This is an author produced version of *The evolutionary ecology of C-4 plants*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/95833/

**Article:**

http://dx.doi.org/10.1111/nph.13033

This is the pre-peer reviewed version of the following article: Christin, P.-A. and Osborne, C. P. (2014), The evolutionary ecology of C4 plants. New Phytologist, 204: 765–781, which has been published in final form at https://dx.doi.org/10.1111/nph.13033. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving (http://olabout.wiley.com/WileyCDA/Section/id-820227.html).
The Evolutionary Ecology of C₄ Plants

Pascal-Antoine Christin¹, Colin P. Osborne¹

¹Department of Animal and Plant Sciences, University of Sheffield, Western Bank, Sheffield S10 2TN, UK.

Author for correspondence: Pascal-Antoine Christin; telephone +44-(0)114-222-0027; fax +44 114 222 0002; email: p.christin@sheffield.ac.uk

Total word count: 16135
Figures: 6 (1 to 5 in color)
Table: 1
Summary

C₄ photosynthesis is a physiological syndrome resulting from multiple anatomical and biochemical components, which function together to increase the CO₂ concentration around Rubisco and reduce photorespiration. It evolved independently multiple times and C₄ plants now dominate many biomes, especially in the tropics and subtropics. The C₄ syndrome comes in many flavours, with numerous phenotypic realizations of C₄ physiology and diverse ecological strategies. In this work, we analyse the events that happened in a C₃ context and enabled C₄ physiology in the descendants, those that generated the C₄ physiology, and those that happened in a C₄ background and opened novel ecological niches. Throughout the manuscript, we evaluate the biochemical and physiological evidence in a phylogenetic context, which demonstrates the importance of contingency in evolutionary trajectories and shows how these constrained the realized phenotype. We then discuss the physiological innovations that allowed C₄ plants to escape these constraints for two important dimensions of the ecological niche, growth rates and distribution along climatic gradients. This review shows that a comprehensive understanding of C₄ plant ecology can be achieved by accounting for evolutionary processes spread over million of years, including the ancestral condition, functional convergence via independent evolutionary trajectories, and physiological diversification.

Keywords: C₄ photosynthesis, physiology, evolution, ecological niche, co-option, contingency
Introduction

C₄ photosynthesis is a complex phenotype, formed from multiple anatomical and biochemical components that together increase the concentration of CO₂ around Rubisco (Hatch, 1987; Figure 1). This evolutionary innovation increases the carbon-fixation efficiency under all conditions that restrict CO₂ supply to Rubisco, and has its greatest effects at high light and temperature (Ehleringer & Bjorkman, 1977; Ehleringer, 1978; Ehleringer et al., 1991, 1997). However, the distributions of C₄ plants cannot be comprehensively explained by individual environmental variables, and C₄ species thrive across a diversity of habitats, ranging from the tropics to the boreal zone, from deserts to submerged conditions, from open grasslands to forest understoreys, and from nutrient-depleted to fertile soils. This ecological diversity results from the rich evolutionary history of this physiological trait, which evolved many times in distantly related groups (Sage et al., 2011).

Since its discovery in the 60s, C₄ photosynthesis has been the subject of many studies, from the fields of biochemistry, physiology, organismal biology, ecology and evolution (reviewed in Langdale, 2011). In the last fifteen years, our understanding of evolutionary aspects of C₄ photosynthesis has been boosted by the accumulation of molecular phylogenies, which have identified more than 62 monophyletic C₄ groups (e.g. Kellogg, 1999; GPWG, 2001; Giussani et al., 2003; Kadereit et al., 2003; McKown et al., 2005; Besnard et al., 2009; Sage et al., 2011; GPWGII, 2012). Phylogenetic trees allow us to disentangle the events that led to the evolution of C₄ physiology (McKown et al., 2007; Christin et al., 2011b, 2013b; Khoshravesh et al., 2012; Griffiths et al., 2013; Box 1), and the accumulated evidence shows that some C₄ constituents evolved in a C₃ context and enabled the transition to C₄ physiology via the gradual addition of other C₄ constituents (Sage, 2001, 2004; Christin & Osborne, 2013). The availability of robust and densely sampled phylogenetic trees has also revolutionized our understanding of C₄ ecology, with the possibility of dating C₄ origins and placing them on the geological timeline (e.g. Christin et al., 2008a; Vicentini et al., 2008; Kadereit et al., 2010), and the capacity to differentiate ecological properties that were inherited from C₃ ancestors from those that represent departures from ancestral conditions (e.g. Edwards et al., 2007, 2008; Osborne & Freckleton, 2009; Edwards & Smith, 2010; Taylor et al., 2010, 2012; Kadereit et al., 2012; Box 1).

In this review, we integrate knowledge acquired during the last 50 years and recent modelling efforts into a phylogenetic context, to infer the most plausible events occurring during the evolutionary transition from C₃ to C₄ photosynthesis, and discuss their physiological and ecological consequences. Throughout, we evaluate the evidence in the context of two non-mutually exclusive
hypotheses. First, that evolutionary trajectories towards novel traits cannot vary in any direction, but are highly constrained by the phenotype and genotype of the organism. Secondly, that evolutionary innovation unlocks new phenotypic opportunities for the organism and shifts the fundamental niche, by removing constraints on the trait space that can be occupied.

1. Which properties are common to all C₄ plants?

C₄ physiology
The main effect of C₄ photosynthesis is an elevated concentration of CO₂ relative to O₂ in the vicinity of Rubisco, increasing the ratio of carboxylation to oxygenation reactions catalyzed by the enzyme, and therefore lowering the rate of photorespiration (Chollet & Ogren, 1975; Hatch & Osmond, 1976). It also near-saturates Rubisco with its CO₂ substrate, which increases the rate of carbon assimilation per unit of Rubisco protein and gives the potential for very rapid photosynthetic rates under high light conditions (Schmitt & Edwards, 1981; Long, 1999). The ratio of oxygenation by Rubisco relative to carboxylation rises with temperature because the solubility of CO₂ decreases relative to O₂, and the specificity of Rubisco declines faster for CO₂ than O₂ (Long, 1991). At high temperatures and low CO₂, the C₄ cycle therefore increases the number of CO₂ molecules fixed per absorbed photon (quantum efficiency), but also per unit of Rubisco protein invested, and consequently improves the photosynthetic nitrogen-use efficiency (Ehleringer & Bjorkman, 1977; Brown, 1978; Skillman, 2008). However, the C₄ cycle consumes metabolic energy, and C₃ plants therefore retain a higher quantum efficiency when photorespiration is low, especially at low light and low temperature (Ehleringer & Bjorkman, 1977). These physiological properties are common to all C₄ plants. However, they emerge through a complex assemblage of anatomical and biochemical components. When investigating the evolution of C₄ photosynthesis, it is useful to distinguish phenotypic characters arising from individual developmental changes or biochemical reactions, from the functional properties that emerge through the coordinated action of several such characters (Table 1).

C₄ phenotypic functions
The C₄ syndrome is defined by the primary fixation of carbon by phosphoenolpyruvate carboxylase (PEPC) during the day and its refixation by Rubisco (Kellogg, 1999). These metabolic functions are achieved via the segregation of PEPC and Rubisco into two distinct compartments within the leaf, with the compartment containing Rubisco largely isolated from the external environment (Hatch & Osmond, 1976). In addition, a number of biochemical functions are required to sustain the C₄ cycle (Figure 1a): a) the action of carbonic anhydrase (CA) for converting CO₂ to HCO₃⁻, and its
fixation into organic acids by PEPC; b) a cascade to transform the oxaloacetate produced by PEPC into other C₄ organic acids, and transport them to the Rubisco compartment; c) a system to release CO₂ in the Rubisco compartment; and d) a cascade to regenerate the acceptor molecules for carbon in the C₄ cycle (Hatch, 1987). Besides these biochemical functions, the fixation of carbon by PEPC and its later refixation by Rubisco requires a series of functions linked to the plant structure that are present in all C₄ plants (Hattersley & Watson, 1975; Edwards & Voznesenskaya, 2011; Lundgren et al., 2014). These include two compartments separated by a short distance, into which PEPC and Rubisco reactions can be segregated (Figure 1).

