly phylogenetically diverse species pool will facilitate robust comparative analyses across multiple CCM origins — in other words, we imagine a proliferation of refined versions of the preliminary 1KP analyses we presented here. Second, we predict that new data and analytical tools in population genomics will enable a renaissance of intraspecific studies of CCM variation and will uncover additional *A. semialata*-like systems with which to explore the evolutionary emergence of a nascent CCM. Taken together, these efforts will disentangle the genetic changes that are required for refinement of CCM evolution from those responsible for the origin of a CCM pathway. From a rich history of biochemical and genetic studies, our knowledge of plant CCM evolution may already be far ahead of other classical examples of convergence. In time, we predict that a growth of genomic resources in non-model systems has the potential to provide a fairly complete explanation of how C_4 and CAM photosynthesis evolved so often in plants, and perhaps it is not so far away.

Karolina Heyduk[1] *, *Jose J. Moreno-Villena*[1], *Ian S. Gilman*[1], *Pascal-Antoine Christin*[2] *and Erika J. Edwards*[1]

[1]*Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, USA.*

[2]*Department of Animal and Plant Sciences, The University of Sheffield, Sheffield, UK.*

**e-mail: karolina.heyduk@yale.edu*

https://doi.org/10.1038/s41576-019-0107-5

Published online: 18 March 2019

1. Fernald, R. D. Casting a genetic light on the evolution of eyes. *Science* **313**, 1914–1918 (2006).
2. Fraser, J. A. et al. Convergent evolution of chromosomal sex-determining regions in the animal and fungal kingdoms. *PLOS Biol.* **2**, e384 (2004).
3. Berens, A. J., Hunt, J. H. & Toth, A. L. Comparative transcriptomics of convergent evolution: different genes but conserved pathways underlie caste phenotypes across lineages of eusocial insects. *Mol. Biol. Evol.* **32**, 690–703 (2015).
4. Smith, S. D. & Kriebel, R. Convergent evolution of floral shape tied to pollinator shifts in Iochrominae (Solanaceae). *Evolution* **72**, 688–697 (2018).
5. Larter, M. et al. Convergent evolution at the pathway level: predictable regulatory changes during flower color transitions. *Mol. Biol. Evol.* **35**, 2159–2169 (2018).
6. Knudsen, J. T. & Tollsten, L. Floral scent in bat-pollinated plants: a case of convergent evolution. *Bot. J. Linn. Soc.* **119**, 45–57 (1995).
7. Conn, C. E. et al. Convergent evolution of strigolactone perception enabled host detection in parasitic plants. *Science* **349**, 540–543 (2015).
8. Tcherkez, G. G. B., Farquhar, G. D. & Andrews, T. J. Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. *Proc. Natl Acad. Sci. USA* **103**, 7246–7251 (2006).
9. Bauwe, H., Hagemann, M. & Fernie, A. R. Photorespiration: players, partners and origin. *Trends Plant Sci.* **15**, 330–336 (2010).
10. Hatch, M. D. C_4 photosynthesis: a unique blend of modified biochemistry, anatomy and ultrastructure. *Biochim. Biophys. Acta* **895**, 81–106 (1987).
11. Osmond, C. B. Crassulacean acid metabolism: a curiosity in context. *Annu. Rev. Plant Physiol.* **29**, 379–414 (1978).
12. Christin, P.-A. et al. Oligocene CO_2 decline promoted C_4 photosynthesis in grasses. *Curr. Biol.* **18**, 37–43 (2008).
13. Arakaki, M. et al. Contemporaneous and recent radiations of the world's major succulent plant lineages. *Proc. Natl Acad. Sci. USA* **108**, 8379–8384 (2011).
14. Edwards, E. J. & Still, C. J. Climate, phylogeny and the ecological distribution of C_4 grasses. *Ecol. Lett.* **11**, 266–276 (2008).
