Lateral Gene Transfer Acts As an Evolutionary Shortcut to Efficient C$_4$ Biochemistry

Chatchawal Phansopa,[1,2] Luke T Dunning,[1] James D Reid,[2] and Pascal-Antoine Christin*,[1]

[1]Department of Animal and Plant Sciences, University of Sheffield, Sheffield, United Kingdom
[2]Department of Chemistry, University of Sheffield, Sheffield, United Kingdom

*Corresponding author: E-mail: p.christin@sheffield.ac.uk.
Associate editor: Juliette de Meaux

Abstract

The adaptation of proteins for novel functions often requires changes in their kinetics via amino acid replacement. This process can require multiple mutations, and therefore extended periods of selection. The transfer of genes among distinct species might speed up the process, by providing proteins already adapted for the novel function. However, this hypothesis remains untested in multicellular eukaryotes. The grass *Alloteropsis* is an ideal system to test this hypothesis due to its diversity of genes encoding phosphoenolpyruvate carboxylase, an enzyme that catalyzes one of the key reactions in the C$_4$ pathway. Different accessions of *Alloteropsis* either use native isoforms relatively recently co-opted from other functions or isoforms that were laterally acquired from distantly related species that evolved the C$_4$ trait much earlier. By comparing the enzyme kinetics, we show that native isoforms with few amino acid replacements have substrate K_M values similar to the non-C$_4$ ancestral form, but exhibit marked increases in catalytic efficiency. The co-option of native isoforms was therefore followed by rapid catalytic improvements, which appear to rely on standing genetic variation observed within one species. Native C$_4$ isoforms with more amino acid replacements exhibit additional changes in affinities, suggesting that the initial catalytic improvements are followed by gradual modifications. Finally, laterally acquired genes show both strong increases in catalytic efficiency and important changes in substrate handling. We conclude that the transfer of genes among distant species sharing the same physiological novelty creates an evolutionary shortcut toward more efficient enzymes, effectively accelerating evolution.

Key words: adaptation, C$_4$ photosynthesis, horizontal gene transfer, positive selection, Poaceae.

Introduction

The evolution of novel traits usually involves the co-option of preexisting genes, which were previously used for different functions (True and Carroll 2002; Jiggins et al. 2017; Fernández and Gabaldón 2020). These genes are often subsequently modified in terms of their expression pattern and/or properties of the encoded enzymes, the extent of which depends on the strength of selection (Toprak et al. 2012; Karageorgi et al. 2019). Mutations required to trigger certain new functions are often restricted to a subset of codon positions, and epistasis can restrict the order in which they can occur (Weinreich et al. 2006; Blount et al. 2012; Studer, Christin, et al. 2014; Kumar et al. 2017; Yang et al. 2019). Because of these complexities, the modification of genes for a new function can require protracted periods of selection, the length of which depends on the mutation rate and demography of the species (Desai et al. 2007; Neher et al. 2010). The transfer of genes among species, via hybridization or lateral gene transfer (LGT), can bypass these extended periods of gradual evolution and boost evolutionary innovation (Ochman et al. 2000; Jain et al. 2003; Arnold and Kunte 2017; Hall et al. 2017). However, the impact of interspecific gene transfer on the speed of adaptation is difficult to directly compare with the iterative adaptation of co-opted native genes in complex multicellular organisms.

C$_4$ photosynthesis offers a tractable system to study the evolutionary paths to new functions. This complex trait, which combines anatomical and biochemical modifications to increase productivity in tropical conditions (Hatch 1987; Atkinson et al. 2016), has evolved >60 times independently in flowering plants (Sage et al. 2011, 2012). All known C$_4$ genes were present in the non-C$_4$ ancestors, and their co-option involved a massive increase in their expression in specific leaf compartments, followed in some cases by kinetic adaptation of the encoded enzymes (Tausta et al. 2002; Engelmann et al. 2003; Gowik et al. 2004; Tanz et al. 2009; Aubry et al. 2011; Moreno-Villena et al. 2018; Alvarez et al. 2019; DiMario and Cousins 2019). In particular, the key C$_4$ enzyme phosphoenolpyruvate carboxylase (PEPC) is highly expressed in all C$_4$ plants, and the C$_4$ forms of this enzyme differ from their non-C$_4$ homologs in their affinities for the substrates as well as their sensitivity to inhibitors (Ting and Osmond 1973; Bauwe and Chollet 1986; Svensson et al. 1997; Gowik et al. 2006; Paulus et al. 2013; DiMario and Cousins 2019). Phylogeny-based sequence comparisons have shown that C$_4$-specific genes for PEPC underwent numerous adaptive
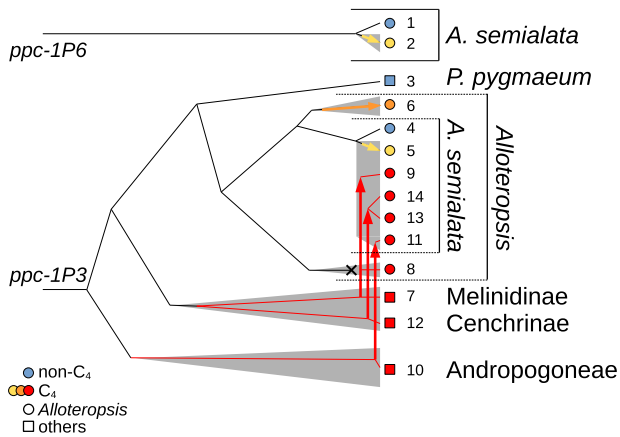
FIG. 1. History of genes encoding PEPC in *Alloteropsis*. This schematic shows previously inferred relationships among the genes for PEPC analyzed (Christin, Edwards, et al. 2012; Dunning et al. 2017). Branching depths are proportional to estimated divergence times. C_4 lineages are represented by gray areas, and arrows pointing to the tips represent modifications for the C_4 function, whereas vertical arrows indicate interspecific gene transfers. The native copy of accession 8 (*A. cimicina*) was pseudogenized, an event indicated with a cross. Genes are numbered as in table 1.

amino changes that were repeated among distant lineages (Christin et al. 2007; Besnard et al. 2009; Paulus et al. 2013; Rosnow et al. 2015). Although the kinetic effects of these mutations remain generally unknown (for exceptions, see Bläsing et al. 2000; Paulus et al. 2013; DiMario and Cousins 2019), the convergence of these C_4-related mutations suggests that the adaptation of PEPC for the C_4 context is similarly constrained in divergent C_4 lineages. Importantly, although most C_4-specific PEPCs originated via novel mutations that followed the co-option of native non-C_4 genes, several instances of interspecific transfers of C_4 PEPC have been reported (Besnard et al. 2009; Christin, Edwards, et al. 2012; Christin, Wallace, et al. 2012).

In grasses, the genus *Alloteropsis* includes plants that use C_4 photosynthesis and others that lack the trait, sometimes within the same species (Ibrahim et al. 2009; Dunning et al. 2017). The C_4 accessions of *Alloteropsis* use various PEPC genes for their C_4 pathway, some of which were co-opted from other functions, whereas others were laterally acquired from distant C_4 lineages (fig. 1; Christin, Edwards, et al. 2012; Dunning et al. 2017). Two different native non-C_4 PEPC genes were co-opted by geographically isolated populations of *Alloteropsis semialata* (fig. 1 and table 1), which have undergone relatively few modifications since the trait evolved as evidenced by their high similarity to PEPC orthologs from non-C_4 *A. semialata*, and a lack of the convergent amino acid replacements observed in older C_4 lineages (Christin, Edwards, et al. 2012; Dunning et al. 2017). The sister species *A. angusta*, which likely evolved the C_4 trait earlier, uses a native gene for PEPC co-opted from other functions that has undergone more amino acid replacements (fig. 1; Christin, Edwards, et al. 2012; Dunning et al. 2017). By contrast, several populations of *A. semialata* and *A. cimicina* use one of three PEPC genes that were laterally acquired from

distantly related C_4 lineages and likely replaced the co-opted native copies (fig. 1; Christin, Edwards, et al. 2012; Dunning et al. 2017). Because these other genes had spent millions of years within C_4 plants before the transfer (fig. 1), they had been adapted for the C_4 context (Christin et al. 2007; Christin, Edwards, et al. 2012). The unrivalled diversity of PEPC isoforms in *Alloteropsis* offers a unique opportunity to assess the biochemical changes conferred by interspecific transfers as opposed to adapting co-opted native genes.