C₄ characters

The anatomical and metabolic functions listed above are present in all C₄ plants, independently of their taxonomic origin, but each of these functions arises from multiple characters, which result from independent modifications in the characteristics of their components (Table 1). Unlike the functions generated, these underlying characters and characteristics vary among C₄ lineages, and each time the C₄ syndrome evolved, it was assembled using one of numerous possible sets of anatomical and biochemical characters (Sinha & Kellogg, 1996; Kellogg, 1999). This leads to a number of important distinctions among C₄ lineages. First, the two compartments used to segregate PEPC and Rubisco reactions vary among C₄ plants, and may be cell types derived from the same or different meristematic tissues, or even different compartments within the same cell (Brown, 1975; Dengler et al., 1985; Edwards et al., 2004). Similarly, the close contact between the PEPC and Rubisco compartments can be achieved by modifying the vein architecture through different developmental pathways (reviewed by Lundgren et al., 2014). The biochemical cascade that transforms and transports the product of PEPC, releases CO₂ and regenerates the intermediate compounds (Figure 1a, steps b-d), is also well known to vary among C₄ lineages, with different enzymes involved, especially in the release of CO₂ from C₄ acids in the Rubisco compartment (Figure 1a, step c; Andrews et al., 1971; Gutierrez et al., 1974). In conclusion, the phenotypic characters that are known to be common to all C₄ plants are a high activity of CA and PEPC in the cytosol of the first compartment and a high activity of Rubisco within chloroplasts in the second compartment (Figure 1), and most, if not all, of the others vary among C₄ lineages (Kellogg, 1999).

2. What is unique to C₄ plants?

Individual C₄ components in non-C₄ plants

The emergent physiological properties associated with the C₄ syndrome are unique to C₄ plants, but several of the underlying functions and all the components can be found in plants using other
photosynthetic pathways. Close contact between the two leaf compartments usually used for PEPC and Rubisco reactions is found in several C₃ grasses (Lundgren et al., 2014), and in many plants that use a C₂ pathway, a low efficiency CO₂-scavenging mechanism based on glycine decarboxylase localization (Sage et al., 2012). Similarly, a concentration of Rubisco in bundle sheath chloroplasts is observed in C₂ plants as well as closely related C₃ taxa (Sage et al., 2013). The biochemical functions that generate the C₄ cycle are not found as such in other plants, except for CAM plants, which use a similar pathway with a temporal segregation of reactions. However, all the enzymes of the C₄ cycle, and the catalyzed reactions, exist in C₃ plants (Aubry et al., 2011). In these species, the enzymes are responsible for different functions in basal metabolism (reviewed by Aubry et al., 2011). Most of these enzymes are encoded by multigene families, and the different isoforms vary in their catalytic properties and expression patterns (Tausta et al., 2002; Svensson et al., 2003). The ancestral functions generally still exist in C₄ plants, but some isoforms now operate in the C₄ cycle, which requires specific spatial and temporal regulation, as well as specific kinetic properties. At least some of these expression and kinetic characteristics however exist in C₃ plants. For instance, decarboxylating enzymes are active around the vascular tissue in a phylogenetically diverse range of C₃ species (Hibberd and Quick, 2002; Osborne & Beerling, 2006; Brown et al., 2010), and most of the genes for the enzymes of the C₄ cycle can be found at significant levels in C₃ leaves (Christin et al., 2013a; Bräutigam et al., 2014).

Gradual C₄ assembly through repeated co-option of components

All of the components that together generate C₄ physiology can therefore be found in other photosynthetic types, but their characteristics vary both quantitatively and qualitatively, and C₄ lineages each present unique combinations of the resulting characters (Table 1). The presence of all components in C₃ or C₂ species implies that the evolution of C₄ photosynthesis required their co-option into a new function and, in many cases, their adaptation for the novel metabolic context. The different C₄ components were not co-opted simultaneously, but must have been added sequentially. The exact order of this process is still to be elucidated and is very likely to vary among lineages (Williams et al., 2013), but recent insights have come from phylogenetic reconstructions (e.g. Christin et al., 2011b; Khoshravesh et al., 2012) and modelling efforts (Heckmann et al., 2013; Williams et al., 2013; Mallmann et al., 2014). These studies differ in the characters that are considered, sometimes modelling the whole C₄ cycle as a simple component (Heckmann et al., 2013) or transforming quantitative traits into discrete binary variables (Christin et al., 2011b; Williams et al., 2013), but they all converge on similar conclusions. For instance, it is now widely accepted that several C₄ characters, especially anatomical ones, were acquired before C₄ physiology (Sage, 2001, 2004; McKown et al., 2007; Christin et al., 2011b; Khoshravesh et al., 2012;
Heckmann et al., 2013; Williams et al., 2013). Similarly, several C₄ characters were probably acquired once plants were already fixing the majority of their carbon via PEPC, thereby optimizing the syndrome and adapting it to diverse environments (Christin et al., 2011b; Heckmann et al., 2013). The whole history of events that led to optimized C₄ descendants was likely spread over many million years (Christin & Osborne, 2013; Figure 2), and the ecological drivers and biological consequences are likely to differ among these events. In the following sections, we discuss first the events that happened in a non-C₄ context and enabled the transition to C₄ physiology (previously referred to as preconditions; Sage, 2001, 2004), then the process that generated the C₄ physiology itself, and finally the modifications that likely happened within a C₄ context. For each of these, the potential physiological and ecological consequences are discussed.

3. What happened before C₄ physiology?

**Origin of enzymes of the C₄ pathway**

All enzymes of the C₄ pathway originated in bacteria, hundreds of millions or billions of years before they were co-opted for C₄ photosynthesis. In angiosperms, they are usually encoded by gene families, with multiple isogenes that appeared through successive whole genome or single gene duplications (Wang et al., 2009; Christin et al., 2013a). The different isoforms generally diversified and came to fulfil a variety of functions, mostly anaplerotic (Drincovich et al., 2001; Lepiniec et al., 2003). This diversification also involved changes in expression patterns (spatial, temporal, and quantitative), as well as kinetic properties and responses to regulators (e.g. Blasing et al., 2002; Tausta et al., 2002; Christin et al., 2013a; John et al., 2014). This functional diversification was not driven by C₄ photosynthesis, but might have predisposed some plants for a later C₃-to-C₄ transition. Indeed, a function in the C₄ cycle requires specific expression patterns as well as catalytic properties (Hibberd & Covshoff, 2010), and the existence in some genomes of genes encoding enzymes with characteristics partially suitable for the C₄ cycle might have facilitated C₄ evolution. This hypothesis is supported by the observation that independent C₄ origins preferentially co-opted specific isogenes, suggesting that these were more suitable for a function in C₄ photosynthesis (Christin et al., 2013a; John et al., 2014). It has been shown that some C₃ plants possess isoforms with C₄-like expression patterns (Hibberd & Quick, 2002; Brown et al., 2010). For instance, genes for bundle sheath-specific glycine decarboxylase were already present in the C₃ ancestors of the genus Flaveria (Schulze et al., 2013), and mechanisms for the cell specificity of NAD-ME and NADP-ME enzymes might have evolved long before the C₄ pathway (Brown et al., 2011). While the drivers of these characters remain to be elucidated, their co-option would drastically reduce the number of steps separating C₃ ancestors from C₄ descendants.
Evolution of $C_4$-like anatomical characters

In most $C_4$ lineages, PEPC and Rubisco functions are segregated within leaves into mesophyll and bundle-sheath cells, respectively (Figure 1b), the latter being specialized cells surrounding the vascular tissue. In this common variant of the $C_4$ syndrome, a short distance between mesophyll and bundle sheath cells is usually achieved via high vein density. Vein density first increased during the early diversification of angiosperms (Feild et al., 2011), and was followed by several further increases in diverse groups of $C_3$ plants (Figure 2; Christin et al., 2013b). In a $C_3$ context, a high density of major veins provides alternative paths for water transport in case of xylem embolism and might confer higher tolerance to damage and drought (Sack et al., 2008, 2012). In addition, higher densities of minor veins enable high rates of photosynthesis and are advantageous in productive environments, such as high irradiance conditions (McKown et al., 2010). High vein density therefore represents an adaptation to high photosynthetic rates or a high risk of xylem embolism or damage. However, vein density is only indirectly relevant to $C_4$ photosynthesis. Indeed, the absolute distance between veins (interveinal distance; IVD) is less important than the number of mesophyll cells separating consecutive vascular bundles (Hattersley & Watson, 1975). This latter characteristic is only partially correlated to IVD, which is also influenced by the size of mesophyll cells, the thickness of the bundle sheath, and the diameter of vascular tissue. Similar IVD values can therefore emerge through different combinations of mesophyll cell size and number (Lundgren et al., 2014), and the environmental drivers of these cellular properties are yet to be identified.

