doi:10.1111/evo.13250

Introgression and repeated co-option facilitated the recurrent emergence of C_4 photosynthesis among close relatives

Luke T. Dunning,[1,*] Marjorie R. Lundgren,[1,*] Jose J. Moreno-Villena,[1,*] Mary Namaganda,[2] Erika J. Edwards,[3] Patrik Nosil,[1] Colin P. Osborne,[1] and Pascal-Antoine Christin[1,4]

[1]*Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, United Kingdom*

[2]*Makerere University, Kampala, Uganda*

[3]*Department of Ecology and Evolutionary Biology, Brown University, Providence, Rhode Island 02912*

[4]*E-mail: p.christin@sheffield.ac.uk*

Received September 6, 2016
Accepted April 4, 2017

The origins of novel traits are often studied using species trees and modeling phenotypes as different states of the same character, an approach that cannot always distinguish multiple origins from fewer origins followed by reversals. We address this issue by studying the origins of C_4 photosynthesis, an adaptation to warm and dry conditions, in the grass *Alloteropsis.* We dissect the C_4 trait into its components, and show two independent origins of the C_4 phenotype via different anatomical modifications, and the use of distinct sets of genes. Further, inference of enzyme adaptation suggests that one of the two groups encompasses two transitions to a full C_4 state from a common ancestor with an intermediate phenotype that had some C_4 anatomical and biochemical components. Molecular dating of C_4 genes confirms the introgression of two key C_4 components between species, while the inheritance of all others matches the species tree. The number of origins consequently varies among C_4 components, a scenario that could not have been inferred from analyses of the species tree alone. Our results highlight the power of studying individual components of complex traits to reconstruct trajectories toward novel adaptations.

KEY WORDS: **Ancestral state, complex trait, co-option, reticulate evolution, species tree.**

Inferences of transitions among character states along species phylogenies provide powerful tools to test specific hypotheses about the timing and rate of functional diversification, correlations among functional and ecological traits (e.g., Pagel 1999; Edwards et al. 2010; Danforth et al. 2013; Moreau and Bell 2013; McGuire et al. 2014; Halliday et al. 2016), and speciation rates (Rabosky et al. 2013; Cantalapiedra et al. 2017; Cooney et al. 2017). However, distinguishing between a single origin of a trait with subsequent losses versus multiple independent origins can be problematic (Whiting et al. 2002; Pagel 2004; Wiens et al. 2006; Gamble et. al. 2012), particularly when some character states affect the rates of speciation and/or extinction, when rates of transitions are high and asymmetrical, or variable among clades and through time (Maddison 2006; Goldberg and Igic 2008; Beaulieu et al. 2013; Igic and Busch 2013; King and Lee 2015). Indeed, transition rates might be higher in taxonomic groups possessing evolutionary precursors that increase the likelihood of evolving a specific trait (Blount et al. 2008, 2012; Marazzi et al. 2012; Christin et al. 2013a, 2015; Werner et al. 2014). This can lead to an unbalanced distribution of character states across the tree, with clusters forming in certain clades. However, a low rate of origins would lead to similar patterns if the rate of reversals is high (Wiens 1999; Danforth et al. 2003; Trueman et al. 2004; Pyron and Burbink 2014). Difficulties worsen if hybridization and introgression disconnect the history of underlying traits from the species tree (Pardo-Diaz et al. 2012; Meier et al. 2017).

An alternative approach to analyzing the phenotypes as different character states is to decompose them into their constituent

*These authors contributed equally to the work

parts, then carefully analyze the evolution of each element independently to understand how the trait has been assembled or lost (Christin et al. 2010; Oliver et al. 2012; Niemiller et al. 2013; Kadereit et al. 2014). The use of distinct components can be interpreted as evidence for multiple origins, while reversals could leave a signature of the lost trait that can be detected when components are compared with those from species that never evolved it (Protas et al. 2006; Christin et al. 2010; Oliver et al. 2012; Niemiller et al. 2013). Identifying the mutations that underlie a trait further helps to distinguish shared origins and reversals (Igic et al. 2006; Shimizu et al. 2008; Niemiller et al. 2013; Meier et al. 2017). Evaluating the number of origins of each component of a complex trait would reconstruct the order of modifications that led to the trait of interest. This approach is applied here to the photosynthetic diversity exhibited within a five-species taxonomic group.

C_4 photosynthesis is a complex phenotype that improves the efficiency of carbon fixation in warm and dry conditions when compared to the ancestral C_3 photosynthetic pathway (Sage et al. 2012; Atkinson et al. 2016). The C_4 advantages are achieved by increasing the concentration of CO_2 around Rubisco, the enzyme responsible for inorganic carbon fixation in the Calvin cycle of all photosynthetic organisms (von Caemmerer and Furbank 2003; Sage et al. 2012). To function, C_4 photosynthesis requires the coordinated action of numerous anatomical and biochemical components that lead to the emergence of a novel biochemical pathway, usually across two types of cells; the mesophyll and bundle sheath cells (Hatch 1987; Prendergast et al. 1987; Gowik et al. 2011; GPWGII 2012; Bräutigam et al. 2014). Besides the increased expression of genes coopted for a C_4 function, several other changes are known to occur during the evolution of C_4 photosynthesis, including an expansion of bundle sheath tissue, a concentration of chloroplasts within it, and the adaptation of the enzymes to the new catalytic context (Fig. 1; Bläsing et al. 2000; von Caemmerer and Furbank 2003; McKown and Dengler 2007; Sage et al. 2012).

Despite its apparent complexity, C_4 photosynthesis evolved multiple times independently, and is present in distantly related groups of plants (Sinha and Kellogg 1996; Kellogg 1999; Sage et al. 2011). As with any complex trait, C_4 photosynthesis likely evolved in incremental steps, via stages that are functionally intermediate and gradually increase carbon assimilation in warm and dry conditions (Fig. 1; Sage et al. 2012; Heckmann et al. 2013; Williams et al. 2013; Mallmann et al. 2014; Christin and Osborne 2014). An increase in bundle sheath size and the relocation of the chloroplasts/Rubisco to these cells can sustain a photorespiratory bypass (Hylton et al. 1988; Bräutigam and Gowik 2016). Subsequent increases in C_4 enzyme abundances can generate a weak C_4 cycle, which assimilates some of the atmospheric CO_2, complementing the C_3 cycle in C_3+C_4 plants (referred to as "type II C_3-C_4 intermediates" in the specialized literature; Fig. 1;

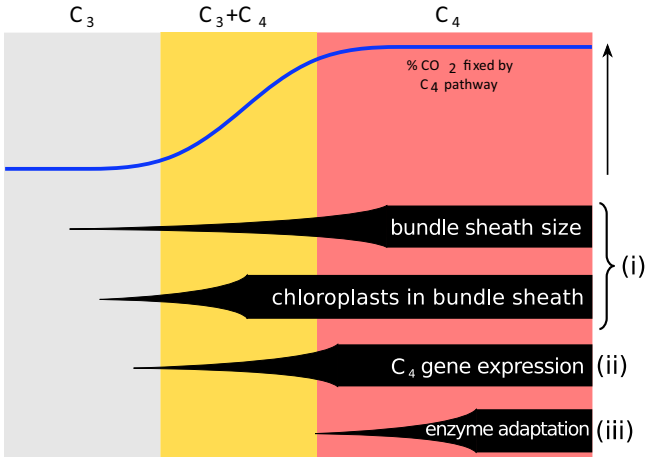

Figure 1. Schematic of expected changes during the transition from C_3 to C_4.
The continuous variation in anatomical and biochemical components can be simplified using three phenotypic categories; C_3 plants, C_3+C_4 intermediates, and C_4 plants. A schematic indicates the proportion of atmospheric CO_2 fixed by the C_4 cycle, and the expected order of modifications is shown at the bottom for four categories of changes, with the number on the right indicating the section of our analyses where they are investigated.

Heckmann et al. 2013; Mallmann et al. 2014). The transition to a full C_4 state involves further increases of the bundle sheath tissue and gene expression, while selective pressures adapt the C_4 enzymes for the new biochemical context (Fig. 1; Bläsing et al. 2000; McKown and Dengler 2007).

In the angiosperm phylogeny, C_4 taxa form clusters, many of which have multiple C_4 clades that are separated by non-C_4 branches (Sage et al. 2011; GPWGII 2012). Thus, establishing past photosynthetic transitions is difficult when photosynthetic type is modeled as a simple binary character (Ibrahim et al. 2009; Christin et al. 2010; Hancock and Edwards 2014; Bohley et al. 2015; Fisher et al. 2015; Washburn et al. 2015). Overall, nonhomology of key C_4 components among some closely related C_4 groups, including the cells, enzymes, and genes modified to generate the C_4 pathway (Prendergast et al. 1987; Soros and Dengler 2001; Bräutigam et al. 2014; Lundgren et al. 2014; Wang et al. 2014), points to a predominance of C_4 origins (Sinha and Kellogg 1996; Christin and Besnard 2009; Christin et al. 2010). However, the possibility of evolutionary reversals to a non-C_4 state is still debated (e.g., Kadereit et al. 2014; Bohley et al. 2015; Washburn et al. 2015). Furthermore, some components of the C_4 phenotype (e.g., expansion of bundle sheaths and migration of chloroplasts; Fig. 1) may have evolved relatively few times, and have then been recurrently used for independent transitions to C_3+C_4 or C_4 photosynthesis (Christin et al. 2011, 2013a).

One of the proposed candidates for an evolutionary reversal from C_4 to C_3 is in the grass genus *Alloteropsis* (Ibrahim et al.

2009). Within this genus, the species *Alloteropsis semialata* contains C_3, C_3+C_4, and C_4 genotypes (Ellis 1974; Brown 1975; Lundgren et al. 2016). In molecular phylogenies based on either plastid or nuclear markers, this species is sister to the C_4 *Alloteropsis angusta*, and the two species form a monophyletic clade sister to the three remaining closely-related C_4 species: *Alloteropsis cimicina*, *A. paniculata,* and *A. papillosa* (Ibrahim et al. 2009; Christin et al. 2012; Olofsson et al. 2016). The C_4 *A. semialata* and *A. cimicina* use different cell types for the segregation of C_4 reactions (Renvoize 1987), which suggests independent realizations of C_4 photosynthesis (Christin et al. 2010). However, the evolutionary origins of C_4 biochemistry and the situation within the *A. angusta*/*A. semialata* group remain largely unexplored.