15. Edwards, E. J. & Smith, S. A. Phylogenetic analyses reveal the shady history of C_4 grasses. *Proc. Natl Acad. Sci. USA* **107**, 2532–2537 (2010).
16. Horn, J. W. et al. Evolutionary bursts in Euphorbia (Euphorbiaceae) are linked with photosynthetic pathway. *Evolution* **68**, 3485–3504 (2014).
17. Crayn, D. M. et al. Photosynthetic pathways in Bromeliaceae: phylogenetic and ecological significance of CAM and C_3 based on carbon isotope ratios for 1893 species. *Bot. J. Linn. Soc.* **178**, 169–221 (2015).
18. McKown, A. D. & Dengler, N. G. Key innovations in the evolution of Kranz anatomy and C_4 vein pattern in Flaveria (Asteraceae). *Am. J. Bot.* **94**, 382–399 (2007).
19. Christin, P.-A. et al. Anatomical enablers and the evolution of C_4 photosynthesis in grasses. *Proc. Natl Acad. Sci. USA* **110**, 1381–1386 (2013).
20. Silvera, K. et al. Crassulacean acid metabolism and epiphytism linked to adaptive radiations in the Orchidaceae. *Plant Physiol.* **149**, 1838–1847 (2013).
21. Crayn, D. M., Winter, K. & Smith, J. A. C. Multiple origins of crassulacean acid metabolism and the epiphytic habit in the Neotropical family Bromeliaceae. *Proc. Natl Acad. Sci. USA* **101**, 3703–3708 (2004).
22. Sage, R. F., Christin, P.-A. & Edwards, E. J. The C_4 plant lineages of planet Earth. *J. Exp. Bot.* **62**, 3155–3169 (2011).
23. Edwards, E. J. & Ogburn, R. M. Angiosperm responses to a low-CO_2 world: CAM and C_4 photosynthesis as parallel evolutionary trajectories. *Int. J. Plant Sci.* **173**, 724–733 (2012).
24. Kellogg, E. A. in *C_4 Plant Biology* (eds Sage, R. F. & Monson, R. K.) 411–444 (Academic Press, 1999).
25. Sage, R. F. Environmental and evolutionary preconditions for the origin and diversification of the C_4 photosynthetic syndrome. *Plant Biol.* **3**, 202–213 (2001).
26. Smith, J. M. et al. Developmental constraints and evolution: a perspective from the Mountain Lake Conference on Development and Evolution. *Q. Rev. Biol.* **60**, 265–287 (1985).
27. Blount, Z. D., Lenski, R. E. & Losos, J. B. Contingency and determinism in evolution: replaying life's tape. *Science* **362**, eaam5979 (2018).
28. Jordan, D. B. & Ogren, W. L. The CO_2/O_2 specificity of ribulose 1,5-bisphosphate carboxylase/oxygenase. *Planta* **161**, 308–313 (1984).
29. Farquhar, G. D. & Sharkey, T. D. Stomatal conductance and photosynthesis. *Annu. Rev. Plant Physiol.* **33**, 317–345 (1982).
30. Mallmann, J. et al. The role of photorespiration during the evolution of C_4 photosynthesis in the genus *Flaveria*. *eLife* **3**, e02478 (2014).
31. Voznesenskaya, E. V. et al. Kranz anatomy is not essential for terrestrial C_4 plant photosynthesis. *Nature* **414**, 543–546 (2001).
32. Dengler, N. G. et al. Quantitative leaf anatomy of C_3 and C_4 grasses (Poaceae): bundle sheath and mesophyll surface area relationships. *Ann. Bot.* **73**, 241–255 (1994).
33. Hattersley, P. W. Characterization of C_4 type leaf anatomy in grasses (Poaceae). Mesophyll: bundle sheath area ratios. *Ann. Bot.* **53**, 163–180 (1984).
34. Hatch, M. D. in *C_4 Plant Biology* (eds Sage, R. F. & Monson, R. K.) 17–46 (Academic Press, 1999).
35. Furbank, R. T. & Hatch, M. D. Mechanism of C_4 photosynthesis: the size and composition of the inorganic carbon pool in bundle sheath cells. *Plant Physiol.* **85**, 958–964 (1987).
36. Christin, P.-A. et al. C_4 eudicots are not younger than C_4 monocots. *J. Exp. Bot.* **62**, 3171–3181 (2011).