In this work, we test the hypothesis that interspecific gene transfer provides an evolutionary shortcut to gene adaptations that would otherwise be achieved after a long period of selection on novel mutations. First, we establish the evolutionary trajectory of co-opted native PEPC enzymes within *Alloteropsis* by comparing the PEPC proteins of non-C_4 and C_4 accessions without any LGT PEPC. Second, we characterize genes from older C_4 lineages that have numerous amino acid changes to test the hypothesis that they encode enzymes with drastically altered biochemical phenotypes when compared with non-C_4 ancestors. Finally, we compare the properties of the enzymes encoded by the laterally acquired genes of *Alloteropsis* with the native copies of both *Alloteropsis* and the donor groups, to determine whether the transfers provided an evolutionary shortcut, and whether any further modifications of the kinetic properties happened after the transfers. Coupled with phylogenetic analyses of coding sequences, this work provides new insights into the evolutionary paths to new biochemical functions in plants, and the impact of gene transfers on physiological adaptations.

Results

Phylogenetic Analyses Confirm Different Amounts of Amino Acid Changes

Genes from *Alloteropsis* were placed within the six distinct lineages of *ppc-1* as expected (supplementary fig. S1, Supplementary Material online; Dunning et al. 2017). The phylogeny inferred from *ppc-1P6* matched the species tree, with *A. angusta* genes sister to *A. semialata*, and the non-C_4 individuals branching first within *A. semialata* (supplementary fig. S1, Supplementary Material online). Most amino acid replacements occurred on the two branches leading to groups of C_4 *A. semialata*, one of which encompasses mainly Asian accessions, whereas the other one includes only African accessions. Many of these genes are pseudogenes, as evidenced by mutations disrupting the reading frame (supplementary fig. S1, Supplementary Material online). However, functional copies are detected in the individuals previously shown to use these genes for their C_4 pathway (i.e., TPE1-10, BUR1-02, and RSA4-01; Dunning et al. 2017). The cloned variants of the C_4 (from TPE1-10) and non-C_4 (from RSA5-03) forms of *ppc-1P6* differ by 13 amino acids (table 2), and in four cases, the C_4 form harbors the ancestral residue as observed in other non-C_4 species (sites 51, 280, 486, and 526; fig. 2). Of the nine replacements that represent novel mutations in the C_4, only one is fixed among C_4 accessions (site 78; fig. 2).

The phylogeny based on the native copy of *ppc-1P3* also recovered the expected relationships among species and

Table 1. Isoform Sampling and Gene Information.

Index	Species	Accession[a]	Gene Lineage[b]	Category	Source
1	*Alloteropsis semialata*	RSA5-03	*ppc-1P6*	Non-C$_4$	Synthesized
2	*Alloteropsis semialata*	TPE1-10	*ppc-1P6*	Native co-opted 1	Isolated[c]
3	*Panicum pygmaeum*	—	*ppc-1P3*	Non-C$_4$	Moody et al. (forthcoming)
4	*Alloteropsis semialata*	RSA5-03	*ppc-1P3*	Non-C$_4$	Synthesized
5	*Alloteropsis semialata*	MAD1-03	*ppc-1P3*	Native co-opted 2	Isolated[d]
6	*Alloteropsis angusta*	AANG4-8	*ppc-1P3*	Native co-opted 2	Synthesized
7	*Megahyrsus maximus*	—	*ppc-1P3_M*	Donor_M	Synthesized
8	*Alloteropsis cimicina*	—	*ppc-1P3_LGT:M*	LGT:M	Isolated[e]
9	*Alloteropsis semialata*	TAN4-08	*ppc-1P3_LGT:M*	LGT:M	Synthesized
10	*Themeda triandra*	—	*ppc-1P3_A*	Donor_A	Synthesized
11	*Alloteropsis semialata*	AUS1-01	*ppc-1P3_LGT:A*	LGT:A	Synthesized
12	*Setaria barbata*	—	*ppc-1P3_C*	Donor_C	Synthesized
13	*Alloteropsis semialata*	RSA3-01	*ppc-1P3_C*	LGT:C	Isolated[f]
14	*Alloteropsis semialata*	RSA4-01	*ppc-1P3_LGT:C*	LGT:C	Isolated[f]

[a]Accession names as in Dunning, Olofsson, et al. (2019).
[b]Genes named as in Bianconi et al. (2018) (M, Melinidinae; C, Cenchrinae; A, Andropogoneae).
[c]5′ primer = AATAGCTAGCATGGCGGGGAAG (*Nhe*I), 3′ primer = AATACTCGAGTTAACCAGTGTT (*Xho*I).
[d]5′ primer = AATACATATGGCGGCGTCC (*Nde*I), 3′ primer = AATAAAGCTTCTAGCCCGTGTT (*Hind*III).
[e]5′ primer = AATACATATGGCGACCTCG (*Nde*I), 3′ primer = AATAGCGGCCGCCTAGCCCGTGTT (*Not*I).
[f]5′ primer = AATACATATGGCGGAGAAG (*Nde*I), 3′ primer = AATAGCTAGCTAGCCAGTGTT (*Nhe*I).

Table 2. Pairwise Number of Amino Acid Differences between Analyzed Sequences[a].

Category	Isoform	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Non-C$_4$	1. RSA5-03	0													
Native co-opted 1	2. TPE1-10	13	0												
Non-C$_4$	3. *Panicum pygmaeum*	154	153	0											
Non-C$_4$	4. RSA5-03	161	161	20	0										
Native co-opted 2	5. MAD1-03	164	164	29	17	0									
Native co-opted 2	6. *Alloteropsis angusta*	187	185	69	59	64	0								
Donor_M	7. *Megahyrsus maximus*	224	223	148	149	147	152	0							
LGT:M	8. *Alloteropsis cimicina*	218	217	142	143	137	147	39	0						
LGT:M	9. TAN4-08	221	220	146	147	140	147	47	22	0					
Donor_A	10. *Themeda triandra*	234	234	180	185	184	178	190	188	190	0				
LGT:A	11. AUS1-01	236	235	181	186	185	179	191	189	191	3	0			
Donor_C	12. *Setaria barbata*	218	217	147	148	151	148	153	159	162	190	193	0		
LGT:C	13. RSA3-01	220	219	149	150	153	150	155	161	164	192	195	2	0	
LGT:C	14. RSA4-01	220	219	149	150	153	150	155	161	164	192	195	2	0	0

[a]Sequences are numbered as in table 1; differences between sequences belonging to the same group (see supplementary fig. S1, Supplementary Material online) are shaded in gray.

accessions. An abundance of amino acid replacements occurred on the branch leading to the *A. cimicina* gene, which is a pseudogene (supplementary fig. S1, Supplementary Material online), and to a lesser extent on the branch leading to *A. angusta* genes, which are functional and used by this species for the C$_4$ pathway (Dunning et al. 2017). Within *A. semialata*, many *ppc-1P3* genes from C$_4$ accessions are pseudogenes, and few amino acid mutations are observed (table 2), mainly on branches leading to genes used by some C$_4$ accessions (e.g., MAD1-03, TPE1-10, and BUR1-02; Dunning et al. 2017). The cloned variants of the C$_4$ (from MAD1-03) and non-C$_4$ (from RSA5-03) genes differ by a total of three amino acid deletions and 17 amino acid substitutions, four of which represent novel mutations in the non-C$_4$ form (sites 628, 708, 715, and 955) and an extra two sites are variable among non-C$_4$ accessions (sites 35 and 567, fig. 2). All of the 11 sites representing new mutations in the C$_4$ forms are polymorphic among C$_4$ accessions (fig. 2), and in many cases within individuals. Three of these 11 amino acid substitutions

are also observed in *A. angusta* (sites 18, 320, and 369), but many more substitutions occurred in this species (table 2). Indeed, the cloned C$_4$ gene from *A. angusta* differs from the cloned non-C$_4$ variant from *A. semialata* by 59 amino acid substitutions, one insertion and one deletion (supplementary data set 1, Supplementary Material online). Nine of the amino acid residues specific to the C$_4$ form of *A. angusta* are among the 21 previously reported as convergent among C$_4$ lineages of grasses (positions 531, 577, 579, 780, 794, 572, 813, 502, 665; Christin et al. 2007; Christin, Edwards, et al. 2012).