In this study, we focus on the genus *Alloteropsis* and its C_3 outgroup, to test the competing hypotheses of multiple origins versus fewer origins followed by reversals, independently for each C_4 component. A C_4 phenotype generated via distinct cells, genes, or amino acid mutations would indicate independent origins. In contrast, a reversal may lead to a derived state that retains traces of its past C_4 state when compared to the ancestral one (i.e., approximated by the C_3 outgroup here). We combined different approaches to investigate different components of the complex C_4 trait. (i) Focusing on anatomical characters, we evaluate the most likely number of episodes of movement of chloroplasts to the bundle sheath, and expansion of this tissue. (ii) Using transcriptome analyses to estimate gene expression, we then determine the most likely number of origins of a C_4 cycle via the upregulation of known C_4 photosynthetic genes. (iii) The number of episodes of enzyme adaptation for the C_4 cycle is estimated using positive selection analyses, with scenarios corresponding to episodes of adaptation along different sets of branches. (iv) Finally, we compare divergence times across genes, to detect potential introgression of C_4 components, as suggested within this genus for two C_4 genes (Olofsson et al. 2016). Our multifaceted effort highlights the power of comparative analyses that directly consider genes and other components involved in the trait of interest, rather than modeling complex phenotypes as states of a single character. Using this approach, we show that recurrent origins of C_4 photosynthesis in *Alloteropsi*s arose via a complex mixture of co-option of traits increasing C_4 accessibility, hybridization, and independent adaptation of the phenotype.

Methods

TAXON SAMPLING

The different datasets were obtained from plants grown under controlled conditions (See Supporting Information Methods 1.1 for detailed description of growth conditions), including one *Alloteropsis cimicina* (C_4) accession, one *A. paniculata* (C_4) accession, two *A. angusta* (C_4) accessions, and up to 10 different *A. semialata* accessions collected from separate populations that encompass the global genetic and photosynthetic diversity of this species (one C_3, two C_3+C_4 intermediates with a weak C_4 cycle, and seven C_4 accessions; Table S1; Lundgren et al. 2016). The over representation of C_4 *A. semialata* accessions mirrors their natural abundance, with C_4 accessions spread throughout Africa, Asia, and Australia, C_3 accessions only reported in Southern Africa, and C_3+C_4 individuals restricted to central East Africa (Lundgren et al. 2015). We also make use of species representing the C_3 sister group to *Alloteropsis* (*Panicum pygmaeum* and *Entolasia marginata*), previously identified using plastid markers (GPWGII 2012). Using the above taxa, we conduct four complementary sets of analyses, each providing insight into the origins or spread of distinct components of C_4 in *Alloteropsis*.

(i) COMPARING LEAF ANATOMIES AMONG PHOTOSYNTHETIC TYPES

Leaf cross-sections were analyzed to identify the leaf compartment being used for the segregation of Rubisco and the modifications that increased the proportion of bundle sheath tissue in C_3+C_4 and C_4 accessions. Co-option of different tissues and distinct modifications among accessions would support independent origins, while a reversal should result in the leaves of C_3 individuals having reverted to a state that retain traces of their past C_4 state when compared to the ancestral condition (e.g., enlarged bundle sheath cells and/or chloroplasts in the bundle sheath).

We generated new anatomical data for nine *A. semialata* accessions and *A. angusta* (Table S2), which supplemented previously published anatomical data for *E. marginata*, *P. pygmaeum*, *A. cimicina*, and *A. paniculata* (Christin et al. 2013a). Images of *A. semialata* and *A. angusta* leaves in cross-section were obtained by fixing the center portion of a mature leaf blade in 4:1 ethanol:acetic acid, embedding them in methacrylate embedding resin (Technovit 7100, Heraeus Kulzer GmbH, Wehrhein, Germany), sectioning on a manual rotary microtome (Leica Biosystems, Newcastle, U.K.), staining with Toluidine Blue O (Sigma-Aldrich, St. Louis, MO), then photographing them with a camera mounted atop a microscope (Olympus DP71 and BX51, respectively, Olympus, Hamburg, Germany), as described in Lundgren et al. (2016).

All species used in this study have two bundle sheath layers, differentiated as inner and outer bundle sheaths, which create concentric circles around each vein (Fig. S1). The sheath co-opted for the segregation of Rubisco was identified by a concentration of chloroplasts producing starch. We also recorded the presence of minor veins, and measured the following traits on one cross-sectional image per accession, as described in Christin et al. (2013a), using ImageJ software (Schneider et al. 2012): the interveinal distance (IVD; the average distance between centers of consecutive veins), the number of mediolateral mesophyll cells

between veins, the average width of all outer and inner bundle sheath cells within a leaf segment, and the ratio of outer to inner bundle sheath cell widths (OS:IS). One leaf cross-section was used per accession, with previous work showing the traits we are measuring exhibit little variation within populations (Lundgren et al. 2016).

(ii) COMPARING GENE EXPRESSION PROFILES AMONG PHOTOSYNTHETIC TYPES

We use RNA-Seq to identify the genes co-opted by the different accessions performing a C_4 cycle, as those encoding C_4-related enzymes that reach high abundance in C_4 leaves. Variation in the co-opted loci would support multiple origins of a weak C_4 cycle, while a reversal might lead to high expression of C_4-related genes in individuals without a C_4 cycle or loss of functions of genes previously used for the C_4 cycle.

For RNA-Seq, we sampled the highly photosynthetically active distal halves of fully expanded new leaves and fresh roots midway into the photoperiod, which were subsequently flash frozen. Two different photoperiods (i.e., 10 and 14 h) were used to ensure that the identification of the most highly expressed genes did not differ among light regimes. Data from root libraries were only used in this study for transcriptome assembly, while all leaf samples were used for both assembly and quantification of transcript abundances. For a full list of individuals, conditions, and tissues sampled see Table S3.

Total RNA was extracted, Illumina TruSeq libraries generated, and sequencing performed using standard laboratory procedures, and transcriptomes were assembled using available pipelines (see Supporting Information Methods 1.2 for a detailed description of RNA-Seq protocol and assembly statistics). For each assembled contig, the transcript abundance was calculated as reads per million of mapped reads (rpm). Using a previously developed phylogenetic annotation pipeline (Christin et al. 2013b, 2015), the transcript abundance was then calculated for each gene lineage encoding C_4-related enzymes. For each gene family, all sequences descending from a single gene in the common ancestor of grasses via speciation and/or duplication were considered as the same gene lineage (i.e., these are grass co-orthologs). These groups include potential lineage specific paralogs (i.e., also known as inparalogs). When different *Alloteropsis* genes were identified within the same group of co-orthologs through detailed phylogenetic analyses, the abundance of each group was estimated independently. In *Alloteropsis*, this is the case only for genes previously shown to have been acquired laterally from distantly related C_4 lineages (Christin et al. 2012; see Results). In short, the reference datasets, composed of *Arabidopsis thaliana* coding sequences annotated as encoding C_4-related enzymes, and homolog sequences from other completely sequenced plants including five grasses, were retrieved from Christin et al. (2013b; 2015), or generated following the same approach for additional C_4-related enzymes identified in more recent studies (Mallmann et al. 2014; Li et al. 2015; Fig. S2). Contigs with similar sequences from the transcriptomes generated here were identified using BLASTn, with a minimal *e*-value of 0.01, and a minimal matching length of 50 bp. Only the portion of the contig matching the references was considered to remove UTRs, potential introns, and other very variable segments. Each sequence retrieved this way was then aligned independently to the reference dataset using Muscle (Edgar 2004), and a phylogenetic tree was inferred using Phyml (Guindon and Gascuel 2003) with a GTR+G+I model, a model that fits the vast majority of genes (e.g., Fisher et al. 2016) and is appropriate to infer a large number of trees. Phylogenetic trees were automatically screened, and each contig was assigned to the previously identified gene lineages in which it was nested. The sum of rpm values of all transcriptome contigs assigned to the same gene lineage produced transcript abundance per group of grass co-orthologs or distinct genes within these groups, which were subsequently transformed into rpm per kilobase (rpkm) values. Rpkm values were then compared among accessions to identify similarities and differences in the expression of C_4 photosynthetic genes.

(iii) GENE TREES AND DETECTION OF ENZYME ADAPTATION FOR C_4 PHOTOSYNTHESIS

Phylogenetic trees were inferred for C_4-related genes that were highly abundant in the leaf transcriptomes of at least two C_4 *Alloteropsis* samples (identified from transcriptome data; see Results) and their co-orthologs in other C_3 and C_4 grasses (see Supporting Information Methods 1.3 for a detailed description of phylogeny construction). The inferred gene trees were used to verify that C_4-related genes were placed as expected based on the species tree, as opposed to a position suggesting an acquisition from distant C_4 relatives. In addition, the gene tree topologies were used for positive selection analyses to detect traces of past episodes of enzyme adaptation for the new catalytic context after the initial emergence of a C_4 cycle (Fig. 1; Blasing et al. 2000; Christin et al. 2007; Besnard et al. 2009; Wang et al. 2009; Heckmann et al. 2013; Mallmann et al. 2014; Huang et al. 2017). Positive selection on branches leading to each C_4 group would support independent transitions to a full C_4 cycle via enzyme adaptation, while an early origin followed by a reversal should result in positive selection in the common ancestor of all C_4 accessions and possibly in the lineages that reversed back to the previous state.

For each set of genes encoding core C_4 enzymes in at least two *Alloteropsis* accessions, identified via transcriptome analyses, we optimized several codon models (site and branch-site models) to test for adaptive evolution using codeml as implemented in PAML (Yang 2007). The best-fit model was identified among those that assume (0) no positive selection (M1a null model), and

the branch-site models that assume shifts in selection pressure, either to relaxed selection (model BSA) or to positive section (model BSA1), at the base of: (1) *Alloteropsis* (one round of enzyme adaptation), (2) both *A. cimicina* and *A. angusta* + *A. semialata* (two rounds of enzyme adaptation), and (3) *A. cimicina*, *A. angusta*, and *A. semialata* (three independent episodes of enzyme adaptation). Foreground branches for all models were specified as the branch leading to the identified node plus all descending branches (i.e., using a "$" sign as opposed to a "#"). Models involving positive selection in only one of the C_4 lineages were also considered (see Supporting Information Methods 1.3 for additional details of positive selection analysis). For each gene lineage, the best-fit model was identified based on the corrected Akaike information criterion (AICc), selecting the model with the lowest AICc after checking that its ΔAICc score was at least 5.22 units below that of the M1a null model. An ΔAICc score = 5.22 corresponds to a *P*-value threshold of 0.01 for a likelihood ratio test comparing these two models using 2 degrees of freedom (df). C_4 species other than *Alloteropsis* were removed prior to analysis to avoid an influence of positive selection in these taxa affecting our conclusion. Analyses were repeated using only codons with fixed nucleotides within each lineage (i.e., *A. angusta,* C_3 *A. semialata,* C_3+C_4 *A. semialata,* and C_4 *A. semialat*a), to verify that short terminal branches with unfixed mutations did not significantly inflate the dN/dS ratio, and therefore alter our conclusion. Finally, to assess the effect of gene tree topology on our conclusions, we repeated the positive selection analyses using 100 bootstrap pseudoreplicate topologies.