Bundle-sheath cells evolved early in the history of vascular plants, with the function of regulating water and metabolite fluxes from and into the leaves, and a variety of additional metabolic tasks (Leegood, 2008; Griffiths et al., 2013; Aubry et al., 2014). The ecological significance of bundle-sheath cell size is still unclear, although it has been proposed that larger cells might provide protection against or rapid repair of cavitation (Sage, 2001; Griffiths et al., 2013), and hence confer an advantage when transpiration exceeds water supply (Osborne & Sack, 2012). However, $C_4$ photosynthesis does not necessarily require large bundle-sheath cells, but only a large relative amount of bundle-sheath tissue (Hattersley, 1984; Dengler et al., 1994), which may be achieved via a proliferation of small bundle-sheath cells, for instance through the development of abundant minor veins (Lundgren et al., 2014). The proportion of bundle-sheath tissue varies among clades of $C_3$ grasses, with large fractions increasing the likelihood of evolving $C_4$ physiology (Christin et al., 2013b; Griffiths et al., 2013). Since this leaf property results from multiple characteristics of distinct components, and in particular the size of bundle-sheath cells and the number of mesophyll cells between consecutive vascular bundles (Christin et al., 2013b), it could be dictated by multiple
drivers, including those that influence vein density.

Concentration of Rubisco activity in bundle-sheath cells and the C₂ pathway

A high Rubisco activity in chloroplasts of the bundle-sheath is probably necessary for the evolution of C₄ photosynthesis, since any C₄ cycle in its absence would be futile. Determinants of the relative abundance of chloroplasts among mesophyll and bundle-sheath cells are poorly understood. However, it has been clearly established that enhanced Rubisco activity in the bundle-sheath can be related to the C₂ pathway (Sage et al., 2012). The C₂ cycle arises through a concentration in the bundle sheath of glycine decarboxylase (GDC), the enzyme responsible for CO₂-liberation in photorespiration (Sage et al., 2012). In Flaveria species, mesophyll and bundle-sheath GDC are encoded by different isogenes, so that a decrease of GDC expression in the mesophyll increases the relative activity of GDC in the bundle-sheath (Schulze et al., 2013). This localization forces photorespiration to release CO₂ in the bundle-sheath cells, meaning that the CO₂ is less likely to diffuse back to the atmosphere before being refixed by Rubisco (Sage et al., 2012). The rate of refixation is higher if Rubisco is abundant in the bundle-sheath cells, and an increased confinement of Rubisco and GDC activities to these cells might co-evolve to optimize the C₂ physiology.

The C₂ pathway has been seen as an intermediate stage between C₃ and C₄ photosynthesis for a long time (Monson et al., 1984; Hylton et al., 1988), a hypothesis later supported by phylogenetic analyses in different taxonomic groups (McKown et al., 2005; Khoshravesh et al., 2012). However, phylogenetic analyses and molecular dating have also shown that the C₂ trait can be stable, having existed in some groups for more than 10 million years without producing any known C₄ descendant (Christin et al., 2011a). Although most plants using the C₂ pathway are limited in range (Sudderth et al., 2009), others, like Mollugo verticillata, are widespread and colonize numerous ecological conditions. While some C₂ plants possess C₄-like biochemical characters (e.g. Mollugo verticillata; Kennedy & Laetsch, 1974), others, such as Mollugo nudicaulis, have no C₄ activity (Kennedy et al., 1980), which shows that C₂ physiology can evolve and be maintained independently of any C₄ cycle. The main physiological effect of the C₂ pathway is to slightly decrease photorespiration, and consequently increase the net carbon gain in conditions where photorespiration is important (Vogan & Sage, 2011; Way et al., 2014).

Selective pressures

The assembly of C₄ physiology via natural selection requires environmental conditions where C₄ photosynthesis is advantageous compared to the ancestral conditions. This is believed to have happened after atmospheric CO₂ reached very low levels some 30 million years ago during the
Oligocene (Pagani et al., 2005; Beerling & Royer, 2011), which exacerbated photorespiration (Ehleringer et al., 1991). Molecular dating places C₄ origins in the last 30 million years (Box 1; Figure 2), and phylogeny-based models have shown that the probability of C₃-to-C₄ transition increased during this time (Christin et al., 2008a, 2011a; Vicentini et al., 2008; Besnard et al., 2009). However, depending on the taxonomic/phylogenetic placement of some microfossils, the earliest C₄ origin, in the grass subfamily Chloridoideae, might have happened in a high-CO₂ world (Prasad et al., 2011; Christin et al., 2014), and fossilized pollen grains from a couple of million years before the Oligocene CO₂ decline have been assigned to C₄ species (Urban et al., 2010). Despite this possibility of some C₄ origins before the Oligocene CO₂ decline, the vast majority of C₄ origins happened in a low-CO₂ world. (Christin et al., 2014). However, a low atmospheric CO₂ level is not sufficient to select for C₄ photosynthesis (Ehleringer and Bjorkman, 1977; Osborne and Beerling, 2006), and other environmental factors that increase photorespiration likely promoted each of the numerous origins of C₄ physiology (Sage, 2001; Roalson, 2008). Comparative analyses have shown that transitions to C₄ physiology occurred in grass lineages from open habitats of warm regions (Osborne & Freckleton, 2009; Edwards & Smith, 2010), while in Chenopodiaceae sensu stricto, the evolution of C₄ photosynthesis was more likely in lineages inhabiting saline and coastal environments (Kadereit et al., 2012).

4. What happened during the transition to C₄ photosynthesis?

Increase of PEPC activity and new selective pressures

If the appropriate leaf functions are in place and a significant fraction of Rubisco activity is concentrated in the BS cells, the C₄ cycle can theoretically evolve through the gradual increase of C₄ reactions (Heckmann et al., 2013). The order in which the C₄ enzymes are incorporated is not known with precision, and the order might differ among lineages (Williams et al., 2013). An increase in the rate of transformation and transport of the C₄ intermediates, release of CO₂, or regeneration of the intermediates would not generate any kind of C₄ cycle in the absence of a sufficiently high concentration of oxaloacetate, the product of the PEPC reaction (Figure 1b). An increased activity of the other enzymes could however evolve before enhanced PEPC activity for reasons unrelated to C₄ photosynthesis (Williams et al., 2013; Mallmann et al., 2014). The very first step in the establishment of a proper C₄ cycle must be an increase in the rate of fixation of atmospheric CO₂ by the coupled action of PEPC and CA. CA is already present at high levels in many C₃ plants, where it plays a role in carbon assimilation (Majeau & Coleman, 1994). An increase of PEPC activity in the mesophyll might thus be sufficient to generate high concentrations of oxaloacetate. This oxaloacetate would however need to be transformed and transported by
several enzymes before feeding Rubisco with released CO2. It has been established that at least some enzymes of the C4 cycle are already present in some C3 plants in the areas of the leaf required for a C4 cycle (Hibberd & Quick, 2002). Their expression levels in C3 plants can moreover be significant, although below those observed in C4 plants (Christin et al., 2013a; Bräutigam et al., 2014). Furthermore, the activities of PPDK and decarboxylating enzymes increase in some C2 plants before PEPC (Williams et al., 2013), potentially to rebalance nitrogen metabolism in C2 plants (Mallmann et al., 2014). The enzymes already present in the cells of some C3 or C2 species may be sufficient to process the oxaloacetate produced by an increased PEPC activity, especially if their activity is induced by an increase in substrate concentrations. Transfer of intermediates between cells could initially be made via simple diffusion, so that increased PEPC activity might, in plants already possessing C4-like characters, be sufficient to generate a C4 physiology.