The close relationships between genes laterally acquired by *Alloteropsis* and some other groups of grasses are confirmed (supplementary fig. S1, Supplementary Material online). The *ppc-1P3_LGT:C* gene of *A. semialata* is almost identical to that of *Setaria barbata* (two differences between the cloned genes; table 2), with very few amino acid differences among *A. semialata* accessions (supplementary fig. S1, Supplementary Material online). A great similarity is also observed between the *ppc-1P3_LGT:A* gene of *A. semialata* and

ppc-1P6

		51	63*	78*	280	365*	386*	486	526	545*	578*	731*	856*	920*
non-C4	other species	D	E	E	A	V	R	R	A	A	L	I	L	T
non-C4	A. semialata others	DF	E	N	AT	V	R	RC	AT	A	L	I	L	T
non-C4	A. semialata RSA5-3 (1)	F	E	N	T	V	R	C	T	A	L	I	L	T
weak C4	A. semialata others	D	E	D	A	VGA	R	R	AT	A	VL	I	L	T
C4	A. semialata TPE1-10 (2)	D	V	D	A	G	H	R	A	G	V	T	F	M
C4	A. semialata others	D	EV	D	AS	VG	HR	R	AT	GA	VML	TI	FL	MT

ppc-1P3

		9*	18*	24*	26*	35	42*	97*	320*	369*	567	628	708	715	751*	845*	884*	955
non-C4	other species	?	A	A	G	V	V	V	F	E	K	A	S	A	Y	D	D	L
non-C4	A. semialata others	P	A	A	G	VI	V	V	TF	E	KR	AS	SF	AK	Y	D	D	LV
non-C4	A. semialata RSA5-3 (4)	P	A	A	G	V	V	V	F	E	K	S	F	K	Y	D	D	Y
weak C4	A. semialata others	P	A	A	G	I	VI	VIT	FV	QE	K	A	S	A	Y	D	DG	L
C4	A. semialata MAD1-3 (5)	T	T	V	S	I	I	I	I	Q	R	A	S	A	F	E	G	L
C4	A. semialata others	TP	ATV	AV	GS	IV	VI	IVA	FIVS	QH	RK	A	S	AT	YFH	DEN	GD	LF

Fig. 2. Amino acid variation of genes co-opted by C_4 *Alloteropsis semialata*. For each of the two native gene lineages co-opted by *A. semialata* (*ppc1-P3* and *ppc-1P6*), the amino acid residues differing between the non-C_4 and C_4 cloned genes (names of accessions indicated with numbers in parentheses corresponding to those in fig. 1 and table 1) are shown, in blue for the non-C_4 and yellow for the C_4 forms. Homologous residues are reported in decreasing frequency for non-C_4 orthologous of *A. semialata* and other species, genes of *A. semialata* with a weak C_4 pathway (see Dunning et al. 2017), and other C_4 *A. semialata*. When fixed within a group, the residues are colored as the cloned gene presenting the same residue. Positions are indicated on the top, numbered based on *Zea mays* sequence CAA33317. Asterisks highlight positions with novel mutations in the C_4 group.

Themeda triandra (three amino acid differences between the cloned genes; table 2 and supplementary fig. S1, Supplementary Material online). By contrast, the *ppc-1P3_LGT:M* genes of *Alloteropsis* are relatively diverged from all sequences available for the group of donors, and highly divergent copies are observed within *A. cimicina* (table 2 and supplementary fig. S1, Supplementary Material online). Frequent amino acid replacement in the *ppc-1P3_LGT:M* genes also occurred within *A. semialata*, and although several copies are pseudogenes, functional versions are observed in accessions previously shown to use this gene for their C_4 pathway (e.g., BUR1-02 and TAN4-08; Dunning et al. 2017).

Gradual Modifications Following the Co-option of Native Genes

We cloned and synthesized proteins encoded by a total of 14 genes from *Alloteropsis* accessions and related grasses (table 1), which capture a diversity of origins of C_4 PEPC (fig. 1). The enzyme encoded by the non-C_4 *ppc-1P6* of *A. semialata* has a low K_M for both substrates (PEP and HCO_3^-) and a low k_{cat} (isoform 1 in fig. 3; supplementary table S1, Supplementary Material online). In comparison, the enzyme encoded by the co-opted native ortholog (isoform 2) has a decreased K_M(PEP), an increased K_M(HCO_3^-), and an increased k_{cat} (1.87-fold; fig. 3 and supplementary table S1, Supplementary Material online). The co-option of native *ppc-1P6* was therefore followed by an increased catalytic efficiency and small alterations of the K_m for each substrate. The non-C_4 enzyme encoded by *ppc-1P6* (isoform 1) showed the lowest sensitivities to both malate and aspartate

(two molecules that are produced downstream in the C_4 pathway) of all assayed enzymes, and the co-opted native enzyme (isoform 2) showed a markedly increased sensitivity to malate inhibition (fig. 4).

The enzymes encoded by the non-C_4 *ppc-1P3* of *A. semialata* and the close relative *Panicum pygmaeum* are kinetically very similar (isoforms 3 and 4, respectively; supplementary table S1, Supplementary Material online). They present the lowest k_{cat} and K_M(PEP) of all isoforms analyzed here, and rank among the lowest K_M(HCO_3^-) (fig. 3). In terms of kinetics, the enzyme encoded by non-C_4 *ppc-1P3* (isoforms 3 and 4) are similar to that encoded by non-C_4 *ppc-1P6* (isoform 1), despite >100 My of divergence and many amino acid differences (table 2). However, enzymes encoded by non-C_4 *ppc-1P3* (isoforms 3 and 4) and *ppc-1P6* (isoform 1) differ strongly in terms of their sensitivity to inhibitors, which exhibit the lowest and highest values, respectively (fig. 4). The enzyme encoded by the native *ppc-1P3* co-opted for C_4 photosynthesis by *A. semialata* (isoform 5) is very similar to those encoded by the non-C_4 orthologs (isoforms 3 and 4) in terms of K_M for both substrates, but has a markedly elevated k_{cat} (2.26-fold; fig. 3) and reduced sensitivities to both malate and aspartate (fig. 4). The k_{cat} and sensitivity to inhibition change in the same direction, but are more marked in the co-opted native form from *A. angusta* (isoform 6; figs. 3 and 4). However, the K_M(PEP) is ~1.7× larger in the co-opted native form from *A. angusta* (isoform 6) as compared with enzymes encoded by both C_4 and non-C_4 orthologs from *A. semialata* (isoforms 4 and 5, fig. 3). These results suggest that the co-option of native *ppc-1P3* was followed by rapid changes in k_{cat}
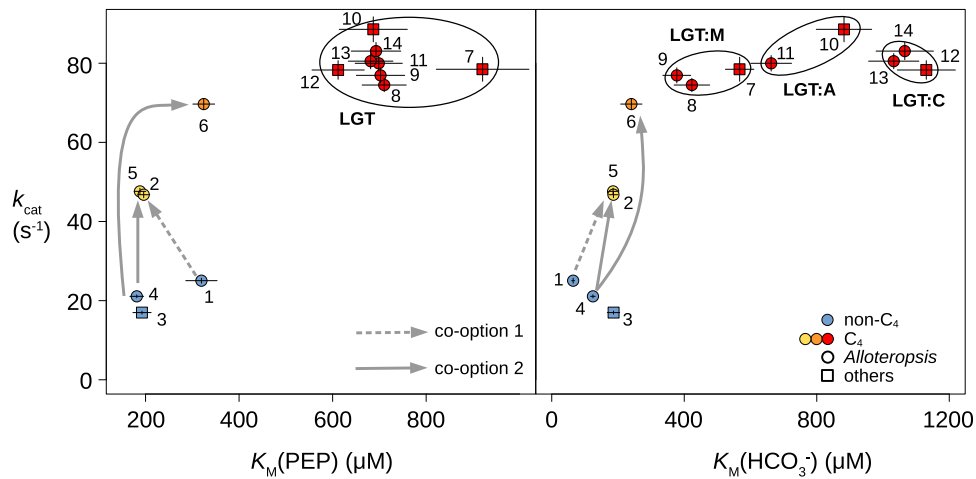
FIG. 3. Comparison of kinetic parameters. Measured values are shown for the 14 assayed enzymes, with error bars showing SDs. Changes following the co-option of native genes are indicated with gray arrows, and ellipses indicate genes involved in lateral gene transfers (LGT). Genes are numbered as in table 1.
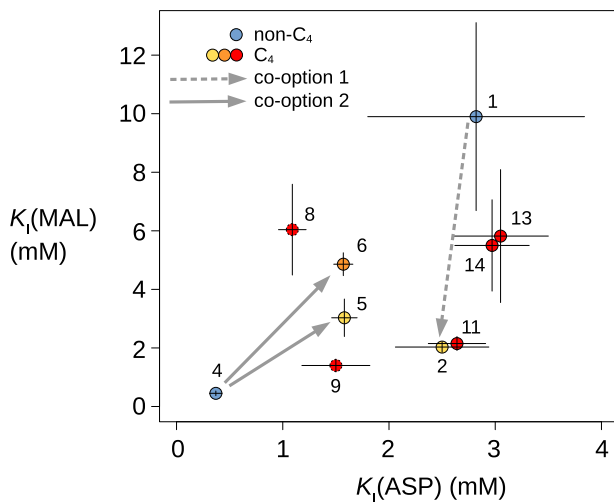
FIG. 4. Comparison of sensitivity to inhibitors. Measured values are shown for different forms of PEPC from *Alloteropsis*, with error bars showing SDs. Genes are numbered as in table 1.