(iv) DATING THE DIVERGENCE OF ADAPTIVE LOCI TO IDENTIFY INTROGRESSION

To determine whether introgression has spread C_4 adaptations among species, we performed molecular dating of markers from across the transcriptomes, including those used for C_4 by at least two *Alloteropsis* accession and their paralogs. The divergence times between species estimated from introgressed genes are expected to be younger than those estimated from other genes (e.g., Smith & Kronforst 2013; Li et al. 2014; Marcussen et al. 2014; Li et al. 2016), resulting either in outliers (if few genes are introgressed) or a multimodal distribution of ages (if many genes are introgressed).

Groups of genes descending from a single gene in the common ancestor of Panicoideae (Panicoideae co-orthologs), the grass subfamily that includes *Alloteropsis*, were identified through phylogenetic analyses of our transcriptomes and completely sequenced genomes that were publicly available. Our automated pipeline started with gene families previously inferred for eight plant genomes (homologs: i.e., all the paralogs and orthologs; Vilella et al. 2009), including two Panicoideae grasses (*Setaria italica* and *Sorghum bicolor*), two non-Panicoideae grasses (*Brachypodium distachyon* and *Oryza sativa*), and four nongrass species (*Amborella trichopoda, A. thaliana*, *Populus trichocarpa*, and *Selaginella moellendorffii*). To ensure accurate annotation, we restricted the analysis to gene families that included at least one *A. thaliana* sequence. The coding sequences (CDS) from the above genomes were then used to identify similar sequences in our transcriptomes using BLASTn with a minimum alignment length of 500 bp.

Stringent alignment and filtering methods were used to ensure reliable alignments of the above sequences for each gene family for phylogenetic inference (see Supporting Information Methods 1.4 for full details). In total, 2,797 1:1 Panicoideae co-ortholog datasets were used for subsequent molecular dating, as implemented in Beast version 1.5.4 (Drummond and Rambaut 2007). For each dataset, divergence times were estimated based on third codon positions, to decrease the risk of selective pressures biasing the outputs. A log-normal relaxed clock was used, with a GTR+G+I substitution model, and a constant coalescent prior. The *Sorghum* sequence was selected as the outgroup and the root of the tree was fixed to 31 Ma (using a normal distribution with a SD of 0.0001), based on estimates from Christin et al. (2014). There is uncertainty around this date, and the low species sampling used here probably leads to overestimation of both divergence times and confidence intervals, but the use of consistent sampling and calibration points among markers allows for the comparison of relative (rather than absolute) ages, which is the point of these analyses. Each Beast analysis was run for 2,000,000 generations, sampling a tree every 1,000 generations after a burn-in period of 1,000,000. For nodes of interest, divergence times were extracted from the posterior distribution as medians.

Divergence times were also estimated for key genes used for C_4 photosynthesis in *Alloteropsis* (identified based on transcriptomes; see Results), using the same parameters. To guarantee a consistent species sampling, the taxa included in the transcriptome-wide analyses were retrieved from manually curated alignments for C_4-specific genes as well as other groups of orthologs from the same gene families, obtained as described above for C_4-specific forms. In addition, plastid genomes for the same species were retrieved from Lundgren et al. (2015), and reanalyzed with the same parameters. For each of these datasets, the median, 95% CI, and 0.25 and 0.75 quantiles were extracted from the posterior distribution, using the R package APE (Paradis et al. 2004).

Results

(i) DIFFERENT REALIZATIONS of C_4 LEAF ANATOMY IN *A. CIMICINA* AND *A. SEMIALATA*/*A. angusta*

Grasses ancestrally possess two concentric rings of bundle sheath cells and either can be co-opted for C_4 photosynthesis (Brown

1975; Lundgren et al. 2014). The closely related C_4 *A. cimicina* and *A. paniculata* co-opted the outer bundle sheath for Rubisco segregation, as evidenced by the proliferation of chloroplasts in this tissue (Fig. S1; Table S2). In these species, the overall proportion of outer bundle sheath tissue within the leaf is increased via enlarged outer bundle sheath cells. Indeed, the outer sheath is 7.8-fold larger than the inner sheath in C_4 *A. cimicina* and *A. paniculata*, compared to a 1.2- to 0.6-fold differences in C_4 *A. semialata* and *A. angusta* (Table S2). This contrasts strongly with the anatomy of the C_4 *A. semialata* and *A. angusta* (Fig. S1). Both of these species use the inner bundle sheath for Rubisco segregation and increase the overall proportion of this tissue via the proliferation of minor veins, and enlargement of the inner sheath cell size (Fig. S1; Table S2).

Staining by Toluidine Blue O indicates some starch production occurs in the inner bundle sheaths of both the C_3 and C_3+C_4 *A. semialata* (Fig. S1), which implies some Rubisco activity in these cells, confirming previous reports (Ueno and Sentoku 2006; Lundgren et al. 2016). The absence of minor veins in the C_3 and C_3+C_4 *A. semialata* results in a larger proportion of mesophyll compared to C_4 *A. semialata* (Table S2; Fig. S1). In the C_3 and C_3+C_4 *A. semialata*, the outer bundle sheath is slightly larger than the inner one (1.2- to 1.8-fold; Table S2), while the C_3 outgroup species *P. pygmaeum* and *E. marginata* have outer bundle sheaths that are considerably larger than their small inner sheaths (4.5- and 5.3-fold; Fig. S1; Table S2).

In summary, our comparative studies of leaf anatomy indicate that the C_4 *A. cimicina* and *A. semialata*/*A. angusta* use different tissues for Rubisco segregation and achieve high bundle sheath proportions via distinct modifications, supporting independent origins of C_4 anatomical components in these two groups. Some Rubisco activity is suggested in the inner sheath of the C_3 *A. semialata*, which supports an early origin migration of chloroplasts to this tissue (Fig. 2). In addition, a slight enlargement of the inner sheath, absent in the C_3 outgroup, is common to all non-C_4 *A. semialata*.

(ii) *A.CIMICINA* USES DIFFERENT ENZYMES AND GENES FOR C_4 BIOCHEMISTRY THAN *A. SEMIALATA*/*A. ANGUSTA*

All *Alloteropsis* C_3+C_4 and C_4 accessions have high expression abundance in their leaves of co-orthologs encoding phosphoenolpyruvate carboxylase (PEPC), the enzyme used for the initial fixation of atmospheric carbon into organic compounds in

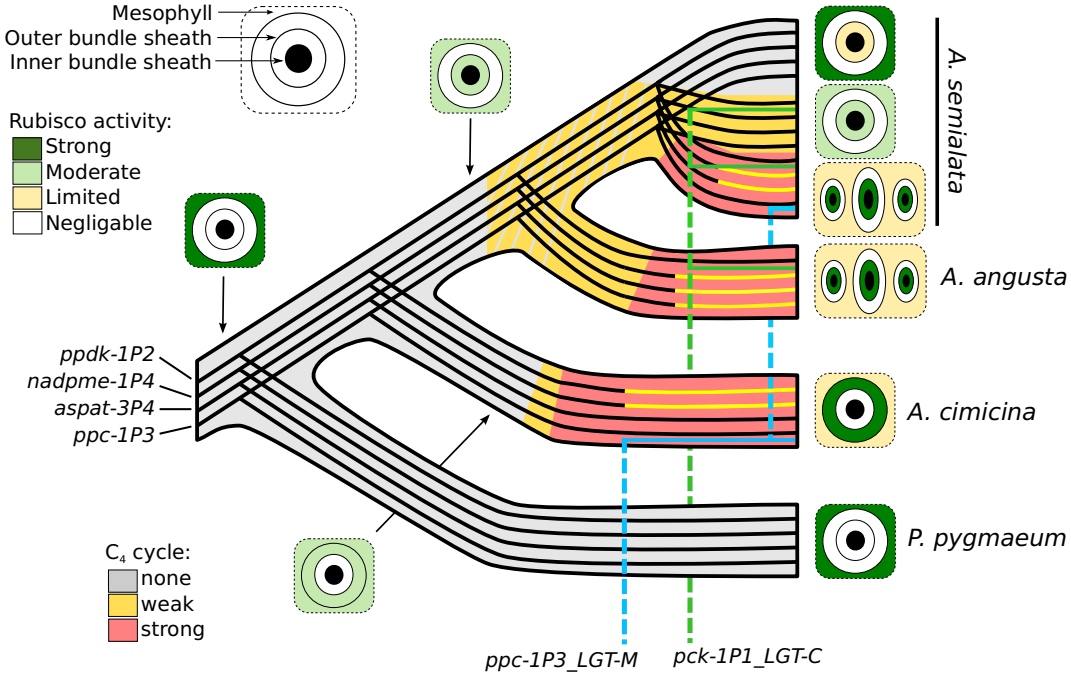

Figure 2. Inferred transitions among C_4 components.
A schematic phylogenetic tree is presented, based on previous genome-wide analyses (Lundgren et al. 2015; Olofsson et al. 2016). Individual lines represent the transmission of individual genes within the species complex. For each of the four genes subject to C_4-related selection, episodes of positive selection are indicated by changes to yellow. Other lines track the spread of two genes that were originally laterally-acquired from distant relatives, and have subsequently been introgressed among *Alloteropsis* species. The inferred phenotype is represented by the background colour, in grey for C_3, in yellow for C_3+C_4, and in red for C_4. The grey hatching indicates uncertainty about the ancestral state. A simplified version of leaf anatomy is represented, for extant taxa and some hypothetical ancestors (see Fig. S1 for details of leaf anatomy of extant accessions).

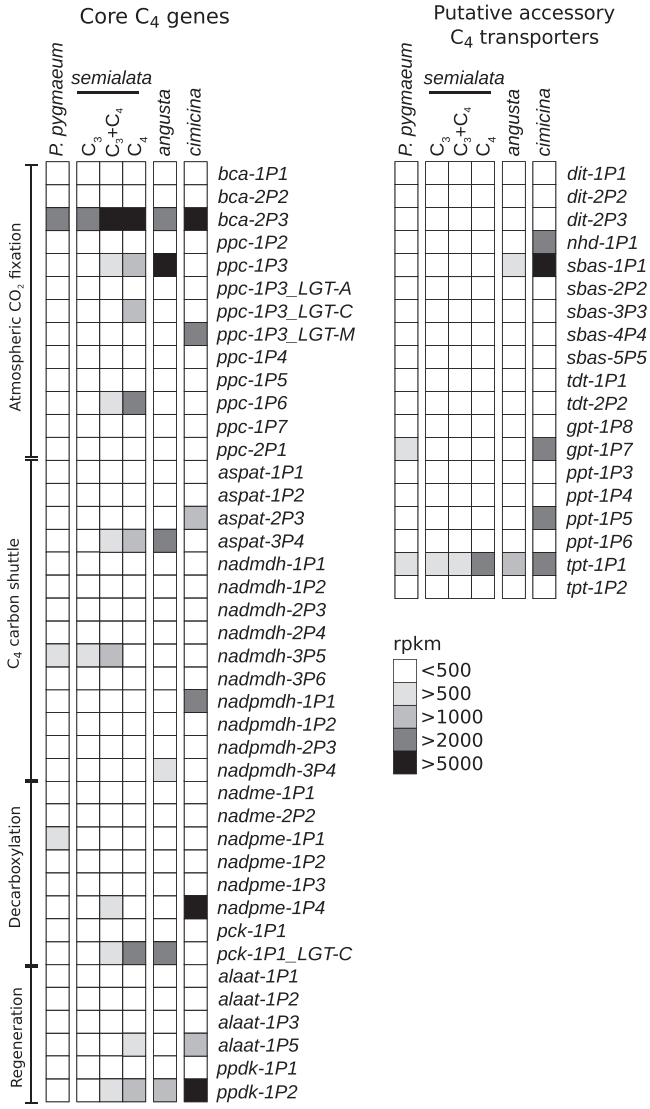

Figure 3. Expression of C_4-related enzymes in *Alloteropsis*.
For each gene encoding a C_4-related enzyme, the shade indicates the category of transcript abundance, using averages per group. For raw values, see Table S4. Note that *ppc* abundance varies among C_4 accessions of *A. semialata* (Fig. 4). The enzymes involved in core C_4 reactions (left column) are grouped by functional property, and gene names are written in italics on the right of the expression values.