37. Christin, P.-A. & Osborne, C. P. The recurrent assembly of C_4 photosynthesis, an evolutionary tale. *Photosynth. Res.* **117**, 163–175 (2013).
38. Long, S. P. in *C_4 Plant Biology* (eds Sage, R. F. & Monson, R. K.) 215–249 (Academic Press, 1999).
39. Cockburn, W., Ting, I. P. & Sternberg, L. O. Relationships between stomatal behavior and internal carbon dioxide concentration in crassulacean acid metabolism plants. *Plant Physiol.* **63**, 1029–1032 (1979).
40. Borland, A. M. et al. Exploiting the potential of plants with crassulacean acid metabolism for bioenergy production on marginal lands. *J. Exp. Bot.* **60**, 2879–2896 (2009).
41. Sage, R. F. et al. Some like it hot: the physiological ecology of C_4 plant evolution. *Oecologia* **187**, 941–966 (2018).
42. Heckmann, D. et al. Predicting C4 photosynthesis evolution: modular, individually adaptive steps on a Mount Fuji fitness landscape. *Cell* **153**, 1579–1588 (2013).
43. Winter, K., Garcia, M. & Holtum, J. A. M. On the nature of facultative and constitutive CAM: environmental and developmental control of CAM expression during early growth of *Clusia*, *Kalanchöe*, and *Opuntia*. *J. Exp. Bot.* **59**, 1829–1840 (2008).
44. Winter, K. & Holtum, J. A. M. Induction and reversal of crassulacean acid metabolism in *Calandrinia polyandra*: effects of soil moisture and nutrients. *Funct. Plant Biol.* **38**, 576–582 (2011).
45. Heyduk, K. et al. Shifts in gene expression profiles are associated with weak and strong crassulacean acid metabolism. *Am. J. Bot.* **105**, 587–601 (2018).
46. Silvera, K., Santiago, L. S. & Winter, K. Distribution of crassulacean acid metabolism in orchids of Panama: evidence of selection for weak and strong modes. *Funct. Plant Biol.* **32**, 397 (2005).
47. Goolsby, E. W. et al. Molecular evolution of key metabolic genes during transitions to C_4 and CAM photosynthesis. *Am. J. Bot.* **105**, 602–613 (2018).
48. Christin, P.-A. et al. Shared origins of a key enzyme during the evolution of C_4 and CAM metabolism. *J. Exp. Bot.* **65**, 3609–3621 (2014).
49. Paterson, A. H. et al. The Sorghum bicolor genome and the diversification of grasses. *Nature* **457**, 551–556 (2009).
50. Cai, J. et al. The genome sequence of the orchid *Phalaenopsis equestris*. *Nat. Genet.* **47**, 65–72 (2015).
51. Yang, X. et al. The Kalanchoë genome provides insights into convergent evolution and building blocks of crassulacean acid metabolism. *Nat. Commun.* **8**, 1899 (2017).
52. Goodstein, D. M. et al. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* **40**, D1178–D1186 (2012).
53. Ding, Z. et al. Identification of photosynthesis-associated C_4 candidate genes through comparative leaf gradient transcriptome in multiple lineages of C_3 and C_4 species. *PLOS ONE* **10**, e0140629 (2015).
54. Wang, L. et al. Comparative analyses of C_4 and C_3 photosynthesis in developing leaves of maize and rice. *Nat. Biotechnol.* **32**, 1158–1165 (2014).
55. Brilhaus, D. et al. Reversible burst of transcriptional changes during induction of crassulacean acid metabolism in *Talinum triangulare*. *Plant Physiol.* **170**, 102–122 (2016).
56. Xu, J. et al. Systems analysis of *cis*-regulatory motifs in C_4 photosynthesis genes using maize and rice leaf transcriptomic data during a process of de-etiolation. *J. Exp. Bot.* **67**, 5105–5117 (2016).
57. Gowik, U. et al. Evolution of C_4 photosynthesis in the genus *Flaveria*: how many and which genes does it take to make C_4? *Plant Cell* **23**, 2087–2105 (2011).