and sensitivity to inhibition, and later by modifications of the $K_m(PEP)$.

Overall, enzymes encoded by the non-C_4 paralogs *ppc-1P3* (isoforms 3 and 4) and *ppc-1P6* (isoform 1) differ in their kinetic properties, as expected from their long divergence (fig. 1 and table 2). The changes consequently happened in slightly different directions after the co-option of the native *ppc-1P6* than following each co-option of native *ppc-1P3* (figs. 3 and 4). However, the kinetic parameters of the enzymes encoded by co-opted native *ppc-1P3* (isoform 5) and *ppc-1P6* (isoform 2) from *A. semialata* are almost identical (figs. 3 and 4), indicating rapid convergence.

Laterally Acquired Genes Are Highly Divergent from the Non-C_4 Forms

The three laterally acquired versions (isoforms 8 + 9, 11, and 13 + 14) are massively different from the native C_4 and non-C_4 enzymes (isoforms 1–6), but are similar to those of the

close relatives of the donors (isoforms 7, 10, and 12). All laterally acquired versions (isoforms 8, 9, 11, 13, and 14) have strikingly convergent k_{cat} and $K_M(PEP)$, the latter of which are 1.8/3.9-fold higher than those of the native versions (isoforms 1–6; fig. 3 and supplementary table S1, Supplementary Material online). Their $K_M(HCO_3^-)$ are more variable, but in all cases above those of the co-opted native isoforms (isoforms 2, 5, and 6), and each *A. semialata* copy clusters with its donor (fig. 3). The sensitivity to inhibitors of the laterally acquired isoforms overlaps with those of the co-opted native C_4 isoforms (fig. 4). Because the LGT replaced the co-opted native versions of *A. semialata* (isoforms 2 and 5; Olofsson et al. 2016; Dunning et al. 2017), the LGTs have led to a >1.5-fold increase of k_{cat}, a >3.1-fold increase of $K_M(PEP)$, and a >2-fold increase of $K_M(HCO_3^-)$, without consistent modifications of the sensitivity to inhibitors (figs. 3 and 4).

Discussion

Rapid Increase in Catalytic Efficiency after the Co-option of Native PEPC for C_4 Photosynthesis

Most C_4 lineages emerged between 5 and 30 Ma, so that the early events of the photosynthetic transitions are blurred by the accumulation of unrelated mutations (Heyduk et al. 2019). As a comparatively young C_4 lineage (<3 Ma; Lundgren et al. 2015), C_4 accessions of *A. semialata* represent an excellent system to pinpoint the exact modifications involved in the early emergence of a C_4 physiology, as previously applied to anatomical traits and gene expression (Dunning, Moreno-Villena, et al. 2019; Lundgren et al. 2019). In the case of PEPC, the non-C_4 enzymes encoded by *ppc-1P3* and *ppc-1P6* likely resemble the ancestral forms, as suggested by the limited number of amino acid changes in non-C_4 plants (table 2 and supplementary fig. S1, Supplementary Material online), and the catalytic similarity between enzymes encoded by the non-C_4 *ppc-1P3* of *A. semialata* and the more distantly related *P. pygmaeum* (figs. 3 and 4). The enzymes encoded by the non-C_4 paralogs vary in their kinetic properties (figs. 3 and 4), as expected given their long divergence time (near the

origin of monocots 140–160 Ma; Deng et al. 2016; Li et al. 2019). However, both isoforms present low K_M for the two substrates, as reported for other non-C_4 isoforms (Dong et al. 1998; Bläsing et al. 2002; Gowik et al. 2006). This might confer rapid responses to small increases of substrate and therefore a tight regulation of the non-C_4 function (O'Leary et al. 2011). Our comparative analyses show that the co-option of both native ppc-1P3 and ppc-1P6 was followed by swift changes to the catalytic efficiency and sensitivity to inhibitors, as observed in the C_4 A. semialata (figs. 3 and 4). Reduced inhibition by the products of PEPC is likely required to allow the enzyme to function in the high-flux C_4 pathway (Svensson et al. 1997, 2003), which leads to massively elevated concentrations of metabolites (Arrivault et al. 2017). Increased catalytic efficiency would directly impact the rate of the cycle providing a selective advantage to emerging C_4 plants (Heckmann et al. 2013).

Although the causal mutations are not known, the characterized C_4-specific native ppc-1P3 and ppc-1P6 of A. semialata differ from their respective non-C_4 orthologs by few amino acids (table 2), some of which are also observed among non-C_4 individuals, whereas almost all others are polymorphic within the C_4 group (fig. 2). This suggests that the C_4-specific properties might have emerged from standing genetic variation, after recombination generated amino acid combinations that altered the properties of the encoded enzyme in synergy. Many of the amino acid differences are moreover polymorphic within C_4 individuals (fig. 2), which suggests that this process is ongoing, potentially as part of the functional diversification of the multiple copies that exist within some of these plants (Bianconi et al. 2018).

Adaptation of the Protein Sequence Leads to Further Biochemical Changes

Alloteropsis angusta diverged from A. semialata ∼7 Ma (Lundgren et al. 2015; Dunning et al. 2017). Its native ppc-1P3 shows signs of positive selection (Dunning et al. 2017), and it presents some of the amino acids that convergently evolved in older C_4 lineages (Christin et al. 2007; Christin, Edwards, et al. 2012; Christin, Wallace, et al. 2012). This co-opted native gene can thus be considered as partially modified for the C_4 context. Because some of the amino acid differences between the native C_4 and non-C_4 isoforms of A. semialata are also observed in A. angusta, it is possible that the adaptation of A. angusta PEPC for the C_4 context initially followed the same path observed within A. semialata. In terms of enzyme phenotype, the C_4 form from A. angusta is even less sensitive to malate than its native C_4 ortholog from A. semialata (fig. 4). It moreover shows a higher catalytic efficiency (fig. 3), which suggests that initial large-effect changes as observed within A. semialata are then followed by further modifications in the same direction. In addition, the C_4 isoform from A. angusta differs from both C_4 and non-C_4 native forms from A. semialata in its increased K_M for PEP (fig. 3). This change has been observed in other C_4 lineages, but its physiological significance remains unknown (Ting and Osmond 1973; Bläsing et al. 2000; Gowik et al. 2006). One hypothesis is that it represents a side effect of selection for

another property, such as reduced inhibition by malate or different affinity for HCO_3^- (Svensson et al. 1997, 2003). Our study argues against this hypothesis as there is a lack of a correlation between these parameters and the K_M for PEP. Instead, it is likely that the increased K_M for PEP evolved in C_4 plants to allow a tighter regulation when substrate concentrations are high (Ting and Osmond 1973; Svensson et al. 2003). Although this hypothesis remains to be tested, our data show that the amino acid replacements observed in the native ppc-1P3 of A. angusta lead to a strengthening of the rapid changes observed in A. semialata, with further alterations of K_M for the substrates.