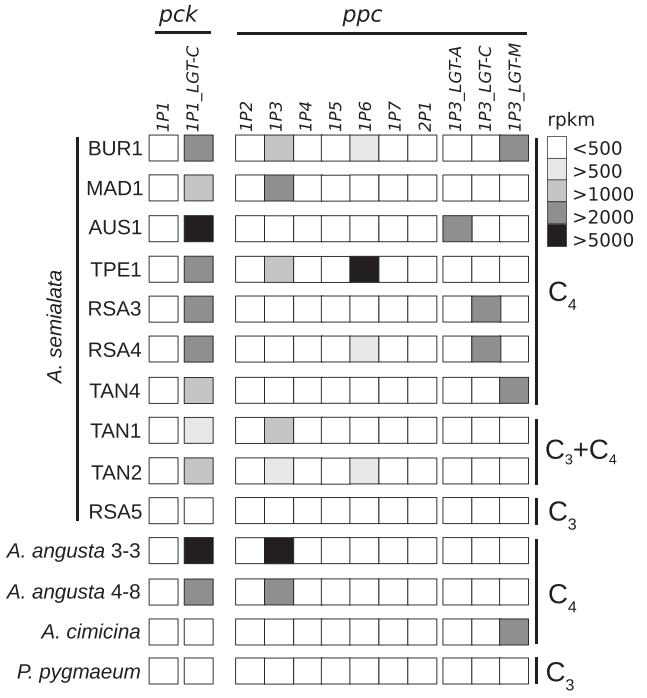

Figure 4. Leaf abundance of *pck* and *ppc* genes in the different accessions.
The shade indicates the relative expression (in rpkm) in the different accessions. For each accession, the averages are used. For raw values, see Table S4.

C_4 plants. However, the gene lineage most highly expressed varies among accessions (Figs. 3 and 4). The close relationships between some of the genes for PEPC and one for phosphoenolpyruvate carboxykinase (PCK) isolated from *Alloteropsis* and those of distantly related C_4 species was confirmed by our phylogenetic analyses (Figs. S3 and S4), supporting the previous conclusion that these genes were acquired by *Alloteropsis* via lateral gene transfer (LGT; Christin et al. 2012). Based on the read abundance, *A. cimicina* uses *ppc-1P3_LGT-M*, while *A. angusta* uses *ppc-1P3* (Fig. 4). There is variation within *A. semialata*, with C_3+C_4 and C_4 accessions using either one or a combination of several gene lineages several gene lineages (Fig. 4).

From the expression profiles (Fig. 3), the carbon shuttle of *A. cimicina* relies on enzymes and transporters associated with the most common form of C_4 photosynthesis (NADP-malic enzyme type; Gowik et al. 2011; Bräutigam et al. 2014; Mallman et al. 2014). This expression profile differs markedly from that observed in the C_4 *A. semialata* and *A. angusta* accessions. These two species mainly use the PCK decarboxylating enzyme, through the high expression of the same gene (*pck-1P1_LGT-C*; Fig. 4). There is little evidence in these species for an involvement of the auxiliary transporters observed in *A. cimicina* (Fig. 3; Table S4), and some of the core enzymes are not shared by *A. cimicina* and *A. semialata*/*A. angusta* (Fig. 3). Furthermore, even when the same enzyme family is used, it is not necessarily encoded by the same locus (e.g., *A. cimicina* expresses *aspat-2P3* and *A. semialata*/*A. angusta* express *aspat-3P4*; Fig. 3).

The transcriptomes of the C_3+C_4 *A. semialata* show elevated levels of some of the genes used by the C_4 *A. semialata*, with a slightly higher abundance of those encoding the NADP-malic enzyme (*nadpme-1P4*; Fig. 3; Table S4). In terms of the expression levels of genes encoding C_4-related enzymes, the transcriptome of the C_3 *A. semialata* is not markedly different from that of the C_3 outgroup *P. pygmaeum* (Fig. 3; Table S4).

Table 1. Results of positive selection analyses inferring the episodes of enzymatic adaptation in *Alloteropsis*[1].

			One origin		Two origins		Three origins		Only *A. cimicina*	
Gene	Number of sequences	Site model M1a	BSA	BSA1	BSA	BSA1	BSA	BSA1	BSA	BSA1
aspat-2P3	14	**0.00***	4.02	4.02	4.02	4.02	3.94	4.02	4.00	4.00
nadpme-1P4	15	35.07	30.44	27.26	26.34	24.28	19.52	13.31	3.34	**0.00***
ppdk-1P2	15	29.45	32.00	26.35	32.31	27.17	26.37	23.37	3.55	**0.00***
alaat-1P5	14	**0.00***	1.74	1.74	2.03	2.03	0.69	0.69	4.02	4.02

[1]The ΔAICc values compared to the best-fit model for that gene are shown. The most appropriate model is indicated with an asterisk, with the null model (M1a) only rejected if the ΔAICc was at least 5.22 (equivalent to a *P*-value of 0.01 with a likelihood ratio test with df = 2). Two branch-site models were used to test for a relaxation of purifying selection (BSA), and potential positive selection (BSA1).

Table 2. Results of positive selection analyses inferring the episodes of enzymatic adaption in the *A. angusta/A. semialata* clade[1].

			One origin		Two origins		Only *A. angusta*	
Gene	Number of sequences	Site model M1a	BSA	BSA1	BSA	BSA1	BSA	BSA1
aspat-3P4	13	12.33	10.20	6.70	6.37	**0.00***	5.45	5.29
nadpme-1P4	14	10.19	14.19	9.66	13.52	**0.00***	14.18	14.18
ppc-1P3	9	72.43	66.62	66.58	11.70	9.85	5.66	**0.00***
ppdk-1P2	14	**0.00***	4.01	4.01	4.01	4.01	3.91	3.91

[1]The ΔAICc values compared to the best-fit model for that gene are shown. The most appropriate model is indicated with an asterisk, with the null model (M1a) only rejected if the ΔAICc was at least 5.22 (equivalent to a *P*-value of 0.01, with a likelihood ratio test with df = 2). Two branch-site models were used to test for a relaxation of purifying selection (BSA), and potential positive selection (BSA1).

Our comparative transcriptomics therefore indicate that *A. cimicina* uses different genes and different enzymes for the C_4 pathway than *A. semialata*/*A. angusta*, suggesting multiple origins of the C_4 cycle (Fig. 2). The only C_4-related genes used by some C_4 *Alloteropsis* that are abundant in the C_3 *A. semialata* (*bca-2P3* and *tpt-1P1*) are also highly expressed in the C_3 outgroup and in other distantly related C_3 taxa (Fig. 3; Külahoglu et al. 2014; Ding et al. 2015), indicating that high levels in leaves is not specific to our group of species. For the C_4-related genes used by the C_4 *Alloteropsis*, but not abundant in the outgroup, there is no evidence for high expression or pseudogenization in the C_3 *A. semialata*. Evidence is thus lacking that the C_3 *A. semialata* represent a reversal from an ancestor with a C_4 cycle.

(iii) INDEPENDENT EPISODES OF C_4-RELATED POSITIVE SELECTION IN EACH C_4 SPECIES

The codon models do not support positive selection on any genes involved in C_4 photosynthesis at the base of *Alloteropsis* or along the branch leading to the *A. angusta*/*A. semialata* group (Table 1). In two cases (*nadpme-1P4* and *ppdk-1P2*), analyses including all *Alloteropsis* accessions clearly point to changes in selective pressures specifically in the branch leading to *A. cimicina* (Table 1; Fig. S5). No evidence of positive selection was found for the two other genes analyzed on the three *Alloteropsis* species (*aspat-2P3* and *alaat-1P5*; Table 1). When testing for selection only in the *A. angusta*/*A. semialata* clade, no positive selection was found on *ppdk-1P2*, while positive selection on *ppc-1P3* was identified only on the branch leading to *A. angusta* (Table 2). For the two other genes (*nadpme-1P4* and *aspat-3P4*), the model that assumes positive selection after the split of the two species was favored (Table 2). A majority of the amino acid sites identified as under positive selection by the Bayes Empirical Bayes analysis overlapped with those previously identified in other C_4 taxa (e.g., site 241 in *nadpme-1P4;* Fig. 5; Christin et al. 2009), or were shared with other C_4 species in our phylogenies (e.g., Fig. 5), supporting their link to C_4 photosynthesis. For *aspat-3P4*, more amino acid substitutions were fixed in *A. angusta* than in *A. semialata.* This variation among *A. semialata* C_4 accessions indicates repeated bouts of positive selection during the diversification of this species (Fig. S6). Conclusions based on the selection tests were also supported using only codons with fixed nucleotides within a lineage (i.e., photosynthetic types in *A. semialata,* and *A. angusta*), with the exception of *nadpme-1P4* for which no positive selection was inferred after removing the unfixed codons (Tables S5 and S6). Furthermore, gene tree topology had no effect on our conclusions, since all bootstrap replicates supported