The establishment of a weak C4 cycle through an increased activity of PEPC and the co-option of other enzymes is a key event, because it can significantly decrease photorespiration and consequently lead to a gradual improvement of the efficiency of the C4 pathway through natural selection (Heckmann et al., 2013), fixing mutations that enhance activities of C4 enzymes and adapt their catalytic properties for the new metabolic context (Nakamoto et al., 1983; Bauwe, 1984; Svensson et al., 2003). In the case of PEPC, the past action of selection left traces as an excess of non-synonymous mutations that are mostly concentrated on branches leading to each C4 group (Christin et al., 2007; Besnard et al., 2009). This distribution of C4-driven amino acid changes suggests that the adaptation of PEPC for the C4 function occurred over a short time that overlaps with changes in the enzyme’s activity (Figure 2). In most phylogenies, the first C4 descendant is separated from its last C3 ancestor by several million years (Christin et al., 2008a, 2011a; Besnard et al., 2009), so that the different characters that together generate C4 physiology cannot be disentangled. However, some exceptional groups maintained a diversity of photosynthetic phenotypes that might represent the footprint of gradual modifications during the evolution of C4 physiology.

Insights from Flaveria

In the genus Flaveria, the transition from the last C3 ancestor to the first C4 descendant spanned about 2-3 million years (Christin et al., 2011a), and extant taxa represent a range of anatomical, biochemical and physiological states (Bauwe, 1984; Ku et al., 1991; McKown & Dengler, 2007; Sudderth et al., 2007; Vogan & Sage 2011). We compiled data from the literature for different C4-related traits and reconstructed their evolution on the time-calibrated phylogeny for the genus (from Christin et al., 2011a). Ancestral reconstructions for nodes separating the C3 ancestor of all Flaveria
from the extant C₄ species *Flaveria trinervia* suggest that C₄ anatomy, biochemistry and physiology were acquired in parallel in this group (Figure 3), although ancestral reconstructions come with large confidence intervals. A higher PEPC activity can be observed in some *Flaveria* species that do not have a typical C₄ metabolism (Bauwe, 1984), as also shown for other groups (Murphy *et al.*, 2007), and this results in an increase in the proportion of carbon fixed as C₄ acids (Monson *et al.*, 1986; Moore *et al.*, 1987; Vogan & Sage, 2011). The increased C₄ activity in these plants might result from a need to rebalance the nitrogen metabolism between bundle sheath and mesophyll cells, putting some C₂ plants on a highway towards C₄ (Mallmann *et al.*, 2014). An effect of this enhancement of C₄ activity on water-use efficiency has not been detected (Vogan & Sage, 2011).

There are, however, indications of a rise in photosynthetic nitrogen-use efficiency (PNUE) in parallel with the enhancement of C₄ activity in *Flaveria*, associated with the clear decrease in CO₂ compensation point that accompanies the accumulation of C₄ functions (Vogan & Sage, 2011; Figure 3).

The C₄ characters that accumulated before the transition to a C₄ physiology are likely to vary among taxonomic groups (Williams *et al.*, 2013). The increase of PEPC activity might happen in plants that already have C₄ functional properties, but the establishment of a weak C₄ cycle might also be possible in plants with components that are more distant from the C₄ requirements. In the former case, few changes might be needed besides the increase in C₄ cycle activity, while in the latter case C₄ functions would be reinforced by selection for a more efficient C₄ cycle, as seen for leaf anatomical characteristics in *Flaveria* (Figure 3). The changes required in both expression patterns and catalytic properties will also depend on the properties of the enzyme inherited from the C₃ ancestor and co-opted for the C₄ cycle. The timing of origin for C₄ characters will consequently vary among C₄ lineages (Williams *et al.*, 2013), with the same changes happening in some cases within a C₃ context, while in other lineages they might happen during the evolution of a C₄ physiology, or even slightly later.

5. **What happened after C₄ evolution?**

*Optimization of Rubisco and PNUE*

The relative specificity of Rubisco for CO₂ compared to O₂ is negatively correlated with its catalytic efficiency, and the two parameters are thought to be finely tuned to allow the highest catalytic rate while minimizing O₂ fixation (Tcherkez *et al.*, 2006). In C₃ plants and a low-CO₂ atmosphere, this trade-off results in more specific but slower enzymes that have to be highly expressed to fix sufficient CO₂, and Rubisco represents up to one third of all leaf soluble proteins and 20% of the
total nitrogen budget (Evans & Poorter, 2001). The higher concentration of CO\textsubscript{2} around Rubisco generated by C\textsubscript{4} physiology relaxed selection for enzymes with a higher specificity for CO\textsubscript{2}, and enabled the evolution of faster Rubiscos (Seeman et al., 1984; Tcherkez et al. 2006; Kubien et al. 2008; Kapralov et al. 2011). A more efficient enzyme, together with increased CO\textsubscript{2} concentrations at its active site, means that fewer protein molecules are needed, and the abundance of Rubisco is reduced by 60-80% in some C\textsubscript{4} species (Ku et al., 1979). Although the C\textsubscript{4} cycle itself requires additional enzymes, large quantities of proteins are not necessary if their catalytic rates are high, and the C\textsubscript{4} cycle thus allows for lower total protein and nitrogen amounts if the proteins are optimized, which increases photosynthetic nitrogen-use efficiency (PNUE; Schmitt & Edwards 1981; Sage & Pearcy 1987; Ghannoum et al., 2005).

Models suggest that the adaptation of Rubisco kinetics started in parallel with increased C\textsubscript{4} enzyme activity, but continued once the plants were in a C\textsubscript{4} physiological state (Heckmann et al., 2013; Williams et al., 2013). In Flaveria, the Rubisco kinetics of C\textsubscript{4} species differ from those of related C\textsubscript{3} taxa, but those of C\textsubscript{3} and intermediate taxa were not consistently different (Kubien et al., 2008). The continuous adaptation of Rubisco after C\textsubscript{4} evolution is supported by the footprint of adaptive evolution on genes encoding Rubisco, with an excess of non-synonymous mutations spread across branches within C\textsubscript{4} lineages in various groups of angiosperms (Christin et al., 2008b; Kapralov et al., 2012). The decreased nitrogen costs of Rubisco thus evolved very gradually, and continued long after the initial diversification of C\textsubscript{4} groups. The ranges of Rubisco kinetics almost overlap between C\textsubscript{3} and C\textsubscript{4} species (Seeman et al., 1984), and variation in the catalytic rate of Rubisco affects PNUE among C\textsubscript{4} grasses, with higher catalytic rates increasing PNUE (Ghannoum et al., 2005). For instance, the PNUE increase in C\textsubscript{4} lineages compared to C\textsubscript{3} sister-groups varies from 25% in the C\textsubscript{4} grass lineage Aristida to 42% in Chloridoideae and 60% in Andropogoneae (Taylor et al., 2010).

The capacity to grow with limited access to nitrogen is key to ecological success on infertile soils, and a more efficient use of nitrogen acquired during the diversification of C\textsubscript{4} lineages might have contributed to the rise to ecological dominance of some C\textsubscript{4} species (Edwards et al., 2010). For example, recovery after fire in mesic savannas requires rapid resprouting in a nitrogen-depleted soil, and these environments are dominated by grasses from the Andropogoneae clade (Forrestel et al., 2014), which have the highest PNUE values among C\textsubscript{4} grasses (Taylor et al., 2010). The number of species for which PNUE has been measured is limited, and it is thus not known whether the evolution of high PNUE coincided with the rise to ecological dominance better than the origin of C\textsubscript{4} photosynthesis. It is however likely that C\textsubscript{4} physiology enabled the evolution of very high PNUE in some cases, and hence the colonization of competitive habitats, like savannas.
Adaptation of stomatal conductance and plant hydraulics

CO₂ partial pressures within the leaf intercellular air spaces are sufficient to saturate the coupled CA-PEPC enzyme system at 25-33% of the atmospheric value, and maximum rates of C₄ photosynthesis can thus be maintained despite large decreases in stomatal conductance (Wong et al., 1979; Long, 1999). C₄ plants consequently evolved lower stomatal conductance for a given rate of photosynthesis, a property that is amongst the most consistently associated with C₄ photosynthesis in grasses (Taylor et al., 2010). Decreased stomatal conductance could theoretically arise directly from the emergence of a C₄ cycle if stomatal aperture is regulated in response to the intercellular CO₂ partial pressure and photosynthetic rate (e.g. Messinger et al., 2006). Changes in the stomatal response to internal CO₂ concentrations are already visible in some C₃–C₄ species of Flaveria (Huxman & Monson, 2003), but in the longer term, the maximum capacity for stomatal conductance is adjusted downwards via developmental changes in the density and/or size of the stomata (Taylor et al., 2012). The diversity of strategies used to decrease stomatal conductance within some C₄ grass lineages (i.e. smaller versus less numerous stomata; Taylor et al., 2012) suggests continuing adjustments after the emergence of a C₄ cycle, although an initial decrease of stomatal number might result directly from the elevated vein density in C₄ species (Way, 2012; Figure 3).