58. Bräutigam, A. et al. An mRNA blueprint for C_4 photosynthesis derived from comparative transcriptomics of closely related C_3 and C_4 species. *Plant Physiol.* **155**, 142–156 (2011).
59. Lauterbach, M. et al. De novo transcriptome assembly and comparison of C_3, C_3–C_4, and C_4 species of Tribe Salsoleae (Chenopodiaceae). *Front. Plant Sci.* **8**, 1939 (2017).
60. Ellis, R. P. Anomalous vascular bundle sheath structure in *Alloteropsis semialata* leaf blades. *Bothalia* **11**, 273–275 (1974).
61. Lundgren, M. R. et al. Evolutionary implications of C_3–C_4 intermediates in the grass *Alloteropsis semialata*. *Plant Cell Environ.* **39**, 1874–1885 (2016).
62. Olofsson, J. K. et al. Genome biogeography reveals the intraspecific spread of adaptive mutations for a complex trait. *Mol. Ecol.* **25**, 6107–6123 (2016).
63. Ueno, O. & Sentoku, N. Comparison of leaf structure and photosynthetic characteristics of C_3 and C_4 *Alloteropsis semialata* subspecies. *Plant Cell Environ.* **29**, 257–268 (2006).
64. Dunning, L. T. et al. Introgression and repeated co-option facilitated the recurrent emergence of C_4 photosynthesis among close relatives. *Evolution* **71**, 1541–1555 (2017).
65. Christin, P.-A. et al. Adaptive evolution of C_4 photosynthesis through recurrent lateral gene transfer. *Curr. Biol.* **22**, 445–449 (2012).
66. Sayre, R. T., Kennedy, R. A. & Pringnitz, D. J. Photosynthetic enzyme activities and localization in *Mollugo verticillata* populations differing in the levels of C_3 and C_4 cycle operation. *Plant Physiol.* **64**, 293–299 (1979).
67. Heyduk, K. et al. Gas exchange and leaf anatomy of a C_3-CAM hybrid, *Yucca gloriosa* (Asparagaceae). *J. Exp. Bot.* **67**, 1369–1379 (2016).
68. Reeves, G. et al. Natural variation within a species for traits underpinning C_4 photosynthesis. *Plant Physiol.* **177**, 504–512 (2018).
69. O'Leary, B., Park, J. & Plaxton, W. C. The remarkable diversity of plant PEPC (phosphoenolpyruvate carboxylase): recent insights into the physiological functions and post-translational controls of non-photosynthetic PEPCs. *Biochem. J.* **436**, 15–34 (2011).
70. Smith, F. A. & Raven, J. A. Intracellular PH and its regulation. *Annu. Rev. Plant Physiol.* **30**, 289–311 (1979).
71. Kunitake, G., Stitt, C. & Saltman, P. Dark fixation of CO_2 by tobacco leaves. *Plant Physiol.* **34**, 123–127 (1959).
72. Dhindsa, R. S., Beasley, C. A. & Ting, I. P. Osmoregulation in cotton fiber: accumulation of potassium and malate during growth. *Plant Physiol.* **56**, 394–398 (1975).
73. Hibberd, J. M. & Quick, W. P. Characteristics of C_4 photosynthesis in stems and petioles of C_3 flowering plants. *Nature* **415**, 451–454 (2002).
74. Cui, H., Kong, D., Liu, X. & Hao, Y. SCARECROW, SCR-LIKE 23 and SHORT-ROOT control bundle sheath cell fate and function in *Arabidopsis thaliana*. *Plant J.* 78, (319–327 (2014).
75. Slewinski, T. L. et al. Scarecrow plays a role in establishing Kranz anatomy in maize leaves. *Plant Cell Physiol.* **53**, 2030–2037 (2012).
76. Yuan, M. et al. A single leaf of *Camellia oleifera* has two types of carbon assimilation pathway, C_3 and crassulacean acid metabolism. *Tree Physiol.* **32**, 188–199 (2012).
77. Blount, Z. D., Borland, C. Z. & Lenski, R. E. Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **105**, 7899–7906 (2008).
78. Marshall, D. M. et al. Cleome, a genus closely related to *Arabidopsis*, contains species spanning a developmental progression from C_3 to C_4 photosynthesis: C_4 photosynthesis in *Cleome*. *Plant J.* **51**, 886–896 (2007).
79. Sage, R. F. A portrait of the C_4 photosynthetic family on the 50th anniversary of its discovery: species number, evolutionary lineages, and Hall of Fame. *J. Exp. Bot.* **67**, 4039–4056 (2016).
80. Males, J. Concerted anatomical change associated with crassulacean acid metabolism in the Bromeliaceae. *Funct. Plant Biol.* **45**, 681–695 (2018).
81. Zambrano, V. A. B. et al. Leaf anatomical traits which accommodate the facultative engagement of crassulacean acid metabolism in tropical trees of the genus *Clusia*. *J. Exp. Bot.* **65**, 3513–3523 (2014).