Lateral Gene Transfer Provides a Shortcut to Adaptation

The enzymes encoded by genes laterally acquired from three different grass lineages representing two C_4 origins (fig. 1) are highly similar in terms of their catalytic efficiency and affinity for PEP, which reflects convergence among the donor species (fig. 3). It is however clear from other studies that not all C_4 PEPC have the exact same properties (Ting and Osmond 1973; Moody et al. forthcoming), and we suggest that the clustering of properties reflects a bias in the genes that successfully transferred into Alloteropsis.

Compared with the co-opted native isoform from A. angusta, the catalytic efficiency of the laterally acquired versions is only slightly higher (fig. 3). However, their K_M values are massively increased (fig. 3). We conclude that the trend observed in A. angusta was continued in other lineages, leading to enzymes with very high K_M for PEP in older C_4 groups. The K_M for HCO_3^- is also strongly increased in the laterally acquired isoforms, which is opposite to differences observed in other C_4 systems (Bauwe 1986; DiMario and Cousins 2019; Moody et al. forthcoming). This might indicate that the optimal interaction with HCO_3^- is context dependent. Indeed, the enzyme catalyzing HCO_3^- production is essential in only some C_4 plants (Studer, Gandin, et al. 2014), suggesting that the substrate is naturally abundant in others. In all cases, the laterally acquired genes show amplified differences with the non-C_4 orthologs when compared with the co-opted native isoform of A. angusta (fig. 3). Because the co-opted native orthologs of A. semialata lack most C_4-specific amino acid modifications, the laterally acquired genes generated an extreme jump in the enzyme catalytic properties (fig. 3). The integration of these isoforms in the C_4 pathway of A. semialata therefore provided a direct shortcut, forgoing the long phase of adaptive evolution observed in A. angusta and other groups. We conclude that LGTs represent a highway to biochemical adaptation in plants.

The leaf anatomy and C_4 biochemistry are similar between the donors of ppc-1P3_LGT:A and ppc-1P3_LGT:C, and A. semialata (Prendergast et al. 1987; Renvoize 1987; Dunning et al. 2017), which might explain why the transfers were not followed by significant modification to the encoded enzyme. The C_4 phenotype is also similar between the donor of ppc-1P3_LGT:M and A. cimicina, which is the original recipient of the gene (Dunning et al. 2017). The ppc-1P3_LGT:M gene was subsequently introgressed from A. cimicina to

A. semialata (Dunning et al. 2017), which despite being closely related markedly differ in their C_4 anatomy (Dunning et al. 2017). Interestingly, this *A. semialata ppc-1P3_LGT:M* was replaced by *ppc-1P3_LGT:C* in several *A. semialata* accessions, and the former has been pseudogenized (supplementary fig. S1, Supplementary Material online; Olofsson et al. 2016). It is possible that the kinetic properties of the latter, including a larger K_M for HCO_3^- and a reduced sensitivity to aspartate (figs. 3 and 4), were advantageous in *A. semialata*, a species whose C_4 cycle relies on an aspartate shuttle (Dunning, Moreno-Villena, et al. 2019). We therefore suggest that the fit of the laterally acquired genes depends on the functional similarity between the donor and recipient species, making some evolutionary shortcuts more advantageous.

Conclusions

The evolution of complex traits, such as C_4 photosynthesis, involves the co-option of numerous genes, often requiring their subsequent modification to adapt the encoded enzymes for the new biochemical context. In the case of PEPC, the massive upregulation in expression of the non-C_4 copies was followed by amino acid replacements that rapidly increased the catalytic efficiency and sensitivity to inhibitors of the enzyme. This process, evidenced within *A. semialata*, likely capitalized on standing genetic variation. The resultant enzyme was able to sustain a functioning C_4 cycle, but was likely suboptimal and over time underwent secondary adaptations. This evolutionary process involved the fixation of novel mutations that are absent from non-C_4 forms and therefore likely necessitated substantial evolutionary time, explaining why the co-opted native isoform from *A. angusta* presents only some of the characteristics of older C_4 lineages. The interspecific transfer of genes already adapted to the C_4 context in these older groups provided a shortcut to evolutionary adaptation, bringing in enzymes that directly improved the novel physiology. Our work therefore shows that LGTs among grasses generated a leap toward the adaptation of emerging physiologies. We predict that such successful transfers will be more prevalent in the case of genes requiring extensive adaptations, as is the case of PEPC for the C_4 context.

Materials and Methods

Phylogenetic Analysis of the *ppc-1* Gene Family

We generated phylogenetic trees for different groups of the gene lineage *ppc-1* containing forms used for C_4 photosynthesis by some *Alloteropsis* (Dunning et al. 2017). Sequences were obtained from published transcriptomes and genomes (Moreno-Villena et al. 2018; Dunning, Olofsson, et al. 2019) or retrieved from NCBI database. In addition, we also included data for *A. semialata* (AUS1-01 accession; Dunning, Olofsson, et al. 2019), *A. angusta* (AANG4-8; unpublished), *A. cimicina* (data from Dunning, Olofsson, et al. 2019 and assembled using the same method), and *T. triandra* (Dunning, Olofsson, et al. 2019). Apart from the chromosome-level assembly of *A. semialata*, these genomes were generated solely using short-read data and as a result, the assemblies are highly

fragmented. We therefore had to assemble the *ppc-1* gene models from multiple contigs, and used *Setaria italica* and *Sorghum bicolor* sequences as a reference. We also generated gene models for two genes from a Zambian *A. semialata* accession (ZAM15-05-10) which were either truncated in AUS1-01 reference (*ppc-1P6*), or absent (*ppc-1P3_C*). Coding sequences were extracted from additional *Alloteropsis* short-read data sets as described in Dunning, Olofsson, et al. (2019). All gene models from each group of interest were then aligned using mafft v7.123b (Katoh and Standley 2013). For each group, a maximum likelihood phylogenetic tree was inferred using the third-codon positions to avoid biases due to convergent adaptive evolution. This was performed with PhyML v.21031022 (Guindon and Gascuel 2003) using the best substitution model identified using Smart Model Selection SMS v.1.8.1 (Lefort et al. 2017). Branch lengths were subsequently also estimated in amino acid substitution on the fixed topology using codeml v.4.7 (Yang 2007) with the M0 model.

Isolation and Cloning of *ppc-1* Genes

Genes representing a diversity of origins (fig. 1 and table 1) were selected for detailed biochemical characterization. This included native copies co-opted for C_4 photosynthesis, non-C_4 forms of the native copies as well as C_4 forms from species closely related to the putative donor for each laterally acquired gene (fig. 1 and table 1). To account for diversity within *Alloteropsis* two different variants were targeted for some genes (*ppc-1P3*, *ppc-1P3_LGT:M*, and *ppc-1P3_LGT:C*). Finally, a non-C_4 ortholog from a close relative of *Alloteropsis* (*P. pygmaeum*) was included using a previously prepared plasmid (Moody et al. forthcoming).

Complete coding sequences corresponding to the most abundantly transcribed copies, as identified based on transcriptome analyses (Dunning et al. 2017; Dunning, Moreno-Villena, et al. 2019), were isolated by PCR from leaf cDNAs. RNA was extracted from mature leaves that had been exposed to 7 h of light, using the RNeasy Plant Mini Kit (Qiagen). The synthesis of cDNA was then performed using the MultiScribe Reverse Transcriptase (Applied Biosystems) and RT random primers, following the manufacturer's instructions. Amplification was performed with the Q5 High-Fidelity DNA Polymerase (New England Biolabs), with primers corresponding to the 5′ and 3′ extremities of each targeted gene (table 1), as determined from previous transcriptomes (Dunning et al. 2017; Dunning, Moreno-Villena, et al. 2019). Each primer includes a digestion site before the start and after the stop codons (table 1), for follow-up cloning. The PCR mixture contained 1× Q5 Reaction Buffer, 200 μM dNTPs, 0.5 μM of each primer, ∼ 900 ng template cDNA, and 0.5 U Q5 DNA polymerase. A denaturing, annealing, and extension temperature of 98 °C (10 s), 57 °C (30 s), and 72 °C (3 min), respectively, were used in the PCR reactions over 35 cycles.