A lower stomatal conductance decreases leaf transpiration relative to hydraulic supply, thereby improving leaf water status if the hydraulic system remains unchanged (Osborne & Sack, 2012). This effect remains if any subsequent reduction in hydraulic conductance is of a smaller magnitude than the change in stomatal conductance. In keeping with this expectation, comparisons within common garden, glasshouse and controlled environments show that soil-leaf water potential gradients are smaller in C₄ grass lineages compared to their close C₃ relatives under well-watered conditions (Taylor et al., 2010, 2011, 2014). This can be advantageous in environments where evaporative demand exceeds hydraulic supply, including conditions of high evaporative potential where solar radiation is high or the atmosphere is dry (Osborne & Sack, 2012). The advantage of reducing stomatal conductance is greater in low CO₂ atmospheres, where the stomatal aperture of both C₃ and C₄ species tend to increase, thereby augmenting the risk of hydraulic failure (Osborne & Sack, 2012).

The effects of stomatal conductance on plant tolerance of water deficits are complex (Ghannoun, 2009). During the initial stages of soil drying, stomatal conductance decreases more sensitively in C₃ than C₄ grasses (Ripley et al., 2010; Taylor et al., 2011, 2014). This observation is consistent
with a hypothesis of hydropassive stomatal control, mediated via a higher ratio of evaporative
demand to hydraulic supply in C₃ than C₄ species (Osborne & Sack, 2012), but may also follow
from differences in the optimization of stomatal aperture relative to photosynthesis in C₃ and C₄
species (Taylor et al., 2014). In a common garden experiment of closely related grasses adapted to
similar habitats in the same regional flora, this difference in stomatal behaviour unexpectedly led to
higher stomatal conductance in C₄ than C₃ species during the early stages of drought (Taylor et al.,
2014). However, during chronic drought, non-stomatal limitation of carbon assimilation becomes
more important in C₄ than closely related C₃ grasses, and may reduce or eliminate the differences in
photosynthesis between them (Ghannoum et al., 2003; Ripley et al., 2007, 2010; Ibrahim et al.,
2008; Ghannoum, 2009; Taylor et al., 2011). The mechanisms underlying this behaviour are
unknown, but seem to correlate with low water potential in C₄ leaves (Ibrahim et al., 2008; Ripley
et al., 2010; Taylor et al., 2014), and could correspond to a failure of the C₄ cycle.

In some C₄ eudicots, modifications in the xylem architecture, including narrower and shorter
vessels, decrease the leaf conductivity, which provides protection against cavitation and thus
enhanced drought tolerance (Kocacinar & Sage, 2003, 2004). It might be assumed that the higher
water-use efficiency conferred by the C₄ physiology enabled decreases in leaf conductivity.
However, xylem modifications are already visible in the C₃-C₄ intermediates of Flaveria that have
water-use efficiencies similar to the C₃ species, suggesting that xylem modifications predated C₄-
related higher water-use efficiency, at least in this genus (Kocacinar et al., 2008). It has been
hypothesized that the decreased conductivity actually drove the evolution of a C₂ pathway in these
species (Kocacinar et al., 2008), and might therefore be seen as a C₂ precondition. This emphasizes
difficulties in generalizing the order of events during the transition from C₃ to C₄ photosynthesis,
such that some modifications might have evolved before C₄ physiology and favored its evolution in
some lineages, while they were enabled by C₄ physiology in others.

Addition of alternative carbon shuttles

The action of a decarboxylase is necessary directly after PEPC becomes responsible for a
significant part of atmospheric CO₂ fixation. The evidence accumulated so far however indicates
that the shuttling of carbon between PEPC and Rubisco (Figure 1a, steps b-d) diversified after
plants were already in a C₄ physiological state. The variation in the carbon shuttles among C₄ plants
belonging to the same C₄ groups (Gutierrez et al., 1974; Wang et al., 2014) indeed indicates either
that some shuttles present in the common C₄ ancestor were lost in some of the descendants, or that
shuttles were added in some descendants only. The second hypothesis receives strong support from
comparative analyses of genes encoding decarboxylating enzymes (Christin et al., 2009a, 2009b).
In particular, strong signatures of positive selection are associated with the evolution of C₄-specific PCK in grasses, and this selection is detected on branches nested within several of the C₄ groups (Christin et al., 2009a; Figure 2).

The C₄ biochemical pathway can be plastic and respond to the environment (Furbank, 2011). For example, leaves of maize change the balance between NADP-ME and PCK shuttles when subject to shade (Bellasio & Griffiths, 2014; Sharwood et al., 2014), and models suggest that the addition of alternative carbon shuttles increases the range of light conditions tolerated by the plant (Wang et al., 2014). These attributes often evolved long after the initial origins of C₄ photosynthesis, and might thus have allowed the colonization of habitats differing in their vegetation cover. These adaptations consequently allowed C₄ plants to expand their niches compared to the ancestors that first acquired a C₄ pathway, and contributed to the ecological diversity found within C₄ groups.

6. Contingency and the ecological diversity of C₄ plants

The evolution of C₄ photosynthesis is a long process, beginning with the acquisition of C₄ anatomical and biochemical functions in a C₃ context, and continuing long afterward with the development of novel attributes enabled by the C₄ pathway (Figure 2). Following the establishment of C₄ physiology, each C₄ lineage has subsequently diversified, in some cases producing more than a thousand extant species (GPWGII, 2012). The diversity of environments occupied by C₄ plants means that the C₄ syndrome cannot be associated with a simply defined ecological strategy, but only partially affects the ecological preference of each plant, which is also influenced by other attributes inherited from the C₃ ancestors or that evolved after C₄ photosynthesis (Stowe & Teeri 1978; Stock et al., 2004; Edwards et al., 2010). The ecological diversity of C₄ species is therefore contingent upon (i) the ecology of ancestral C₃ lineages, which has subsequently been modified by (ii) physiological changes imparted by C₄ photosynthesis and then (iii) radiation into new niche space. In recent years, a phylogenetic perspective has enabled these three interacting factors to be teased apart, to bring a deeper understanding of the ecological diversity of C₄ species. In the next two sections, we illustrate how these processes have operated, using the examples of growth rate and sorting along environmental gradients.

*Phenotypic integration – the example of growth*

Growth rate varies significantly among plant species, with fast growth especially important for the persistence of species in resource-rich or disturbed habitats, and slow growth associated with persistence in resource-limited environments (Grime & Hunt, 1975; Grime et al., 1997). C₄
photosynthesis increases the efficiency of canopy photosynthesis across a range of temperatures (Long, 1999), especially in open environments, and allows a higher maximum conversion efficiency of intercepted light energy into biomass compared with C₃ photosynthesis (Monteith, 1978). If all else were equal, the acquisition of C₄ photosynthesis would therefore increase the rate of plant growth under hot, sunny conditions. However, experimental comparisons have surprisingly failed to discern a clear general difference in growth between C₃ and C₄ species.

Snaydon (1991) compiled published aboveground productivity data for 34 herbaceous species across 88 sites, and found no significant difference between C₃ and C₄ species when latitude (and, by proxy, temperature and growing season length) were taken into account. The most productive species in this analysis were however all C₄, consistent with previous results (Monteith, 1978), and supporting the hypothesis that C₄ photosynthesis confers the potential for higher maximum productivity than in C₃ species (Hatch, 1999; Long, 1999). Indeed, work by Piedade et al. (1991) showed that productivity in the C₄ hydrophyte *Echinocloa polystachya* growing in nutrient-rich Amazon floodwaters approaches the theoretical limit predicted from the efficiencies of physiological processes. However, in general, direct comparisons between C₃ and C₄ plants have failed to show consistently faster growth in C₄ species under controlled environments (e.g. Öztürk et al., 1981; Pearcy et al., 1981; Hunt et al., 1996; Reich et al., 2003), natural climate conditions (e.g. Öztürk et al., 1981; Gebauer et al., 1987; Reich et al., 2001), or in comparisons between closely related C₃ and C₄ species (Slatyer, 1970; Rajendrudu & Das, 1982; Taylor et al., 2010). For example, Taylor et al (2010) compared 34 closely related species of C₃ and C₄ grass, sampling multiple independent C₄ lineages. Although leaf photosynthesis was higher in the C₄ species, as expected, there were no differences in relative growth and net assimilation rates between these C₃ and C₄ species. The evidence from multiple experiments is clear: the large differences in leaf photosynthesis typically observed between C₃ and C₄ species do not generally translate into faster rates of growth.