82. Heyduk, K. et al. Evolution of CAM anatomy predates the origins of crassulacean acid metabolism in the Agavoideae (Asparagaceae). *Mol. Phylogenet. Evol.* **105**, 102–113 (2016).
83. Gowik, U. et al. *cis*-Regulatory elements for mesophyll-specific gene expression in the C_4 plant *Flaveria trinervia*, the promoter of the C_4 phosphoenolpyruvate carboxylase gene. *Plant Cell* **16**, 1077–1090 (2004).
84. Matsuoka, M. et al. The promoters of two carboxylases in a C_4 plant (maize) direct cell-specific, light-regulated expression in a C_3 plant (rice). *Plant J.* **6**, 311–319 (1994).
85. Akyildiz, M. et al. Evolution and function of a *cis*-regulatory module for mesophyll-specific gene expression in the C_4 dicot *Flaveria trinervia*. *Plant Cell* **19**, 3391–3402 (2007).
86. Yin, H. et al. Diel rewiring and positive selection of ancient plant proteins enabled evolution of CAM photosynthesis in *Agave*. *BMC Genomics* **19**, 588 (2018).
87. Ming, R. et al. The pineapple genome and the evolution of CAM photosynthesis. *Nat. Genet.* **47**, 1435–1442 (2015).
88. Moreno-Villena, J. J. et al. Highly expressed genes are preferentially co-opted for C_4 photosynthesis. *Mol. Biol. Evol.* **35**, 94–106 (2018).
89. Silvera, K. et al. Multiple isoforms of phosphoenolpyruvate carboxylase in the Orchidaceae (subtribe Oncidiinae): implications for the evolution of crassulacean acid metabolism. *J. Exp. Bot.* **65**, 3623–3636 (2014).
90. Brown, N. J. et al. Independent and parallel recruitment of preexisting mechanisms underlying C_4 photosynthesis. *Science* **331**, 1436–1439 (2011).
91. Borba, A. R., Serra, T. S. & Gõrska, A. Synergistic binding of bHLH transcription factors to the promoter of the maize NADP-ME gene used in C_4 photosynthesis is based on an ancient code found in the ancestral C_3 state. *Mol. Biol.* **35**, 1690–1705 (2018).
92. Adwy, W., Laxa, M. & Peterhansel, C. A simple mechanism for the establishment of C_2-specific gene expression in Brassicaceae. *Plant J.* **84**, 1231–1238 (2015).
93. Reyna-Llorens, I. et al. Ancient duons may underpin spatial patterning of gene expression in C_4 leaves. *Proc. Natl Acad. Sci. USA* **115**, 1931–1936 (2018).
94. Kausch, A. P. et al. Mesophyll-specific, light and metabolic regulation of the C_4 PPCZm1 promoter in transgenic maize. *Plant Mol. Biol.* **45**, 1–15 (2001).
95. Burgess, S. J. et al. Ancestral light and chloroplast regulation form the foundations for C_4 gene expression. *Nat. Plants* **2**, 16161 (2016).
96. Cao, C. et al. Evidence for the role of transposons in the recruitment of *cis*-regulatory motifs during the evolution of C_4 photosynthesis. *BMC Genomics* **17**, 201 (2016).
97. Wang, X. et al. Comparative genomic analysis of C_4 photosynthetic pathway evolution in grasses. *Genome Biol.* **10**, R68 (2009).
98. Christin, P.-A. et al. Evolution of C_4 phosphoenolpyruvate carboxykinase in grasses, from genotype to phenotype. *Mol. Biol. Evol.* **26**, 357–365 (2009).
99. Wang, P. et al. Evolution of GOLDEN2-LIKE gene function in C_3 and C_4 plants. *Planta* **237**, 481–495 (2013).
100. Bianconi, M. E. et al. Gene duplication and dosage effects during the early emergence of C_4 photosynthesis in the grass genus *Alloteropsis*. *J. Exp. Bot.* **69**, 1967–1980 (2018).
101. Emms, D. M. et al. Independent and parallel evolution of new genes by gene duplication in two origins of C_4 photosynthesis provides new insight into the mechanism of phloem loading in C_4 species. *Mol. Biol. Evol.* **33**, 1796–1806 (2016).
102. Huang, P. et al. Cross species selection scans identify components of C_4 photosynthesis in the grasses. *J. Exp. Bot.* **68**, 127–135 (2017).
103. Matasci, N. et al. Data access for the 1,000 Plants (1KP) project. *Gigascience* **3**, 17 (2014).
104. Wickett, N. J. et al. Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proc. Natl Acad. Sci. USA* **111**, E4859–E4868 (2014).
105. Deng, H. et al. Evolutionary history of PEPC genes in green plants: implications for the evolution of CAM in orchids. *Mol. Phylogenet. Evol.* **94** (Suppl. B), 559–564 (2016).