Successful PCR products were gel extracted using the QIAquick Gel Extraction Kit (Qiagen), and the purified products were digested with the appropriate restriction endonucleases (table 1). The digested products were ligated into pET-

28a(+) expression vectors (Novagen), using a T4 DNA ligase (New England Biolabs). The vectors had been previously digested with the appropriate enzymes, so that genes were cloned in-frame with the T7 promoter, lacO, ribosome-binding site, and N-terminal hexa-Histidine tag. The cloned constructs were Sanger sequenced using the T7 promoter and terminator primers and compared with the transcriptome data to verify the identity of the cloned genes. For several genes, PCR amplification failed, potentially because of low gene expression. In other cases, the unavailability of live plants prevented RNA isolation. These genes were therefore synthesized by GeneArt (LifeTechnologies) and directly cloned into the pET100/D-TOPO expression vector for codon-optimized expression in *Escherichia coli*.

Heterogeneous Expression and Purification of Recombinant PEPC

The 14 *ppc* constructs were used in the transformation of competent *E. coli* BL21λDE3 (Novagen) cells. Successfully transformed cells were selected for using either 50 mg ml^{-1} ampicillin (Sigma–Aldrich) or 30 mg ml^{-1} kanamycin (Sigma–Aldrich) depending on the plasmid vector. Bacterial cells were cultured in 2×TY media (1.6% [*w/v*] tryptone, 1% [*w/v*] yeast extract, 0.5% [*w/v*] NaCl, adjusted to pH 7.0 with NaOH and sterilized by autoclaving) at 25 °C with vigorous agitation and appropriate antibiotic added. At the mid-log phase (A$_{600}$ = ~0.6), the cultures were chilled at 4 °C for 1 h, then induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG; filter-sterilized; Melford) at 16 °C for a further 39 h. Cells were harvested by centrifugation at 4 °C (10 min; 14,000 × g), resuspended in lysis buffer (0.2 M Tris–HCl, 0.5 M NaCl, pH 8.0, with either pefabloc SC or Roche complete mini [EDTA free] protease inhibitors at the manufacturers recommended concentrations), and disrupted using a French pressure cell press (Constant Systems). The suspension was clarified by two sequential centrifugations at 4 °C (31,000 × g) for 15 min and 30 min, and the supernatants were passed through a 0.45-μm filter (Millipore) before it was fractionated on a 1-ml His-Trap HP column (GE Healthcare) at 1 ml min^{-1} on the ÄKTA pure (GE Healthcare), which was preequilibrated in the Binding Buffer (0.2 M Tris–HCl, 0.5 M NaCl, 50 mM imidazole [Sigma–Aldrich], pH 8.0). After washing with 60× column volumes of Wash Buffer (0.2 M Tris–HCl, 0.5 M NaCl, 100 mM imidazole, pH 8.0), recombinant PEPC was gradient-eluted with Elution Buffer (0.2 M Tris–HCl, 0.5 M NaCl, 400 mM imidazole, pH 8.0). Fractions containing eluted protein were then pooled and desalted using a 5-ml HiTrap Desalting Column (GE Healthcare) that had been preequilibrated with Storage Buffer (0.2 M Tris–HCl, 50 mM NaCl, 10% [*v/v*] glycerol, pH 8.0). Upon elution, the purified protein, as judged pure by resolving on a 10% Mini-Protean TGX precast gel (Bio-Rad) via SDS–PAGE and Coomassie Blue (Sigma–Aldrich) staining, was snap-frozen in aliquots and stored at −80 °C. The concentration of PEPC was determined using a NanoDrop UV-Vis spectrophotometer (ThermoFisher) whereby the A$_{280}$ measurements (subtracted by A$_{310}$) were divided by the predicted extinction coefficient of the amino acid sequence of a PEPC fused to the N-terminal hexa-Histidine tag (according to the ProtParam tool on the ExPASy server; web.expasy.org/protparam/).

Kinetic Analyses

Rates of PEPC catalyzed formation of oxaloacetate were measured spectroscopically by coupling to malate dehydrogenase where oxidation of the NADH cofactor can be monitored at 340 nm. Assays with a high, fixed, concentration of bicarbonate (HCO$_3^-$) were observed using a FLUOstar plate reader (BMG Labtech) through a 340 ± 5 nm bandpass filter in absorbance mode with a reaction volume of 150 μl. Assays where bicarbonate concentrations were varied were observed at 340 nm using a Cary spectrophotometer (Agilent Technologies) in a 1-ml volume. All reactions were at 25 °C and followed for at least 15 min. NADH concentrations in the plate reader were determined using a standard curve. All assays were performed using three or more independently purified PEPC with three technical replicates. Initial rates were corrected for blank rates, determined in the absence of PEPC.

Assays typically contained 50 mM Tris–HCl (pH 7.4), 5 mM MgCl$_2$, 6 Uml^{-1} malate dehydrogenase (porcine heart; Sigma), 0.2 mM NADH, 10 μM–5 mM PEP, 10 μM–10 mM KHCO$_3$, and were initiated by addition of PEPC (2–9 nM, final concentration). When the concentration of bicarbonate was varied KCl was added to maintain a constant ionic strength, background bicarbonate was removed by extensive sparging with N$_2$ and residual bicarbonate was determined by assay in the absence of added bicarbonate.

Inhibition parameters were determined for *Alloteropsis* genes at fixed bicarbonate (10 mM), variable PEP, and inhibitor (L-malate and L-aspartate) concentrations between 0 and 25 mM.

Kinetic Data Analysis

Kinetic parameters were determined by nonlinear regression analysis in Igor Pro (Version 8; Wavemetrics Inc.). In the absence of inhibitor, data were analyzed with equation (1), where $K_{iA}K_B$ was held at 50 μM^2 and with a correction factor for differences in activity between runs.

$$v_0 = (V_{max} \cdot [A] \cdot [B]) / ([A] \cdot [B] + K_A \cdot [B] + K_B \cdot [A] + K_{iA}K_B). \quad (1)$$

Estimates of the SE values for k_{cat} (i.e., $V_{max}/[E]_T$) and the two K_m values (i.e., K_A and K_B) were produced directly from the nonlinear regression analysis.

Inhibition parameters (K_I) were determined from secondary plots of $(k_{cat}/K_m)^{app}$ against inhibitor concentration fitted to equation (2).

$$(k_{cat}/K_m)^{app} = (k_{cat}/K_m)/(1 + [I]/K_I). \quad (2)$$

In Vivo Enzymatic Assays

Enzymes purified from leaves of the plants used to isolate the genes were characterized to determine whether posttranscriptional modification affects the kinetic patterns. The plants were maintained under greenhouse conditions with

supplementary lightings (Agrolux), temperature control ($25\,^\circ$C in the day and $20\,^\circ$C at night; Mitsubishi Electric), and a light pollution screen (CambridgeHOK) at The Arthur Willis Environment Centre, The University of Sheffield. They were maintained in 11-l, free-draining pots containing M3 compost (Levington) and perlite (Sinclair), mixed in a 2:1 volume ratio, under well-watered and suitably fertilized (Scotts Evergreen Lawn Food; The Scotts Company) conditions. They grew in ambient CO_2 and received 15 h daylight at the time of harvesting, with light intensities at the leaf levels measured using a light meter (LI-250A; LI-COR) at ≥ 500 and $\leq 12\ \mu$mol m^{-2} s^{-1} photosynthetic photon flux density for light and dark photoperiods, respectively. After a minimum of 30 days under the above conditions, 1.28 cm^2 mid-sections of leaf tissues were harvested after 7.5 h of exposure to daylight and after 7.5 h of dark, flash-frozen in liquid nitrogen, and disrupted by grinding to homogeneity when frozen using a mortar and pestle. To extract their protein contents, the ground tissues were resuspended in Extraction Buffer (200 mM bicine-KOH, pH 9.8, 5 mM dithiothreitol [DTT], with 1 tablet cOmplete protease inhibitor cocktail tablets [Roche] per 10 ml), snap-frozen in aliquots, stored in $-80\,^\circ$C, and used within 30 days. Proteins were colorimetrically quantitated ($\lambda = 562$ nm) via the BSA assay (Pierce) with BSA standards.