This apparent paradox might result from the way that C₄ photosynthesis is integrated into the phenotype of the whole organism. In particular, interactions among processes operating at the organismal scale mean that growth often does not depend strongly on area-normalized leaf photosynthesis (Poorter et al., 1990). First, a limited number of pairwise comparisons between ecologically similar or closely related species have shown that the leaves of C₄ plants may be shorter-lived than those in C₃ species (reviewed by Long, 1999), suggesting that higher photosynthesis may be associated with more rapid leaf turnover, with a negative effect on growth. In addition, the allocation of growth to leaves versus heterotrophic tissues (e.g. roots and stems) and the area to mass ratio of leaves (specific leaf area), each have major effects on growth that may
partially offset or fully obscure the effects of higher rates of leaf photosynthesis (Körner, 1991). These effects are illustrated by work on the recently diverged C₃ and C₄ subspecies of *Alloteropsis semialata*. Leaf photosynthetic rates differ between these taxa as expected from theory (Osborne et al., 2008). However, the associated differences in growth rates are partially offset by a lower allocation of growth to leaves, and a smaller specific leaf area in the C₄ than C₃ subspecies (Ripley et al., 2008), which both tend to oppose the effects of C₄ photosynthesis. More generally, comparative work indicates that each of these growth traits may show phylogenetic patterns (e.g. Burns & Strauss, 2012), which means that closely related species share similar attributes, and the growth rates of C₄ species may be contingent upon characters inherited from their C₃ ancestors.

An altered partitioning of growth from leaves to roots in C₄ plants has been noted in a number of pairwise comparisons between ecologically similar or closely related species (Slatyer, 1970; Long & Mason, 1983; Ripley et al., 2008; Taylor et al., 2010). In each documented case, the shift in partitioning is achieved alongside similar or faster rates of growth in the C₄ species. It has been hypothesized that this shift in allocation could arise from the higher PNUE of C₄ plants and may depend on the ecological context (Long, 1999). C₄ species of fertile and/or disturbed habitats may use the same investment of nitrogen to produce a larger leaf area than their C₃ counterparts, thereby promoting more rapid growth. In contrast, C₄ plants of infertile habitats may adopt a more conservative strategy by producing the same leaf area as their C₃ counterparts with less nitrogen, but investing the resultant surplus of nitrogen into root development to better acquire this limiting resource. The hypothesis is supported by studies of growth allocation in plants adapted to fertile and infertile habitats (reviewed by Long, 1999). In summary, although C₄ photosynthesis offers the potential for faster growth, there is little published evidence for a consistent general translation of higher rates of leaf photosynthesis into greater productivity. Instead, the effects of C₄ photosynthesis on growth are mediated by changes in allocation and turnover, and may depend on the ecological context in which they evolve.

*Ecological sorting at the global scale – temperature and water availability*

Temperature is the primary determinant of species distributions at the global scale (Woodward, 1987), and hot conditions have long been considered important for C₄ plant ecology (Black, 1971). Global distribution patterns in relation to temperature are especially strong for grasses, where the classic pattern is turnover from C₄ to C₃ species with declining temperature, along both latitudinal (Teeri & Stowe, 1976) and altitudinal (Rundel, 1980) gradients. However, phylogenetic analyses show that C₃ grasses closely related to C₄ lineages also inhabit warm environments, which is the ancestral condition for this taxonomic group (Edwards & Still, 2008; Edwards & Smith, 2010;
Figure 4). Differences in land surface temperature can be detected between the habitats of closely related C₃ and C₄ grasses (Still et al., 2013), but the classical global patterns arise largely because one lineage of C₃ grasses, the Pooidaeae, acquired cold adaptations in the Oligocene and subsequently diversified at high latitudes and altitudes (Edwards & Still, 2008; Sandve et al., 2008; Edwards & Smith, 2010; Pau et al., 2013; Visser et al., 2014; Figure 4). These observations have prompted a re-evaluation of how C₄ taxa are distributed in relation to climate.

Because of the extra metabolic cost of C₄ photosynthesis, net leaf photosynthesis under light-limited conditions is lower for C₄ than C₃ plants at low temperatures, where the energetic benefit of suppressing photorespiration is limited (Ehleringer & Bjorkman, 1977; Collatz et al., 1998). Model simulations of leaf or canopy photosynthesis that account for this effect therefore predict a “crossover temperature” below which C₃ plants outperform their C₄ counterparts (Ehleringer, 1978; Collatz et al., 1998). However, under light-saturated conditions, energy is absorbed in excess of that required to drive the C₄ cycle and, for a given investment in Rubisco, leaf photosynthesis is higher at all temperatures in a C₄ than C₃ leaf (Long, 1999). As a consequence, a more complex photosynthesis model accounting for the penetration of direct light as sunflecks into the canopy shows that photosynthesis may be higher in a C₄ than C₃ canopy at temperatures down to 10 °C (Long, 1999). However, a lower concentration of Rubisco in C₄ than C₃ leaves leads to a temperature trade-off in light-saturated photosynthesis, with a crossover temperature similar to that observed under light-limitation (Still et al., 2003). Thus, according to theory, if C₄ plants maintain a high investment in Rubisco, there is no intrinsic energetic cost that would prevent them from colonizing open habitats in cool environments, particularly if they also have an open canopy.

C₄ physiology evolved in warm climatic regions of the subtropics (Ehleringer et al., 1991; Sage, 2004; Edwards & Smith, 2010), and the leaves of many C₄ species suffer chilling and freezing damage in common with other tropical and subtropical plants (Pearce, 2001). However, after evolving the C₄ syndrome, a number of plant lineages migrated into cool climate regions (Edwards & Smith, 2010; Figure 5), and now inhabit high temperate latitudes (Bjorkman et al., 1975; Long et al., 1975) and montane habitats (Sage & Sage, 2002). Absolute minimum winter temperatures impose a stringent climatic filter on the species that can persist in these environments, and adaptation requires the prevention or tolerance of ice formation within tissues during extreme low temperature episodes (Woodward, 1987). Many C₄ species of cold environments survive winter freezing events by either adopting an annual life history (e.g. weeds) or being deciduous (e.g. prairie grasses), in both cases overwintering in a dormant state, which is a common strategy adopted by plants to avoid episodic freezing (Zanne et al., 2013). However, there seems to be no intrinsic
barrier to freezing tolerance in a C₄ leaf, with species developing protection via constitutive or facultative cold acclimation mechanisms (Sage & Sage, 2002; Liu & Osborne, 2008, 2013). The leaves of other C₄ species are intolerant of freezing, but have physiological mechanisms for protection against light-mediated damage during chilling events in the range 0-10 °C (Long, 1983; Naidu et al., 2003). In conclusion, C₄ photosynthesis evolved in hot environments because there was a strong selective pressure for decreased photorespiration in these conditions. However, it can offer smaller benefits at low temperatures under high light conditions, so that C₄ plants can colonize cooler regions following the acquisition of cold adaptations, increasing the ecological diversity within C₄ groups (Figure 5).

The water-saving and hydraulic benefits of the C₄ syndrome outlined in Section 5 lead to the prediction that C₄ species should occupy drier habitats and environments with higher potential evaporation than C₃ species. It has long been known that C₄ eudicots sort into drier climate regions than their C₃ counterparts ( Ehleringer et al., 1997). There seems however to be a phylogenetic effect, with C₄ eudicots that are extremely well adapted to arid conditions having evolved from C₃ ancestors that already inhabited dry conditions (Stowe & Teeri, 1978), and, in several groups of eudicots, the distributions of related C₃ and C₄ lineages along environmental gradients largely overlap (Sudderth et al. 2009; Edwards & Ogburn, 2012; Figure 4). Similarly, in the Chenopodiaceae group, C₃ plants that were more tolerant of salinity gave rise to C₄ halophytes (Kadereit et al., 2012). Early studies failed to detect an overall relationship between the distribution of C₄ grasses and rainfall (Hattersley, 1983; Ehleringer et al., 1997), despite the clear differences in water relations between C₄ and C₃ grass species. A phylogenetic perspective has resolved this paradox by revealing a complex picture in which contingency, physiological innovation, and subsequent ecological radiation have each played important parts.