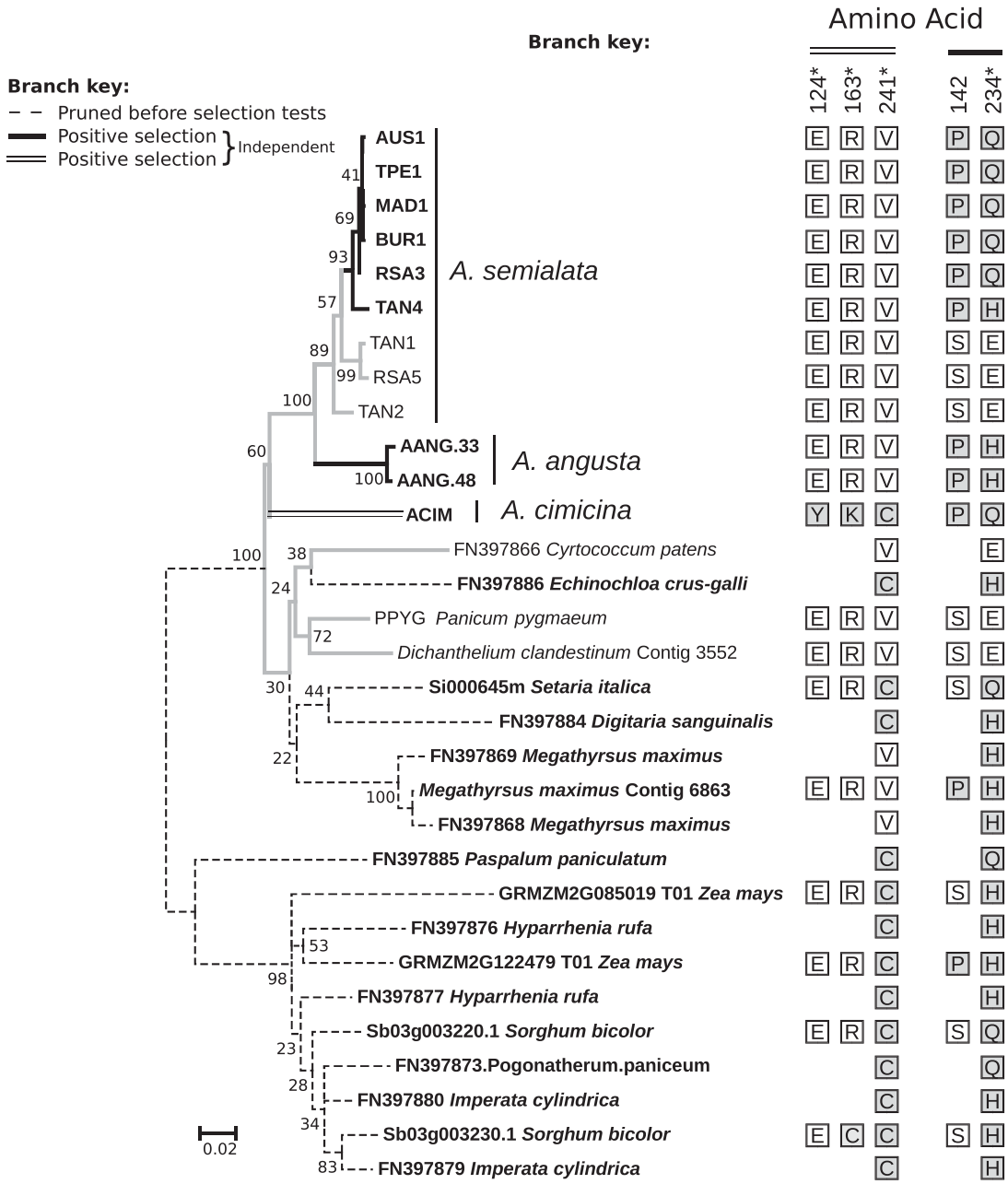

Figure 5. **Evolution of *nadpme-1P4* genes in *Alloteropsis* and other Panicoideae.**
This phylogenetic tree was inferred on 3rd positions of codons of *nadpme-1P4* genes of Panicoideae. Bootstrap values are indicated near branches. Names of C_4 accessions are in bold. Amino acid at positions under positive selection are indicated on the right, with those associated with C_4 accessions in gray. Positions are indicated on the top, based on *Sorghum* gene Sb03g003220.1. Amino acid positions with a posterior probability >0.90 of being under positive selection are indicated on the right, asterisks indicate positions with a posterior probability >0.95.

the same model, with the exception of 2% of *nadpme-1P4* bootstrap replicates (Tables S7 and S8).

Overall, our positive selection tests point to independent episodes of enzyme adaptation for the C_4 context in each of the C_4 groups (Fig. 2). None of the models that included adaptive evolution on branches leading to C_3 and/or C_3+C_4 *A. semialata* were favored, suggesting no evolutionary loss of a full C_4 cycle.

(iv) GENES FOR PEPC AND PCK WERE SPREAD ACROSS SPECIES BOUNDARIES

The 2,797 groups of orthologs extracted from genomes and transcriptomes led to a wide range of estimated divergence times, with 95% of the medians falling between 6.51 and 17.92 Ma for the crown of *Alloteropsis*, and between 4.17 and 11.27 Ma for the split of *A. semialata* and *A. angusta* (Fig. 6). The peak of values

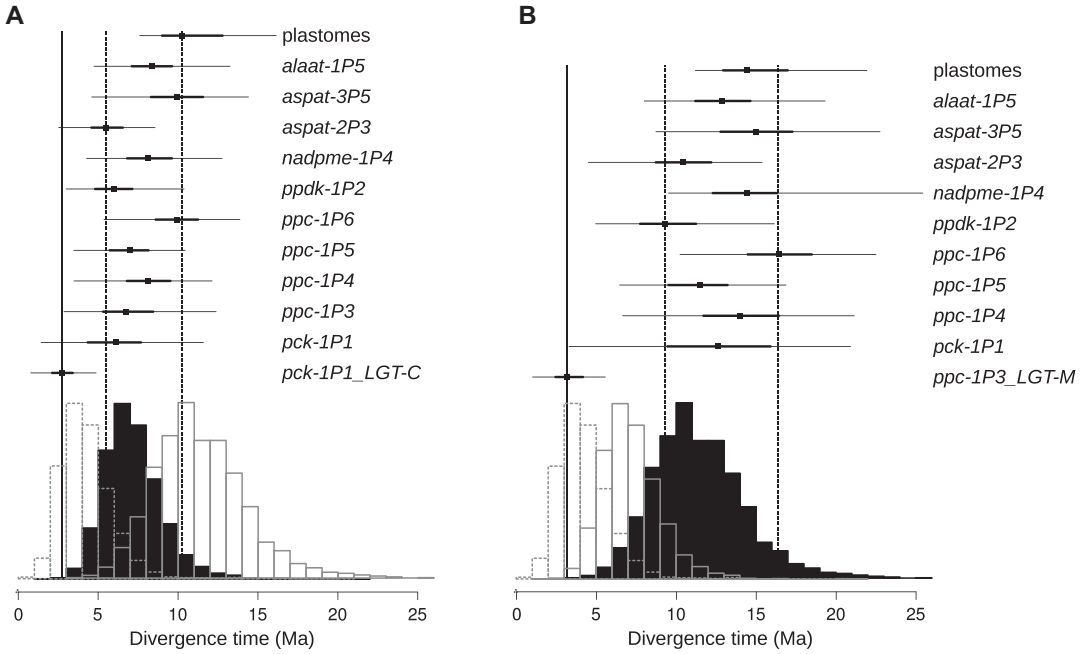

Figure 6. Estimates of divergence times.
On the top, divergence times are shown for selected nuclear genes and plastomes for A) the split of *A. angusta* and *A. semialata* and B) the crown of *Alloteropsis*. For each marker, the median of the estimates is indicated by a square, with thick bars connecting the 25 and 75 percentiles and thin bars connect the 2.5 and 97.5 percentiles. The distribution of medians for the crown of *A. semialata* (left), the split of *A. angusta* and *A. semialata* (middle), and the crown of *Alloteropsis* (right) over 2,797 markers extracted from the transcriptomes is shown at the bottom. The scale is given in million years ago (Ma).

(i.e., 50% of the points) ranged between 9.38 and 13.07 Ma for the crown of *Alloteropsis* and 5.93 and 8.18 Ma for the split of *A. semialata* and *A. angusta* (Fig. 6). Finally, 95% of the markers estimated the crown of *A. semialata* between 1.88 and 7.77 Ma, with a peak between 3.12 and 5.07 Ma (Fig. 6). Note that monophyly of the groups was not enforced, and various combinations of *A. semialata* accessions were included across markers, contributing to the observed variation.

Most of the C_4-related genes, as well as the plastomes, provided age estimates ranging from 5.54 to 10.32 Ma for the split of *A. semialata* and *A. angusta*, which matches the distribution of estimates from the transcriptome-wide data (Fig. 6A), and indicates their transmission followed the species tree. The only exception is the gene *pck-1P1_LGT-C*, for which the last common ancestor of *A. semialata* and *A. angusta* was estimated at 2.77 Ma (Fig. 6A), which is smaller than all but four of the 2,797 estimates from the transcriptome-wide markers. While the confidence intervals of the estimate for this gene do overlap with those of almost all other markers, this estimate matches more closely the diversification of *A. semialat*a accessions (Fig. 6A).

The different markers selected for detailed analyses similarly yielded estimates for the crown of *Alloteropsis* matching those obtained from transcriptome-wide data, between 9.38 and 16.46 Ma (Fig. 6B). The only exception is the gene *ppc-1P3_LGT-M,* for which the last common ancestor of *A. cimicina* and *A. semialata* is estimated at 3.25 Ma (Fig. 6B), which is smaller than all estimates based on markers extracted from the transcriptomes. The 95% CI of the divergence estimate based on this gene does not overlap with many of those based on other markers, and again matches closely with the diversification of *A. semialat*a accessions (Fig. 6B).

Overall, our dating analyses support an introgression of these two genes among *Alloteropsi*s species after their divergence, while the other genes were transmitted following the species tree (Fig. 2).

Discussion

TWO INDEPENDENT TRANSITIONS FROM C_3 TO C_4

The earliest split in *Alloteropsis* separates the lineage containing *A. cimicina* from *A. angusta* and *A. semialata* (Fig. 2). These two lineages co-opted different tissues for the segregation of Rubisco activity and achieved a large proportion of bundle sheath tissue via different modifications (Fig. S1). The evidence therefore strongly supports two independent origins of C_4 anatomical properties, which is generally accepted as the first step during the C_3 to C_4 transition (Fig. 1; Sage et al. 2012; Heckmann et al. 2013). Gene expression analyses show that the two clades use

different enzymes for parts of the C_4 cycle, express different genes encoding the same enzyme family when there is an overlap (Fig. 3), and positive selection analyses show that the enzymes were independently adapted for their C_4 function (Table 1). We therefore conclude that the different transitions to C_4 biochemistry occurred independently after the split of these two lineages (Fig. 2). The only exception to the distinctiveness of *A. cimicina* and the two other C_4 species is the gene *ppc-1P3_LGT-M*, used by both *A. cimicina* and some C_4 *A. semialata* accessions (Fig. 4). This gene is absent from other accessions (Olofsson et al. 2016) and, as such, we previously concluded that it was acquired early during the diversification of the group and then recurrently lost (Christin et al. 2012). This hypothesis is falsified by our dating analyses here, which show that this gene was only recently transferred among species boundaries, likely as a result of a rare hybridization event (Fig. 6).