Enzyme assays were conducted as described above for the cloned genes, but only K_M(PEP) values were collected from the in vivo samples as absolute PEPC and HCO_3^- concentrations are difficult to estimate from leaf extracts. The in vivo measurements of non-C_4 accessions are difficult to compare with cloned genes, as non-C_4 individuals express multiple isoforms at low levels (Dunning et al. 2017). Focusing on the C_4 accessions, there is an overall good correlation between the in vivo and in vitro measurements of K_M(PEP), although more variation exists in leaf extracts (supplementary fig. S2 and table S2, Supplementary Material online). These results indicate that, despite important posttranscriptional regulations of PEPC (Jiao and Chollet 1991; Chollet et al. 1996; O'Leary et al. 2011), our comparisons of kinetic parameters are physiologically meaningful.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

Author Contributions

J.D.R. and P.A.C. designed the project, C.P. did the experimental work, C.P. and J.D.R. analyzed the experimental data, L.T.D. analyzed the gene sequences, and C.P. and P.A.C. wrote the article with the help of all coauthors.

References

Alvarez CE, Bovdilova A, Höppner A, Wolff CC, Saigo M, Trajtenberg F, Zhang T, Buschiazzo A, Nagel-Steger L, Drincovich MF, et al. 2019. Molecular adaptations of NADP-malic enzyme for its function in C_4 photosynthesis in grasses. *Nat Plants*. 5(7):755–765.

Arnold ML, Kunte K. 2017. Adaptive genetic exchange: a tangled history of admixture and evolutionary innovation. *Trends Ecol Evol*. 32(8):601–611.

Arrivault S, Obata T, Szecówka M, Mengin V, Guenther M, Hoehne M, Fernie AR, Stitt M. 2017. Metabolite pools and carbon flow during C_4 photosynthesis in maize: $^{13}CO_2$ labeling kinetics and cell type fractionation. *J EXp Bot*. 68(2):283–298.

Atkinson RR, Mockford EJ, Bennett C, Christin PA, Spriggs EL, Freckleton RP, Thompson K, Rees M, Osborne CP. 2016. C_4 photosynthesis boosts growth by altering physiology, allocation and size. *Nat Plants*. 2:1–5.

Aubry S, Brown NJ, Hibberd JM. 2011. The role of proteins in C_3 plants prior to their recruitment into the C_4 pathway. *J Exp Bot*. 62(9):3049–3059.

Bauwe H. 1986. An efficient method for the determination of K_m values for HCO_3^- of phosphoenolpyruvate carboxylase. *Planta* 169(3):356–360.

Bauwe H, Chollet R. 1986. Kinetic properties of phosphoenolpyruvate carboxylase from C_3, C_4, and C_3-C_4 intermediate species of *Flaveria* (Asteraceae). *Plant Physiol*. 82(3):695–699.

Besnard G, Muasya AM, Russier F, Roalson EH, Salamin N, Christin PA. 2009. Phylogenomics of C_4 photosynthesis in sedges (Cyperaceae): multiple appearances and genetic convergence. *Mol Biol Evol*. 26(8):1909–1919.

Bianconi ME, Dunning LT, Moreno-Villena JJ, Osborne CP, Christin PA. 2018. Gene duplication and dosage effects during the early emergence of C_4 photosynthesis in the grass genus *Alloteropsis*. *J Exp Bot*. 69(8):1967–1980.

Bläsing OE, Ernst K, Streubel M, Westhoff P, Svensson P. 2002. The nonphotosynthetic phosphoenolpyruvate carboxylases of the C_4 dicot *Flaveria trinervia*–implications for the evolution of C_4 photosynthesis. *Planta* 215(3):448–456.

Bläsing OE, Westhoff P, Svensson P. 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(36):27917–27923.

Blount ZD, Barrick JE, Davidson CJ, Lenski RE. 2012. Genomic analysis of a key innovation in an experimental Escherichia coli population. *Nature* 489(7417):513–518.

Chollet R, Vidal J, O'Leary MH. 1996. Phosphoenolpyruvate carboxylase: a ubiquitous, highly regulated enzyme in plants. *Annu Rev Plant Physiol Plant Mol Biol*. 47(1):273–298.

Christin PA, Edwards EJ, Besnard G, Boxall SF, Gregory R, Kellogg EA, Hartwell J, Osborne CP. 2012. Adaptive evolution of C_4 photosynthesis through recurrent lateral gene transfer. *Curr Biol*. 22(5):445–449.

Christin PA, Salamin N, Savolainen V, Duvall MR, Besnard G. 2007. C_4 photosynthesis evolved in grasses via parallel adaptive genetic changes. *Curr Biol*. 17(14):1241–1247.

Christin PA, Wallace MJ, Clayton H, Edwards EJ, Furbank RT, Hattersley PW, Sage RF, Macfarlane TD, Ludwig M. 2012. Multiple photosynthetic transitions, polyploidy, and lateral gene transfer in the grass subtribe Neurachninae. *J Exp Bot*. 63(17):6297–6308.

Deng H, Zhang LS, Zhang GQ, Zheng BQ, Liu ZJ, Wang Y. 2016. Evolutionary history of PEPC genes in green plants: implications for the evolution of CAM in orchids. *Mol Phylogenet Evol*. 94:559–564.

Desai MM, Fisher DS, Murray AW. 2007. The speed of evolution and maintenance of variation in asexual populations. *Curr Biol*. 17(5):385–394.

DiMario RJ, Cousins AB. 2019. A single serine to alanine substitution decreases bicarbonate affinity of phospho*enol*pyruvate carboxylase in C_4 *Flaveria trinervia*. *J Exp Bot*. 70(3):995–1004.

Dong LY, Masuda T, Kawamura T, Hata S, Izui K. 1998. Cloning, expression, and characterization of a root-form phosphoenolpyruvate carboxylase from *Zea mays*: comparison with the C_4-form enzyme. *Plant Cell Physiol*. 39(8):865–873.

Dunning LT, Lundgren MR, Moreno-Villena JJ, Namaganda M, Edwards EJ, Nosil P, Osborne CP, Christin PA. 2017. Introgression and repeated co-option facilitated the recurrent emergence of C_4 photosynthesis among close relatives. *Evolution* 71(6):1541–1555.

Dunning LT, Moreno-Villena JJ, Lundgren MR, Dionora J, Salazar P, Adams C, Nyirenda F, Olofsson JK, Mapaura A, Grundy IM, et al. 2019. Key changes in gene expression identified for different stages of C_4 evolution in *Alloteropsis semialata*. *J Exp Bot*. 70(12):3255–3268.

Dunning LT, Olofsson JK, Parisod C, Choudhury RR, Moreno-Villena JJ, Yang Y, Dionora J, Quick WP, Park M, Bennetzen JL, et al. 2019. Lateral transfers of large DNA fragments spread functional genes among grasses. *Proc Natl Acad Sci U S A*. 116(10):4416–4425.

Engelmann S, Bläsing OE, Gowik U, Svensson P, Westhoff P. 2003. Molecular evolution of C_4 phosphoenolpyruvate carboxylase in the genus *Flaveria* – a gradual increase from C_3 to C_4 characteristics. *Planta* 217(5):717–725.

Fernández R, Gabaldón T. 2020. Gene gain and loss across the metazoan tree of life. *Nat Ecol Evol*. 4(4):524–533.

Gowik U, Burscheidt J, Akyildiz M, Schlue U, Koczor M, Streubel M, Westhoff P. 2004. cis-Regulatory elements for mesophyll-specific gene expression in the C_4 plant Flaveria trinervia, the promoter of the C_4 phosphoenolpyruvate carboxylase gene. *Plant Cell* 16(5):1077–1090.

Gowik U, Engelmann S, Bläsing OE, Raghavendra AS, Westhoff P. 2006. Evolution of C_4 phosphoenolpyruvate carboxylase in the genus *Alternanthera*: gene families and the enzymatic characteristics of the C_4 isozyme and its orthologues in C_3 and C_3/C_4 *Alternantheras*. *Planta* 223(2):359–368.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol*. 52(5):696–704.

Hall JP, Brockhurst MA, Harrison E. 2017. Sampling the mobile gene pool: innovation via horizontal gene transfer in bacteria. *Philos Trans R Soc B*. 372(1735):20160424.

Hatch MD. 1987. C_4 photosynthesis: a unique blend of modified biochemistry, anatomy and ultrastructure. *Biochim Biophys Acta*. 895(2):81–106.