Phylogenetic patterns in the precipitation (Edwards & Smith, 2010) and habitat water requirements of grasses (Osborne & Freckleton, 2009) mean that closely related species tend to occupy similar environments, and both the global and regional distributions of major grass lineages thus differ in relation to precipitation (Taub, 2000; Edwards & Smith, 2010; Visser et al., 2012, 2014; Figure 4). This latter pattern has long been recognized in the differing geographical and climate space occupied by different taxonomic groups (Hartley, 1950). When C₄ photosynthesis evolved against this background, it modified physiological relationships with the environment, but plants nonetheless tended to retain attributes of their ancestors (Figure 4). The variation among groups of C₄ grasses might therefore result from the ecological diversification of grasses before C₄ evolution. For instance, the groups of C₄ grasses that prosper in more arid conditions, such as Aristidoideae
and Chloridoideae (Edwards & Smith, 2010; Visser et al., 2012, 2014), have C₃ relatives that inhabit similarly arid habitats (Gibbs Russell & Le Roux, 1990; Cerros-Tlatilpa et al., 2011). Despite this phylogenetic effect, the transition to C₄ physiology was still accompanied by changes in the ecological niche. Ancestral state reconstructions show that C₄ evolution in grasses led to consistent shifts into drier and more seasonal niche space (Edwards & Smith, 2010), and C₄ grasses are more likely to migrate into arid or saline habitats than their C₃ counterparts (Osborne & Freckleton, 2009; Brohman & Bennett, 2014), suggesting that C₄ photosynthesis facilitates adaptation to conditions of low soil water potential, probably through the continuous adaptation of stomatal conductance and plant hydraulics, and thereby allows plants to more readily access dry niche space (Edwards & Donoghue, 2013; Figure 5). In sedges however, many clades of C₃ species that prosper in more humid habitats produced C₄ descendants that share this preference (Stock et al., 2004). Water-use efficiency is likely irrelevant for sedges of infertile wetlands, where the C₄ advantage might result from the associated nitrogen-use efficiency (Li et al., 1999; Stock et al., 2004). On the other hand, a high maximum rate of growth may be critical for sedges of fertile wetlands (Muthuri et al., 1989), highlighting the diversity of ecological strategies enabled by the C₄ syndrome.

In summary, phylogenetic analyses show that contingency has played an important role in shaping the ecological niche of C₄ plants. This is classically illustrated by island colonists, like the C₃ Scaevola and C₄ Euphorbia lineages of Hawaii. Each is likely derived from a single island colonist, but has radiated into a similar diversity of habitats ranging from wet, closed forest to dry, open scrub, irrespective of the difference in their photosynthetic pathway (Robichaux & Pearcy, 1984). However, ecological diversification into the vacant niches offered by volcanic islands represents a special case. Generally, the ecological preferences inherited from C₃ ancestors have been affected by C₄ physiological novelty in subsequent diversification. This process of diversification is exemplified by the large C₄ group of Paniceae, which evolved from a C₃ ancestor inhabiting tropical seasonal forests but came to colonize diverse conditions after the evolution of C₄ physiology (Figure 5). Despite similar evolutionary times (Figure 5, left panel), the C₃ species in this group remained in a relatively small portion of the environmental space, with the exception of members of the Dichanthelium genus, which adapted to cold habitats (Figure 5). The transitions between C3 and C4 photosynthesis (blue branches leading to red branches in Figure 5) are associated with a slight shift to drier habitats in the same temperature range. This shift has already been reported and interpreted as a migration from forests in the aseasonal moist tropics to more open habitats in the seasonal subtropics, such as woodlands and savannas (Edwards & Smith, 2010; Figure 5).

Following this shift, the C₄ species from this group rapidly dispersed into habitats ranging from dry
and hot deserts to temperate grasslands and deciduous forests, and tropical rainforests (Figure 5). This pattern highlights the niche-opening effect of C₄ photosynthesis, which enables adaptation to new environments, probably through the adaptive integration of other attributes of the plants with the C₄ syndrome.

7. Conclusions

The evolutionary history of each C₄ taxon is rich and unique. It starts with the acquisition by its ancestors of characters that are required to build a C₄ system, but which evolve for completely unrelated reasons. Once all the characters exist in a given plant, these can be co-opted to create a weak C₄ cycle following an increase of PEPC activity. This key event creates new selective pressures toward the optimization of the C₄ pump, but it is not the end of the evolutionary process. The ecological preference of each C₄ group initially depends on the attributes inherited from its C₃ ancestors, but changes that happened during and after the transition to C₄ physiology allow plants to escape this heritage. The ecological strategies of specific C₄ plants are best understood by considering their whole evolutionary history, including the characters that were present in the C₃ ancestors, the way the C₄ apparatus was assembled, and the modifications to this apparatus that happened during the diversification of the C₄ group.

Acknowledgements

We are thankful to Douglas Ibrahim for his literature review of growth rates, Erika Edwards for ecological data on grasses and R. Matthew Ogburn for ecological data on Molluginaceae. PAC is funded by a Royal Society Research Fellowship URF120119.
**Box 1. Phylogenetic analyses and the evolution of complex phenotypes**

Comparisons among groups of species that differ in specific traits is complicated by two factors: i) other attributes of each species alter the effects of the traits; and ii) species are not statistically independent, because of their shared evolutionary history. These problems can be partially solved by taking the evolutionary history into account in comparative analyses. Phylogenetic trees are primarily used to reconstruct the relationships among species, but have also become important in comparative analyses. Their integration into statistical tests of differences among species can remove the variance due to shared evolutionary history, and thus identify properties that are associated with given traits independently of this history (Freckleton et al., 2002). In the case of C₄ photosynthesis, this approach can differentiate attributes that are directly conferred by the C₄ physiology from those that are usually associated with it, but might be inherited from their C₃ ancestors (Edwards & Smith, 2010). The origin of a trait on a phylogenetic tree can be mapped through different ancestral reconstruction methods, which estimate the character state for each speciation event, represented by each node in a phylogenetic tree (Figure 6). For instance, parsimony methods identify scenarios that minimize the number of transitions between character states, and methods based on likelihood estimate the most likely scenario given a set of assumptions (Figure 6). While these are powerful for testing specific hypotheses, such as the statistical association between sets of traits (e.g. Pagel, 1994; Osborne and Freckleton, 2009; Kadereit et al., 2012), the inferred ancestral states are dependent on the underlying model (Maddison, 2006; Christin et al., 2010). This problem can be partially solved by decomposing a complex trait into its constituents, so that the modelled entities are relatively simple properties and not complex phenotypes that result from multiple underlying characters (Christin et al., 2010; Roalson, 2011).

Changes in discrete or quantitative characteristics can be estimated with different methods (Christin et al., 2013b; Figure 6). The timing of these changes can then be estimated either relative to each other, by comparing the order of nodes (Figure 6), or in absolute terms, based on the age associated with the branch on which they happened (Figure 6). In addition, phylogenetic analyses of DNA sequences encoding genes of interest can identify past episodes of adaptive evolution (Zhang et al., 2005), and their positioning on phylogenetic trees can highlight periods of protein adaptation linked to an adaptive shift (Figure 2). Each of these methods comes with caveats, and considering multiple sources of information is crucial when inferring the evolutionary history of complex traits.
References


Hattersley PW. 1984. Characterization of C₄ type leaf anatomy in grasses (Poaceae). Mesophyll:


**Kadereit G, Ackerly D, Pirie MD.** 2012. A broader model for C₄ photosynthesis evolution in plants inferred from the goosefoot family (Chenopodiaceae s.s.). *Proceedings of the Royal Society B* 279: 3304-3311.

**Kapralov MV, Kubien DS, Andersson I, Filatov DA.** 2011. Changes in Rubisco kinetics during
the evolution of C₄ photosynthesis in *Flaveria* (Asteraceae) are associated with positive selection on genes encoding the enzyme. *Molecular Biology and Evolution* **28**: 1491-1503.