ONE INDEPENDENT C_3 TO C_4 TRANSITION INCLUDES TWO SEPARATE C_3+C_4 TO C_4 SHIFTS

The C_4 phenotype is realized in *A. angusta* and *A. semialata* via identical anatomical modifications, using the same enzymes, and the same genes encode these enzymes. Chloroplasts are present in the inner sheaths of all *A. semialata* and *A. angusta* accessions, independent of their photosynthetic type, which suggests that this characteristic represents the ancestral condition for the clade (Fig. 2). The C_4 cycle is realized using the same set of genes in *A. angusta* and *A. semialata*, which can be explained by convergent evolution (e.g., as indicated for other C_4 grasses; Christin et al. 2013b) or a single origin of a weak C_4 cycle (C_3+C_4), followed by a reversal to expression levels that resemble the ancestral condition in the C_3 accessions (Fig. 2). Differentiating these two scenarios would require retracing the origin of the mutations responsible for the increased expression of C_4 enzymes to identify where they occurred on the phylogeny. Unfortunately, the molecular mechanisms controlling C_4 gene expression are poorly known, and can involve both cis- and transacting elements (Gowik et al. 2004; Brown et al. 2011; Williams et al. 2016).

The positive selection analyses indicate that enzyme adaptation happened independently in *A. angusta* and *A. semialata* (Table 2). Together with the variation observed within the C_4 *A. semialata* (Fig. 4, S6), this evidence strongly suggests that the biochemical adaptation allowing the transition to a full C_4 cycle happened recently, and independently in the two species (Fig. 2). The dramatic increase in the proportion of the inner bundle sheath tissue via the proliferation of minor veins is limited to the C_4 *A. semialata* and *A. angusta* (Fig. S1). The genetic control of these features is unknown, preventing a comparison of the causal mutations. However, the distribution of anatomical characters among grasses indicates that the vast majority of C_4 lineages that co-opted the inner bundle sheath increased its proportion via the addition of minor veins (Renvoize 1987; Christin et al. 2013a).

With the current state of knowledge, we hypothesize that the common ancestor of *A. semialata* and *A. angusta* had chloroplasts in the inner bundle sheath, and that this facilitated the emergence of a weak C_4 cycle via the upregulation of some enzymes. Following their split, *A. angusta* strengthened its C_4 anatomy via the proliferation of minor veins, and enzyme adaptations led to a strong C_4 cycle (Fig. 2). In the *A. semialata* lineage, some isolated populations acquired mutations that added minor veins and adapted the enzymes, leading to a C_4 cycle. Other populations, potentially under pressures linked to the colonization of colder environments (Lundgren et al. 2015), might have lost the weak C_4 cycle by downregulating the genes (Fig. 2). However, the details of the changes leading to C_3 photosynthesis in some *A. semialata* will need to be confirmed by comparative genomics, when mutations regulating expression of C_4 enzymes and anatomy are identified.

INTROGRESSION OF C_4 COMPONENTS AMONG SPECIES

Our dating analyses suggest that the gene *pck-1P1_LGT-C* that encodes the decarboxylating enzyme PCK was introgressed among some members of *A. semialata* and *A. angusta* (Figs. 2 and 6). The C_4 cycle carried out before this event was likely based on NADP-malic enzyme, an enzyme still abundant in the C_3+C_4 *A. semialata* and some C_4 accessions (Fig. 4; Frean et al. 1983). The acquisition of *pck-1P1_LGT-C*, a gene already adapted for the C_4 context, probably added a PCK shuttle, which alters the stoichiometry of the pathway and the spatial distribution of its energy requirements, increasing its efficiency under some conditions (Bellasio and Griffiths 2014; Wang et al. 2014). This important component of the C_4 cycles of extant *A. semialata* and *A. angusta* populations first evolved its C_4-specific properties in the distantly related *Cenchrus* (Fig. S3; Christin et al. 2012), and therefore never evolved within *Alloteropsis*. Instead, it represents the spread of a component of a complex physiology across multiple species boundaries. Therefore, in addition to the possibility that the sequential steps generating a complex physiology can happen on different branches of a species phylogeny (Fig. 2), introgression among close relatives can disconnect the origins of key components from the species tree.

ON THE INFERENCE OF TRANSITIONS AMONG CHARACTER STATES

Inferences of transitions among character states are a key component of numerous macroevolutionary studies (e.g., Cantalapiedra et al. 2017; Cooney et al. 2017). However, species trees per se are not always able to disentangle the complex scenarios underlying the appearance or losses of multicomponent adaptations,

especially when complex phenotypes are modeled as different states of a single character (e.g., Goldberg and Igic 2008; Pardo-Diaz et al. 2012; Niemiller et al. 2013; Igic and Busch 2013; King and Lee 2015). In the case of photosynthetic transitions within *Alloteropsis* depicted here, considering the photosynthetic type as a binary character would lead to a single C_4 origin as the most plausible scenario (Ibrahim et al. 2009), and modeling photosynthetic types based on their category of C_4 cycle does not improve the inference (Washburn et al. 2015). For traits assumed to evolve via sequential stages, the accepted sequence of changes can be incorporated in the model (e.g., Marazzi et al. 2012). However, the power of character modeling remains inherently limited by the small number of informative characters. Decomposing the phenotype into its components can solve this problem, especially when the underlying genetic determinism is considered (Oliver et al. 2012; Niemiller et al. 2013; Glover et al. 2015; Meier et al. 2017), and good mechanistic models exist for the evolution of DNA sequences (Liberles et al. 2013). Violation of model assumptions can still mislead the conclusions, but the multiplication of sources of information, coupled with the possibility to track the history of specific genes independently of the species tree, limits the risks of systematic errors. We therefore suggest that efforts to reconstruct the transitions leading to important traits should integrate as many underlying components as possible. As progresses in genome biology increase data availability and improve our understanding of causal mutations, modeling phenotypes as the results of cumulative changes in genomes will be able to solve the problems raised by the paucity of informative characters.

Conclusions

In this study, we dissect the genetic and anatomical components of C_4 photosynthesis in *Alloteropsis*, a genus of grasses with multiple photosynthetic types. Our comparative efforts strongly support at least two independent origins of C_4 photosynthesis within this genus. The C_4 phenotype within these separate origins is realized via divergent anatomical modifications, the upregulation of distinct sets of genes, and independent enzyme adaptations. One of these lineages includes a range of photosynthetic types, and based on our analyses, we suggest that some C_4 components in this group evolved in the shared common ancestor, while others were acquired independently after the lineages diverged. The history of photosynthetic transitions within *Alloteropsis* is furthermore complicated by the introgression of C_4 genes across species boundaries. This disconnects the spread of C_4 components from the species tree, and means that the number of origins varies among the different components of the complex C_4 trait. This scenario is unlikely to have been inferred from traditional macroevolutionary approaches based on species trees alone. We suggest that integrating genomic data and phenotypic details in future studies of character transitions might resolve similarly complicated scenarios in other groups, enabling a better understanding of the trajectories followed during the evolution of novel adaptations.

AUTHOR CONTRIBUTIONS

PAC, CPO, PN, and EJE designed the study, MRL, MN, and PAC secured plant material, MRL generated the anatomical data, JJMV generated the transcriptome data, LTD, MRL, JJMV, and PAC analysed the data, LTD and PAC wrote the paper with the help of all authors.

ACKNOWLEDGMENTS

This work was funded by a Royal Society Research Grant (grant number RG130448) to PAC. PAC and PN are supported Royal Society University Research Fellowships (grant numbers URF120119 and URF130423, respectively), LTD is supported by a NERC grant (grant number NE/M00208X/1), and MRL is supported by an ERC grant (grant number ERC-2014-STG-638333).

DATA ARCHIVING

All raw RNA-Seq data have been deposited in the NCBI Sequence Read Archive (project identifier SRP072730), and transcriptome assemblies are deposited in the NCBI Transcriptome Shotgun Assembly repository (Bioproject PRJNA310121).

LITERATURE CITED

Atkinson, R. R., E. J. Mockford, C. Bennett, P. A. Christin, E. L. Spriggs, R. P. Freckleton, K. Thompson, M. Rees, and C. P. Osborne. 2016. C_4 photosynthesis boosts growth by altering physiology, allocation and size. Nat. Plants 2:16038.

Beaulieu, J. M., B. C. O'Meara, and M. J. Donoghue. 2013 Identifying hidden rate changes in the evolution of a binary morphological character: the evolution of plant habit in campanulid angiosperms. Syst. Biol. 62:725–737.

Bellasio, C., and H. Griffiths. 2014. The operation of two decarboxylases, transamination, and partitioning of C_4 metabolic processes between mesophyll and bundle sheath cells allows light capture to be balanced for the maize C_4 pathway. Plant Physiol. 164:466–480.

Besnard, G., A. M. Muasya, F. Russier, E. H. Roalson, N. Salamin, and P. A. Christin. 2009. Phylogenomics of C4 photosynthesis in sedges (Cyperaceae): multiple appearances and genetic convergence. Mol. Biol. Evol. 26:1909–1919.

Bläsing, O. E., P. Westhoff, and P. Svensson 2000. Evolution of C_4 phosphoenolpyruvate carboxylase in *Flaveria*, a conserved Serine residue in the carboxyl-terminal part of the enzyme is a major determinant for C_4-specific characteristics. J. Biol. Chem. 275:27917–27923.

Blount, Z. D., C. Z. Borland, and R. E. Lenski. 2008. Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A. 105:7899–7906.

Blount, Z. D., J. E. Barrick, C. J. Davidson, and R. E. Lenski. 2012. Genomic analysis of a key innovation in an experimental *Escherichia coli* population. Nature 489:13–518.

Bohley, K., O. Joos, H. Hartmann, R. F. Sage, S. Liede-Schumann, and G. Kadereit. 2015. Phylogeny of Sesuvioideae (Aizoaceae)–Biogeography, leaf anatomy and the evolution of C_4 photosynthesis. Perspect. Plant Ecol. Evol. Syst. 17:116–130.

Bräutigam, A., and U. Gowik. 2016. Photorespiration connects C_3 and C_4 photosynthesis. J. Exp. Bot. 67:2953–2962.

Bräutigam, A., S. Schliesky, C. Külahoglu, C. P. Osborne, and A. P. Weber. 2014. Towards an integrative model of C_4 photosynthetic subtypes: insights from comparative transcriptome analysis of NAD-ME, NADP-ME, and PEP-CK C_4 species. J. Exp. Bot. 65:3579–3593.

Brown, W. V. 1975. Variations in anatomy, associations, and origins of Kranz tissue. Am. J. Bot. 62:395–402.