Heckmann D, Schulze S, Denton A, Gowik U, Westhoff P, Weber AP, Lercher MJ. 2013. Predicting C_4 photosynthesis evolution: modular, individually adaptive steps on a Mount Fuji fitness landscape. *Cell* 153(7):1579–1588.

Heyduk K, Moreno-Villena JJ, Gilman IS, Christin PA, Edwards EJ. 2019. The genetics of convergent evolution: insights from plant photosynthesis. *Nat Rev Genet*. 20(8):485–493.

Ibrahim DG, Burke T, Ripley BS, Osborne CP. 2009. A molecular phylogeny of the genus *Alloteropsis* (Panicoideae, Poaceae) suggests an evolutionary reversion from C_4 to C_3 photosynthesis. *Ann Bot*. 103(1):127–136.

Jain R, Rivera MC, Moore JE, Lake JA. 2003. Horizontal gene transfer accelerates genome innovation and evolution. *Mol Biol Evol*. 20(10):1598–1602.

Jiao JA, Chollet R. 1991. Posttranslational regulation of phosphoenolpyruvate carboxylase in C_4 and Crassulacean acid metabolism plants. *Plant Physiol*. 95(4):981–985.

Jiggins CD, Wallbank RW, Hanly JJ. 2017. Waiting in the wings: what can we learn about gene co-option from the diversification of butterfly wing patterns? *Philos Trans R Soc B*. 372(1713):20150485.

Karageorgi M, Groen SC, Sumbul F, Pelaez JN, Verster KI, Aguilar JM, Hastings AP, Bernstein SL, Matsunaga T, Astourian M, et al. 2019. Genome editing retraces the evolution of toxin resistance in the monarch butterfly. *Nature* 574(7778):409–412.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 30(4):772–780.

Kumar A, Natarajan C, Moriyama H, Witt CC, Weber RE, Fago A, Storz JF. 2017. Stability-mediated epistasis restricts accessible mutational pathways in the functional evolution of avian hemoglobin. *Mol Biol Evol*. 34(5):1240–1251.

Lefort V, Longueville JE, Gascuel O. 2017. SMS: smart model selection in PhyML. *Mol Biol Evol*. 34(9):2422–2424.

Li HT, Yi TS, Gao LM, Ma PF, Zhang T, Yang JB, Gitzendanner MA, Fritsch PW, Cai J, Luo Y, et al. 2019. Origin of angiosperms and the puzzle of the Jurassic gap. *Nat Plants*. 5(5):461–470.

Lundgren MR, Besnard G, Ripley BS, Lehmann CE, Chatelet DS, Kynast RG, Namaganda M, Vorontsova MS, Hall RC, Elia J, et al. 2015. Photosynthetic innovation broadens the niche within a single species. *Ecol Lett*. 18(10):1021–1029.

Lundgren MR, Dunning LT, Olofsson JK, Moreno-Villena JJ, Bouvier JW, Sage TL, Khoshravesh R, Sultmanis S, Stata M, Ripley BS, et al. 2019. C_4 anatomy can evolve via a single developmental change. *Ecol Lett*. 22(2):302–312.

Moody NR, Christin PA, Reid JD. Forthcoming. Kinetic modifications of C_4 PEPC are qualitatively convergent, but larger in Panicum than in *Flaveria*. *Front Plant Sci*.

Moreno-Villena JJ, Dunning LT, Osborne CP, Christin PA. 2018. Highly expressed genes are preferentially co-opted for C_4 photosynthesis. *Mol Biol Evol*. 35(1):94–106.

Neher RA, Shraiman BI, Fisher DS. 2010. Rate of adaptation in large sexual populations. *Genetics* 184(2):467–481.

Ochman H, Lawrence JG, Groisman EA. 2000. Lateral gene transfer and the nature of bacterial innovation. *Nature* 405(6784):299–304.

O'Leary B, Park J, Plaxton WC. 2011. The remarkable diversity of plant PEPC (phosphoenolpyruvate carboxylase): recent insights into the physiological functions and post-translational controls of non-photosynthetic PEPCs. *Biochem J*. 436(1):15–34.

Olofsson JK, Bianconi M, Besnard G, Dunning LT, Lundgren MR, Holota H, Vorontsova MS, Hidalgo O, Leitch IJ, Nosil P, et al. 2016. Genome biogeography reveals the intraspecific spread of adaptive mutations for a complex trait. *Mol Ecol*. 25(24):6107–6123.

Paulus JK, Niehus C, Groth G. 2013. Evolution of C_4 phosphoenolpyruvate carboxylase: enhanced feedback inhibitor tolerance is determined by a single residue. *Mol Plant*. 6(6):1996–1999.

Prendergast HDV, Hattersley PW, Stone NE. 1987. New structural/biochemical associations in leaf blades of C_4 grasses (Poaceae). *Funct Plant Biol*. 14(4):403–420.

Renvoize SA. 1987. A survey of leaf-blade anatomy in grasses XI. *Paniceae*. *Kew Bull*. 42(3):739–768.

Rosnow JJ, Evans MA, Kapralov MV, Cousins AB, Edwards GE, Roalson EH. 2015. Kranz and single-cell forms of C_4 plants in the subfamily Suaedoideae show kinetic C_4 convergence for PEPC and Rubisco with divergent amino acid substitutions. *J Exp Bot*. 66(22):7347–7358.

Sage RF, Christin PA, Edwards EJ. 2011. The C_4 plant lineages of planet Earth. *J Exp Bot*. 62(9):3155–3169.

Sage RF, Sage TL, Kocacinar F. 2012. Photorespiration and the evolution of C_4 photosynthesis. *Annu Rev Plant Biol*. 63(1):19–47.

Studer AJ, Gandin A, Kolbe AR, Wang L, Cousins AB, Brutnell TP. 2014. A limited role for carbonic anhydrase in C_4 photosynthesis as revealed by a ca1ca2 double mutant in maize. *Plant Physiol*. 165(2):608–617.

Studer RA, Christin PA, Williams MA, Orengo CA. 2014. Stability-activity tradeoffs constrain the adaptive evolution of RubisCO. *Proc Natl Acad Sci U S A*. 111(6):2223–2228.

Svensson P, Biasing OE, Westhoff P. 1997. Evolution of the enzymatic characteristics of C_4 phospho*enol*pyruvate carboxylase: a comparison of the orthologous *PPCA* phospho*enol*yruvate carboxylases of *Flaveria trinervia* (C_4) and *Flaveria pringlei* (C3). *Eur J Biochem*. 246(2):452–460.

Svensson P, Bläsing OE, Westhoff P. 2003. Evolution of C_4 phosphoenolpyruvate carboxylase. *Arch Biochem Biophys*. 414(2):180–188.

Tanz SK, Tetu SG, Vella NG, Ludwig M. 2009. Loss of the transit peptide and an increase in gene expression of an ancestral chloroplastic carbonic anhydrase were instrumental in the evolution of the

cytosolic C_4 carbonic anhydrase in *Flaveria*. *Plant Physiol*. 150(3):1515–1529.

Tausta SL, Coyle HM, Rothermel B, Stiefel V, Nelson T. 2002. Maize C_4 and non-C_4 NADP-dependent malic enzymes are encoded by distinct genes derived from a plastid-localized ancestor. *Plant Mol Biol*. 50(4/5):635–652.

Ting IP, Osmond CB. 1973. Photosynthetic phosphoenolpyruvate carboxylases: characteristics of alloenzymes from leaves of C_3 and C_4 plants. *Plant Physiol*. 51(3):439–447.

Toprak E, Veres A, Michel JB, Chait R, Hartl DL, Kishony R. 2012. Evolutionary paths to antibiotic resistance under dynamically sustained drug selection. *Nat Genet*. 44(1):101–105.

True JR, Carroll SB. 2002. Gene co-option in physiological and morphological evolution. *Annu Rev Cell Dev Biol*. 18(1):53–80.

Weinreich DM, Delaney NF, DePristo MA, Hartl DL. 2006. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* 312(5770):111–114.

Yang G, Anderson DW, Baier F, Dohmen E, Hong N, Carr PD, Kamerlin SCL, Jackson CJ, Bornberg-Bauer E, Tokuriki N. 2019. Higher-order epistasis shapes the fitness landscape of a xenobiotic-degrading enzyme. *Nat Chem Biol*. 15(11):1120–1128.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 24(8):1586–1591.