106. Lundgren, M. R., Osborne, C. P. & Christin, P.-A. Deconstructing Kranz anatomy to understand C_4 evolution. *J. Exp. Bot.* **65**, 3357–3369 (2014).
107. Freitag, H. & Kadereit, G. C_3 and C_4 leaf anatomy types in Camphorosmeae (Camphorosmoideae, Chenopodiaceae). *Plant Syst. Evol.* **300**, 665–687 (2014).
108. Lundgren, M. R. et al. C_4 anatomy can evolve via a single developmental change. *Ecol. Lett.* **22**, 302–312 (2019).
109. Rosnow, J. J. et al. Kranz and single-cell forms of C_4 plants in the subfamily Suaedoideae show kinetic C_4 convergence for PEPC and Rubisco with divergent amino acid substitutions. *J. Exp. Bot.* **66**, 7347–7358 (2015).
110. Rosnow, J. J., Edwards, G. E. & Roalson, E. H. Positive selection of Kranz and non-Kranz C_4 phosphoenolpyruvate carboxylase amino acids in Suaedoideae (Chenopodiaceae). *J. Exp. Bot.* **65**, 3595–3607 (2014).
111. Besnard, G. et al. Phylogenomics of C_4 photosynthesis in sedges (Cyperaceae): multiple appearances and genetic convergence. *Mol. Biol. Evol.* **26**, 1909–1919 (2009).
112. Christin, P.-A. et al. Evolutionary switch and genetic convergence on *rbcL* following the evolution of C_4 photosynthesis. *Mol. Biol. Evol.* **25**, 2361–2368 (2008).
113. Christin, P.-A. et al. C_4 Photosynthesis evolved in grasses via parallel adaptive genetic changes. *Curr. Biol.* **17**, 1241–1247 (2007).
114. Besnard, G. et al. Herbarium genomics retraces the origins of C_4-specific carbonic anhydrase in Andropogoneae (Poaceae). *Bot. Lett.* **165**, 419–433 (2018).
115. Caird, M. A., Richards, J. H. & Donovan, L. A. Nighttime stomatal conductance and transpiration in C_3 and C_4 plants. *Plant Physiol.* **143**, 4–10 (2007).
116. Snyder, K. A., Richards, J. H. & Donovan, L. A. Night-time conductance in C_3 and C_4 species: do plants lose water at night? *J. Exp. Bot.* **54**, 861–865 (2003).
117. Tallman, G. et al. Induction of CAM in *Mesembryanthemum crystallinum* abolishes the stomatal response to blue light and light-dependent zeaxanthin formation in guard cell chloroplasts. *Plant Cell Physiol.* **38**, 236–242 (1997).
118. Lee, D. M. & Assmann, S. M. Stomatal responses to light in the facultative Crassulacean acid metabolism species, *Portulacaria afra*. *Physiol. Plant.* **85**, 35–42 (1992).
119. Abraham, P. E. et al. Transcript, protein and metabolite temporal dynamics in the CAM plant *Agave*. *Nat. Plants* **2**, 16178 (2016).
120. Heyduk, K. et al. Altered gene regulatory networks are associated with the transition from C_3 to Crassulacean acid metabolism in *Erycina* (Oncidiinae: Orchidaceae). *Front. Plant Sci.* **9**, 2000 (2019).
121. Hartwell, J. The co-ordination of central plant metabolism by the circadian clock. *Biochem. Soc. Trans.* **33**, 945–948 (2005).
122. Ceusters, J. et al. Light quality modulates metabolic synchronization over the diel phases of Crassulacean acid metabolism. *J. Exp. Bot.* **65**, 3705–3714 (2014).
123. Harris, L. W. & Davies, T. J. Data from: A complete fossil-calibrated phylogeny of seed plant families as a tool for comparative analyses: an example testing the 'time for speciation' hypothesis. *Knowledge Network for Biocomplexity* https://doi.org/10.5063/F13T9F5P (2016).

Acknowledgements
The authors thank members of the Edwards laboratory for their thoughtful discussions on this manuscript. K.H. is supported by a Donnelley Postdoctoral Fellowship through the Yale Institute of Biospheric Studies. Additional support came from US National Science Foundation awards DEB-1252901 and IOS-1754662 to E.J.E. P.-A.C. is supported by a Royal Society Research Fellowship (grant number URF120119).

Author contributions
The authors contributed equally to all aspects of the article.

Competing interests
The authors declare no competing interests.

Publisher's note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Reviewer information
Nature Reviews Genetics thanks R. VanBuren and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Supplementary information
Supplementary information is available for this paper at https://doi.org/10.1038/s41576-019-0107-5.