Brown, N. J., C. A. Newell, S. Stanley, J. E. Chen, A. J. Perrin, K. Kajala, and J. M. Hibberd. 2011. Independent and parallel recruitment of preexisting mechanisms underlying C4 photosynthesis. Science 331:1436–1439.

Cantalapiedra, J. L., J. L. Prado, M. H. Fernández, and M. T. Alberdi. 2017. Decoupled ecomorphological evolution and diversification in Neogene-Quaternary horses. Science 355:627–630.

Christin, P. A., and G. Besnard. 2009. Two independent C_4 origins in Aristidoideae (Poaceae) revealed by the recruitment of distinct phosphoenolpyruvate carboxylase genes. Am. J. Bot. 96:2234–2239.

Christin, P. A. and C. P. Osborne. 2014. The evolutionary ecology of C_4 plants. New Phytol. 204:765–781.

Christin, P. A., N. Salamin, V. Savolainen, M. R. Duvall, and G. Besnard. 2007. C4 photosynthesis evolved in grasses via parallel adaptive genetic changes. Curr. Biol. 17:1241–1247.

Christin, P. A., E. Samaritani, B. Petitpierre, N. Salamin, and G. Besnard. 2009. Evolutionary insights on C4 photosynthetic subtypes in grasses from genomics and phylogenetics. Genome Biol. Evol. 1:221–230.

Christin, P. A., R. P. Freckleton, and C. P. Osborne. 2010. Can phylogenetics identify C_4 origins and reversals? Trends Ecol. Evol. 25:403–409.

Christin, P. A., T. L. Sage, E. J. Edwards, R. M. Ogburn, R. Khoshravesh, and R. F. Sage. 2011. Complex evolutionary transitions and the significance of C_3–C_4 intermediate forms of photosynthesis in Molluginaceae. Evolution 65:643–660.

Christin, P. A., E. J. Edwards, G. Besnard, S. F. Boxall, R. Gregory, E. A. Kellogg, J. Hartwell, and C. P. Osborne. 2012. Adaptive evolution of C_4 photosynthesis through recurrent lateral gene transfer. Curr. Biol. 22:445–449.

Christin, P. A., C. P. Osborne, D. A. Chatelet, J. T. Columbus, G. Besnard, T. R. Hodkinson, L. M. Garrison, M. S. Vorontsova, and E. J. Edwards. 2013a. Anatomical enablers and the evolution of C_4 photosynthesis in grasses. Proc. Natl. Acad. Sci. U. S. A. 110:1381–1386.

Christin, P. A., S. F. Boxall, R. Gregory, E. J. Edwards, J. Hartwell, and C. P. Osborne. 2013b. Parallel recruitment of multiple genes into C_4 photosynthesis. Genome Biol. Evol. 5:2174–2187.

Christin, P. A., E. Spriggs, C. P. Osborne, C. A. Strömberg, N. Salamin, and E. J. Edwards. 2014. Molecular dating, evolutionary rates, and the age of the grasses. Syst. Biol. 63:153–165.

Christin, P. A., M. Arakaki, C. P. Osborne, and E. J. Edwards. 2015. Genetic enablers underlying the clustered evolutionary origins of C_4 photosynthesis in angiosperms. Mol. Biol. Evol. 32:846–858.

Cooney, C. R., J. A. Bright, E. J. Capp, A. M. Chira, E. C. Hughes, C. J. Moody, L. O. Nouri, Z. K. Varley and G. H. Thomas. 2017. Mega-evolutionary dynamics of the adaptive radiation of birds. Nature 542:344–347.

Danforth, B. N., L. Conway, and S. Ji. 2003. Phylogeny of eusocial Lasioglossum reveals multiple losses of eusociality within a primitively eusocial clade of bees (Hymenoptera: Halictidae). Syst. Biol. 52:23–36.

Danforth, B. N., S. Cardinal, C. Praz, E. A. B. Almeida, and D. Michez. 2013. The impact of molecular data on our understanding of bee phylogeny and evolution. Annu. Rev. Entomol. 58:57–78.

Ding, Z., S. Weissmann, M. Wang, B. Du, L. Huang, L. Wang, X. Tu, S. Zhong, C. Myers, T. P. Brutnell, et al. 2015. 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.

Drummond, A. J. and A. Rambaut. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7:214.

Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797.

Edwards, E. J., C. P. Osborne, C. A. E. Strömberg, S. A. Smith, and C4 Grasses Consortium. 2010. The origins of C_4 grasslands: integrating evolutionary and ecosystem science. Science 328:587–591

Ellis, R. P. 1974. The significance of the occurrence of both Kranz and non-Kranz leaf anatomy in the grass species Alloteropsis semialata. S. Afr. J. Sci. 70:169–173.

Fisher, A. E., L. A. McDade, C. A. Kiel, R. Khoshravesh, M. A. Johnson, M. Stata, T. L. Sage, and R. F. Sage. 2015. Evolutionary history of *Blepharis* (Acanthaceae) and the origin of C_4 photosynthesis in section *Acanthodium*. Int. J. Plant Sci. 176:770–790.

Fisher, A. E., K. M. Hasenstab, H. L. Bell, E. Blaine, A. L. Ingram, and J. T. Columbus. 2016. Evolutionary history of chloridoid grasses estimated from 122 nuclear loci. Mol. Phylogenet. Evol. 105:1–14.

Frean, M. L., D. Ariovich, and C. F. Cresswell. 1983. C3 and C4 photosynthetic and anatomical forms of *Alloteropsis semialata* (R. Br.) Hitchcock 2. A comparative investigation of leaf ultrastructure and distribution of Chlorenchyma in the two forms. Ann. Bot. 51:811–821.

Gamble, T., E. Greenbaum, T. R. Jackman, A. P. Russell, and A. M. Bauer. 2012. Repeated origin and loss of adhesive toepads in Geckos. PLoS One 7:e39429

Glover, B. J., C. A. Airoldi, S. F. Brockington, M. Fernandez-Mazuecos, C. Martinez-Perez, G. Mellers, E. Moyroud, and L. Taylor. 2015. How have advances in comparative floral development influenced our understanding of floral evolution? Int. J. Plant Sci. 176:307–323.

Goldberg, E. E., and B. Igić. 2008. On phylogenetic tests of irreversible evolution. Evolution 62:2727–2741.

Gowik, U., J. Burscheidt, M. Akyildiz, U. Schlue, M. Koczor, M. Streubel, and P. Westhoff. 2004. *cis*-Regulatory elements for mesophyll-specific gene expression in the C4 plant *Flaveria trinervia*, the promoter of the C4 phosphoenolpyruvate carboxylase gene. Plant Cell 16:1077–1090.

Gowik, U., A. Bräutigam, K. L. Weber, A. P. M. Weber, and P. Westhoff. 2011. 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.

Grass Phylogeny Working Group II. 2012. New grass phylogeny resolves deep evolutionary relationships and discovers C_4 origins. New Phytol. 193:304–312.

Guindon, S., and O. Gascuel. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52:696–704.

Halliday, T. J. D., P. Upchurch, and A. Goswami. 2016. Eutherians experienced elevated evolutionary rates in the immediate aftermath of the Cretaceous–Palaeogene mass extinction. Proc. R. Soc. B 283:20153026.

Hancock, L., and E. J. Edwards, 2014. Phylogeny and the inference of evolutionary trajectories. J. Exp. Bot. 65:3491–3498.

Hatch, M. D. 1987. C_4 photosynthesis: a unique blend of modified biochemistry, anatomy and ultrastructure. Biochim. Biophys. Acta 895:81–106.

Heckmann, D., S. Schulze, A. Denton, U. Gowik, P. Westhoff, A. P. M. Weber, and M. J. Lercher. 2013. Predicting C_4 photosynthesis evolution: modular, individually adaptive steps on a Mount Fuji fitness landscape. Cell 153:1579–1588.

Huang, P., A. J. Studer, J. C. Schnable, E. A. Kellogg, and T. P. Brutnell. 2017. Cross species selection scans identify components of C4 photosynthesis in the grasses. J. Exp. Botany 68:127–135.

Hylton, C. M., S. Rawsthorne, A. M. Smith, D. A. Jones, and H. W. Woolhouse. 1988. Glycine decarboxylase is confined to the bundle-sheath cells of leaves of C_3-C_4 intermediate species. Planta 175:452–459.

Ibrahim, D. G., T. Burke, B. S. Ripley, and C. P. Osborne 2009. A molecular phylogeny of the genus *Alloteropsis* (Panicoideae, Poaceae) suggests an evolutionary reversion from C_4 to C_3 photosynthesis. Ann. Bot. 103:127–136.

Igic, B., and J. W. Busch. 2013. Is self-fertilization an evolutionary dead end? New Phytol. 198:386–397.

Igic, B., L. Bohs, and J. R. Kohn. 2006. Ancient polymorphism reveals unidirectional breeding system shifts. Proc. Natl. Acad. Sci. U. S. A. 103:1359–1363.

Kadereit, G., M. Lauterbach, M. D. Pirie, A. Rami, and F. Helmut. 2014. When do different C_4 leaf anatomies indicate independent C_4 origins? Parallel evolution of C_4 leaf types in Camphorosmeae (Chenopodiaceae). J. Exp. Bot. 65:3499–3511.

Kellogg, E. A. 1999. Phylogenetic aspects of the evolution of C_4 photosynthesis. Pp. 411–444 in R. F. Sage and R. K. Monson, eds. C_4 plant biology. Academic Press, San Diego, CA

King, B., and M. S. Y. Lee. 2015. Ancestral state reconstruction, rate heterogeneity, and the evolution of reptile viviparity. Syst. Biol. 64:532–544.

Külahoglu, C., A. K. Denton, M. Sommer, J. Mass, S. Schliesky, T. J. Wrobel, B. Berckmans, E. Gongora-Castillo, C. R. Buell, R. Simon, et al. 2014. Comparative transcriptome atlases reveal altered gene expression modules between two Cleomaceae C_3 and C_4 plant species. Plant Cell 26:3243–3260.

Li, F. W., J. C. Villarreal, S. Kelly, C. J. Rothfels, M. Melkonian, E. Frangedakis, M. Ruhsam, E. M. Sigel, J. P. Der, J. Pittermann, and D. O. Burge. 2014. Horizontal transfer of an adaptive chimeric photoreceptor from bryophytes to ferns. PNAS 111:6672–6677.

Li, G., B. W. Davis, E. Eizirik, and W. J. Murphy. 2016. Phylogenomic evidence for ancient hybridization in the genomes of living cats (Felidae). Genome Res. 26:1–11.

Li, Y., X. Ma, J. Zhao, J. Xu, J. Shi, X. G. Zhu, Y. Zhao, and H. Zhang. 2015. Developmental genetic mechanisms of C_4 syndrome based on transcriptome analysis of C_3 cotyledons and C_4 assimilating shoots in *Haloxylon ammodendron*. Plos One 10:e0117175.