**Sage RF, Peary RW. 1987.** The nitrogen use efficiency of C₃ and C₄ plants. II. Leaf nitrogen effects on the gas exchange characteristics of Chenopodium album (L.) and Amaranthus retroflexus (L.). *Plant Physiology* **84**: 959-963.


Table 1: Hierarchical deconstruction of the C4 syndrome into different phenotypic levels, from the cell or enzyme to the whole organism.

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Niche</td>
<td>Environmental conditions in which the organism grows naturally</td>
<td>Warm and open environments</td>
</tr>
<tr>
<td>Physiology</td>
<td>Attribute of the whole organism that is generated by a combination of functions</td>
<td>C4 photosynthesis, growth rate, water-use efficiency</td>
</tr>
<tr>
<td>Function</td>
<td>Action at the cellular or tissue level that is enabled by a combination of underlying characters</td>
<td>Rapid transport of C4 intermediates, fixation of atmospheric carbon by CA + PEPC</td>
</tr>
<tr>
<td>Character</td>
<td>Emergent property of one component that is determined by multiple characteristics</td>
<td>Distance between consecutive bundles, activity of PEPC in the mesophyll</td>
</tr>
<tr>
<td>Characteristic</td>
<td>Property of one component that is theoretically independent from the others</td>
<td>Length of bundle-sheath cells, expression level of PEPC</td>
</tr>
<tr>
<td>Component</td>
<td>One cellular or enzymatic element</td>
<td>Bundle-sheath cell, PEPC</td>
</tr>
</tbody>
</table>
Figure captions

Figure 1. Schematic of the C₄ cycle.
(a) Simplified diagram representing the functional properties of the C₄ cycle (Table 1), which is consequently applicable to all C₄ plants. The main biochemical steps are indicated by circled letters. Atmospheric CO₂ enters the first compartment (dashed grey line) by diffusion. It is fixed into the C₄ cycle (a), which results in C₄ acids (in red) that are transformed and transported (b) to the second compartment (grey line), where CO₂ is released (c). The C₄ cycle is completed by the regeneration of the resulting C₃ acid (d). (b) One of the realizations of the C₄ cycle, with the example of the grass Zea mays, based on Tausta et al. (2014). As in most C₄ species, reactions are segregated between the mesophyll and bundle-sheath tissues of the leaf. The C₄ acids are in red, and the circled numbers represent enzymes. The black circles indicate enzymes that are involved in all C₄ types. Ala = alanine, Asp = aspartate, mal = malate, OAA = oxaloacetate, PEP = phosphoenolpyruvate, pyr = pyruvate, 1 = carbonic anhydrase (CA), 2 = PEP carboxylase (PEPC), 3 = NADP malate dehydrogenase, 4 = NADP-malic enzyme (NADP-ME), 5 = alanine aminotransferase (ALA-AT), 6 = pyruvate, phosphate dikinase (PPDK), 7 = aspartate aminotransferase (ASP-AT), 8 = phosphoenolpyruvate carboxykinase (PCK), 9 = Rubisco and the C₃ cycle (Calvin-Benson cycle).

Figure 2. Gradual accumulation of C₄ characters inferred for grasses.
The dated phylogenetic tree for grasses was obtained from Christin et al. (2013b), with the time scale in million years ago (Ma). All groups containing only C₃ or C₂ species are compressed and in black. Monophyletic C₄ groups are compressed in red, with their numbering on the right following GPWGII (2012). The two main grass clades are delimited on the right (BEP and PACMAD). Important changes in anatomical characters are reported based on Christin et al. (2013b). Episodes of adaptive evolution of C₄ enzymes are based on Christin et al. (2007, 2009a, 2009b). The changes shown here represent only a fraction of all changes linked to C₄ evolution and their positioning is approximate because the species sampling was not identical in the different studies. The grey box represent the last 30 million years, when atmospheric CO₂ stayed below 500 ppm. OS = outer bundle sheath, BSD = distance between consecutive bundle sheaths.

Figure 3. Changes inferred during the transition from a C₃ ancestor to the C₄ species Flaveria trinervia.
Six different variables were reconstructed on the time-calibrated phylogeny for Flaveria from Christin et al. (2011a). The values inferred for each node between the root of the tree and Flaveria trinervia are plotted against the estimated age of the node. Dashed lines indicate the 95%
confidence interval for the reconstructed ancestral values. The coloured background indicates the estimated photosynthetic state through time, with C$_3$ in white, C$_{3}$-C$_{4}$ intermediate in yellow, C$_{4}$-like in orange and C$_{4}$ in red. The vein density values (in mm/mm$^2$) come from McKown et al. (2007), the PEPC activities (in µmol/mg Chl*h) come from Bauwe (1984) and Sudderth et al. (2007) for F. kochiana, the percentages of carbon fixed to C$_{4}$ acids were summarized from various sources by Vogan & Sage (2011), the CO$_2$ compensation points come from Ku et al. (1991) and Sudderth et al. (2007) for F. kochiana, and the photosynthetic water-use efficiency (PWUE; in mmol CO$_2$/mol H$_2$O) and photosynthetic nitrogen-use efficiency (PNUE; in µmol CO$_2$/mmol N*sec) come from Vogan and Sage (2011).

Figure 4. Ecological distribution of some C$_4$ taxa compared to their C$_3$ relatives.
For two distantly related groups that contain C$_4$ taxa (grasses and Molluginaceae), the mean annual temperature (MAT; in °C) is plotted against the mean annual precipitation (MAP; in mm year$^{-1}$). For grasses, environmental variables were extracted from Edwards & Smith (2010), with one point per species. For Molluginaceae, environmental variables were taken from Edwards & Ogburn (2012), with multiple localities per species. Grey points represent localities for C$_3$ species that belong to the sister-group of the clade with C$_4$ species (the BEP clade of grasses and the Portulacineae clade, respectively). Localities for C$_3$ species that are closely related to C$_4$ taxa are in black (C$_3$ grasses from the PACMAD clade and C$_3$ Molluginaceae, respectively), and those C$_4$ taxa in each group are in red.

Figure 5. Ecological diversity in C$_3$ and C$_4$ Paniceae.
The mean annual temperature (MAT; in °C) and mean annual precipitation (MAP; in mm year$^{-1}$) were extracted from the ecological dataset of Edwards & Smith (2010) for those members of the grass tribe Paniceae that were also present in the time-calibrated phylogeny of Christin et al. (2013b). In the phylogenetic tree on the left, dots at the tips are coloured according to the species means for MAT on the left and MAP on the right. Branches are coloured based on photosynthetic types, with C$_4$ clades in red and C$_3$ branches in blue. The phylogenetic relationships are projected into climatic space on the right. For clarity, the lower part of the tree that includes the C$_4$ clade 'MCP' (Melinidinae, Cenchrinae and Panicinae; GPWGII, 2012). In the righthand panels, each segment connects the values estimated for two consecutive nodes in the phylogenetic tree (see Box 1). The blue point indicates the root (also indicated on the phylogeny), while tips are indicated by blue arrows when C$_3$ and red arrows when C$_4$. The major biomes are approximately delimited with dashed grey lines. They follow Ricklefs (2008), and are numbered in
the lower panel; 1 = temperate rain forest, 2 = temperate deciduous forest, 3 = temperate grassland and desert, 4 = tropical rainforest, 5 = tropical seasonal forest, 6 = savanna, 7 = subtropical desert.

**Figure 6: Examples of phylogenetic inference.**

**A.** Hypothetical time-calibrated phylogenetic tree for a group of C₄ species nested within a C₃ clade.

**B.** Hypothetical quantitative character mapped onto the tree using a maximum likelihood method. The estimated value for each node comes with confidence intervals, but only the optimum is presented as the dot size. **C.** Hypothetical binary character mapped on the tree using a maximum likelihood method. The likelihood of each state at each node is represented by pie charts. In the most parsimonious scenario, the origin of C₄ photosynthesis in this group could be estimated between time units 4 and 3 (bold branch, Figure 6A). The increase in the quantitative trait happened between time units 6 and 4 (bold branch, Figure 6B), before the change in the binary trait, which would be estimated between time units 4 and 3 based on a maximum likelihood model (bold dashed branch, Figure 6C) or between time units 3 and 1 based on a maximum parsimony approach (bold solid branch, Figure 6C).