Liberles, D. A., A. I. Teufel, L. Liu, and T. Stadler. 2013. On the need for mechanistic models in computational genomics and metagenomics. Genome Biol. Evol. 5:2008–2018.

Lundgren, M. R., C. P. Osborne, and P. A. Christin. 2014. Deconstructing Kranz anatomy to understand C4 evolution. J. Exp. Bot. 65:3357–3369.

Lundgren, M. R., G. Besnard, B. S. Ripley, C. E. R. Lehmann, D. S. Chatelet, R. G. Kynast, M. Namaganda, M. S. Vorontsova, R. C. Hall, J. Elia, et al. 2015. Photosynthetic innovation broadens the niche within a single species. Ecol. Lett. 18:1021–1029.

Lundgren, M. R., P. A. Christin, E. Gonzalez Escobar, B. S. Ripley, G. Besnard, C. M. Long, P. W. Hattersley, R. P. Ellis, R. C. Leegood, and C. P. Osborne. 2016. Evolutionary implications of C_3-C_4 intermediates in the grass *Alloteropsis semialata*. Plant Cell Environ. 39:1874–85

Maddison, W. P. 2006. Confounding asymmetries in evolutionary diversification and character change. Evolution 60:1743–1746.

Mallmann, J., D. Heckmann, A. Bräutigam, M. J. Lercher, A. P. M. Weber, P. Westhoff, and U. Gowik. 2014. The role of photorespiration during the evolution of C_4 photosynthesis in the genus *Flaveria*. Elife 3:e02478.

Marazzi, B., C. Ané, M. F. Simon, A. Delgado-Salinas, M. Luckow, and M. J. Sanderson. 2012. Locating evolutionary precursors on a phylogenetic tree. Evolution 66:3918–3930.

Marcussen, T., S. R. Sandve, L. Heirer, M. Spannagl, M. Pfeifer, The International What Genome Sequencing Consortium, K. S. Jakobsen, B. B. H. Wulff, B. Steuernagel, K. F. X. Mayer, O. A. Olsen. 2014. Ancient hybridizations among the ancestral genomes of bread wheat. Science 345:1250092.

McGuire, J. A., C. C. Witt, J. V. Remsen, A. Corl, D. L. Rabosky, D. L. Altshuler, and R. Dudley. 2014. Molecular phylogenetics and the diversification of hummingbirds. Curr. Biol. 24:910–916.

McKown, A. D., and N. G. Dengler. 2007. Key innovations in the evolution of Kranz anatomy and C_4 vein pattern in *Flaveria* (Asteraceae). Am. J. Bot. 94:382–399.

Meier, J. I., D. A. Marques, S. Mwaiko, C. E. Wagner, L. Excoffier, and O. Seehausen. 2017. Ancient hybridization fuels rapid cichlid fish adaptive radiations. Nat. Commun. 8:14363.

Moreau, C. S., and C. D. Bell. 2013. Testing the museum versus cradle tropical biological diversity hypothesis: phylogeny, diversification, and ancestral biogeographic range evolution of the ants. Evolution 67:2240–2257.

Niemiller, M. L., B. M. Fitzpatrick, P. Shah, L. Schmitz, T. J. Near. 2013. Evidence for repeated loss of selective constraint in rhodopsin in amblyopsid cavefishes (Teleostei: Amblyopsidae). Evolution 67:732–748.

Oliver, J. C., X. L. Tong, L. F. Gall, W. H. Piel, and A. Monteiro. 2012. A single origin for nymphalid butterfly eyespots followed by widespread loss of associated gene expression. PLoS Genet. 8:e1002893.

Olofsson, J. K., M. Bianconi, G. Besnard, L. T. Dunning, M. R. Lundgren, H. Holota, M. S. Vorontsova, O. Hidalgo, I. J. Leitch, P. Nosil, C. P. Osborne, and P. A. Christin. 2016. Genome biogeography reveals the intraspecific spread of adaptive mutations for a complex traitptive mutations for a complex trait. Mol. Ecol. 25:6107–6123

Pagel, M. 1999. Inferring the historical patterns of biological evolution. Nature 401:877–884.

———. 2004. Limpets break Dollo's Law. Trends Ecol. Evol. 19:278–280.

Paradis, E., J. Claude, and K. Strimmer. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20:289–290.

Pardo-Diaz, C., C. Salazar, S. W. Baxter, C. Merot, W. Figueiredo-Ready, M. Joron, W. O. McMillan, and C. D. Jiggins. 2012. Adaptive introgression across species boundaries in Heliconius butterflies. PLoS Genet. 8:e1002752.

Prendergast, H. D. V., P. W. Hattersley, and N. E. Stone. 1987. New structural/biochemical associations in leaf blades of C_4 grasses (Poaceae). Funct. Plant Biol. 14:403–420.

Protas, M. E., C. Hersey, D. Kochanek, Y. Zhou, H. Wilkens, W. R. Jeffery, L. I. Zon, R. Borowsky, and C. J. Tabin. 2006. Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. Nature Genet. 38:107–111.

Pyron, R. A., and F. T. Burbink. 2014. Early origin of viviparity and multiple reversions to oviparity in squamate reptiles. Ecol. Lett. 17:13–21.

Rabosky, D. L., F. Santini, J. Eastman, S. A. Smith, B. Sidlauskas, J. Chang, and M. E. Alfaro. 2013. Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. Nat. Commun. 6:4.

Renvoize, S. A. 1987. A survey of leaf-blade anatomy in grasses XI. Paniceae. Kew Bull. 739–768.

Sage, R. F., P. A. Christin, and E. J. Edwards. 2011. The C_4 plant lineages of planet Earth. J. Exp. Bot. 62:3155–3169.

Sage, R. F., T. L. Sage, and F. Kocacinar. 2012. Photorespiration and the evolution of C_4 photosynthesis. Annu. Rev. Plant Biol. 63:19–47.

Schneider, C. A., W. S. Rasband, and K. W. Eliceiri. 2012. NIH Image to ImageJ: 25 years of image analysis. Nat. Methods 9:671–675.

Shimizu, K. S., R. Shimizu-Inatsugi, T. Tsuchimatsu, and M. D. Purugganan. 2008. Independent origins of self-incompatibility in *Arabidopsis thaliana*. Mol. Ecol. 17:704–714.

Sinha, N. R., and E. A. Kellogg. 1996. Parallelism and diversity in multiple origins of C_4 photosynthesis in the grass family. Am. J. Bot. 83:1458–1470.

Smith, J., and M. R. Kronforst. 2013. Do *Heliconius* butterfly species exchange mimicry alleles?. Biol. Lett. 9:20130503.

Soros, C. L., and N. G. Dengler. 2001. Ontogenetic derivation and cell differentiation in photosynthetic tissues of C_3 and C_4 Cyperaceae. Am. J. Bot. 88:992–1005.

Trueman, J. W. H., B. E. Pfeil, S. A. Kelchner, and D. K. Yeates. 2004. Did stick insects really regain their wings?. Syst. Entomol. 29:138–139.

Ueno, O., and N. Sentoku. 2006. Comparison of leaf structure and photosynthetic characteristics of C_3 and C_4 *Alloteropsis semialata* subspecies. Plant Cell Environ. 29:257–268.

Vilella, A. J., J. Severin, A. Ureta-Vidal, L. Heng, R. Durbin, and E. Birney, 2009. EnsemblCompara GeneTrees: complete, duplication-aware phylogenetic trees in vertebrates. Genome Res. 19:327–335.

von Caemmerer, S., and R. T. Furbank. 2003. The C_4 pathway: an efficient CO_2 pump. Photosynth. Res. 77:191–207.

Wang, X., U. Gowik, H. Tang, J. E. Bowers, P. Westhoff, and A. H. Paterson. 2009. Comparative genomic analysis of C4 photosynthetic pathway evolution in grasses. Genome Biol. 10:R68.

Wang, Y., A. Bräutigam, A. P. M. Weber, and X. G. Zhu. 2014. Three distinct biochemical subtypes of C_4 photosynthesis? A modelling analysis. J. Exp. Bot. 65:3579–3593.

Washburn, J. D., J. C. Schnable, G. Davidse, and J. C. Pires. 2015. Phylogeny and photosynthesis of the grass tribe Paniceae. Am. J. Bot. 102:1493–1505.

Werner G. D. A., W. K. Cornwell, J. I. Sprent, J. Kattge, and E. T. Kiers. 2014. A single evolutionary innovation drives the deep evolution of symbiotic N_2-fixation in angiosperms. Nature Commun 5:4087.

Whiting, M. F., S. Bradler, and T. Maxwell. 2002. Loss and recovery of wings in stick insects. Nature 421:264–267.

Wiens, J. J. 1999. Phylogenetic evidence for multiple losses of a sexually selected character in phrynosomatid lizards. Proc. R. Soc. B 266:1529–1535.

Wiens, J. J., M. C. Brandley, and T. W. Reeder. 2006. Why does a trait evolve multiple times within a clade? Repeated evolution of snakelike body form in squamate reptiles. Evolution 60:123–141.

Williams, B. P., I. G. Johnston, S. Covshoff, and J. M. Hibberd. 2013. Phenotypic landscape inference reveals multiple evolutionary paths to C_4 photosynthesis. Elife 2:e00961.

Williams, B. P., S. J. Burgess, I. Reyna-Llorens, J. Knerova, S. Aubry, S. Stanley, and J. M. Hibberd. 2016. An untranslated cis-element regulates the accumulation of multiple C_4 enzymes in *Gynandropsis gynandra* mesophyll cells. Plant Cell 28:454–465.

Yang, Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24:1586–1591.

Associate Editor: M. Streisfeld
Handling Editor: P. Tiffin

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Figure S1. Comparisons of leaf anatomy in Alloteropsis and relatives.
Figure S2. Phylogenetic trees for genes encoding three C4-related enzymes.
Figure S3. Phylogeny of pck-1P1 genes in Panicoideae.
Figure S4. Evolution of ppc-1P3 genes in Alloteropsis and other Panicoideae.
Figure S5. Evolution of ppdk-1P2 genes in Alloteropsis and other Panicoideae.
Figure S6. Evolution of aspat-3P4 genes in Alloteropsis and other Panicoideae.
Table S1. Alloteropsis semialata accessions used in this study.
Table S2. Leaf anatomical data for the study species and accessions.
Table S3. RNA-Seq data, NCBI SRA accession numbers, and growth conditions.
Table S4. Transcript abundance (in rpkm) for each C4-related gene and sample.
Table S5. Results of positive selection analyses inferring the episodes of enzymatic adaptation in Alloteropsis using only fixed differences.
Table S6. Results of positive selection analyses inferring the episodes of enzymatic adaptation in the *A. angusta*/*A. semialata* clade using only fixed differences.
Methods 1.1. Plant growth conditions.
Methods 1.2. RNA-Seq protocol.
Methods 1.3. Positive selection analysis.
Methods 1.4. Alignment and filtering