Evolution of C_4 Phospho*enol*pyruvate Carboxykinase in Grasses, from Genotype to Phenotype

Pascal-Antoine Christin, Blaise Petitpierre,* Nicolas Salamin,*† Lucie Büchi,* and Guillaume Besnard**[1]

*Department of Ecology and Evolution, Biophore, University of Lausanne, Lausanne, Switzerland; and †Swiss Institute of Bioinformatics, Lausanne, Switzerland

C_4 photosynthesis is an adaptation over the classical C_3 pathway that has evolved multiple times independently. These convergences are accompanied by strong variations among the independent C_4 lineages. The decarboxylating enzyme used to release CO_2 around Rubisco particularly differs between C_4 species, a criterion used to distinguish three distinct biochemical C_4 subtypes. The phospho*enol*pyruvate carboxykinase (PCK) serves as a primary decarboxylase in a minority of C_4 species. This enzyme is also present in C_3 plants, where it is responsible for nonphotosynthetic functions. The genetic changes responsible for the evolution of C_4-specific PCK are still unidentified. Using phylogenetic analyses on PCK sequences isolated from C_3 and C_4 grasses, this study aimed at resolving the evolutionary history of C_4-specific PCK enzymes. Four independent evolutions of C_4-PCK were shown to be driven by positive selection, and nine C_4-adaptive sites underwent parallel genetic changes in different C_4 lineages. C_4-adaptive residues were also observed in C_4 species from the nicotinamide adenine dinucleotide phosphate–malic enzyme (NADP–ME) subtype and particularly in all taxa where a PCK shuttle was previously suggested to complement the NADP–ME pathway. Acquisitions of C_4-specific PCKs were mapped on a species tree, which revealed that the PCK subtype probably appeared at the base of the Chloridoideae subfamily and was then recurrently lost and secondarily reacquired at least three times. Linking the genotype to subtype phenotype shed new lights on the evolutionary transitions between the different C_4 subtypes.

Introduction

C_4 photosynthesis is an adaptive trait that reduces photorespiration under low CO_2 and high temperature (von Caemmerer and Furbank 2003; Sage 2004). It is a strongly convergent phenotype, having evolved more than 50 times in at least 19 plant families (Muhaidat et al. 2007). The C_4 trait is a combination of numerous biochemical and anatomical modifications, some of which are common to all C_4 plants, such as the use of phospho*enol*pyruvate carboxylase (PEPC) for the initial fixation of CO_2 (Sinha and Kellogg 1996; Kellogg 1999). However, most of the other characters that together create the C_4-specific CO_2 pump vary among C_4 lineages. In particular, the Kranz anatomy, which is found in most C_4 plants (for exceptions, see Voznesenskaya et al. 2001; Edwards et al. 2004), displays great structural variations (e.g., Sinha and Kellogg 1996; Dengler and Nelson 1999; Soros and Dengler 2001; Kadereit et al. 2003; Muhaidat et al. 2007). The C_4 biochemical machinery is also variable, with well-documented differences in the decarboxylating enzymes used to liberate CO_2 from C_4 acids (Gutierrez et al. 1974; Sinha and Kellogg 1996; Kanai and Edwards 1999). Three different decarboxylating enzymes have been identified, and their use defines C_4 subtypes, which are generally associated with a suite of anatomical and biochemical characteristics (Gutierrez et al. 1974; Prendergast et al. 1987). The subtypes using the nicotinamide adenine dinucleotide–malic enzyme (NAD–ME) or nicotinamide adenine dinucleotide phosphate–malic enzyme (NADP–ME) are the most widespread and are found both among dicots and monocots (Gutierrez et al. 1974; Sage et al. 1999). The third C_4 subtype uses the adenosine triphosphate (ATP)-dependent phospho*enol*pyruvate carboxykinase (PCK) as decarboxylating enzyme (Edwards et al. 1971). C_4 species belonging to the PCK subtype have been found so far only in two grass (Poaceae) subfamilies (Panicoideae and Chloridoideae; Sage et al. 1999).

The PCK enzyme seems to be ubiquitous in plants, whatever their photosynthetic type (Leegood and Walker 2003). It plays a role in gluconeogenesis, liberating carbon stored in lipids and making this energy available for seedling tissues (Leegood and ap Rees 1978), but it could be involved in other functions, such as the metabolism of nitrogenous compounds in seeds (Delgado-Alvarado et al. 2007) and the catabolism of citrate and/or malate in ripening fruits (Famiani et al. 2005). The PCK isoform involved in C_4 photosynthesis (hereafter called C_4-PCK) is expressed in bundle-sheath cells where it releases CO_2 from oxaloacetate for the Calvin cycle. This enzyme is highly expressed in leaves of C_4 species from the PCK subtype (Edwards et al. 1971; Gutierrez et al. 1976; Finnegan et al. 1999; Voznesenskaya et al. 2006). A significant PCK leaf expression has also been detected in some species using the C_4-NADP–ME subtype, which led authors to suggest that it could complement the NADP–ME C_4 pathway of some taxa (Prendergast et al. 1987; Lin et al. 1993; Wingler et al. 1999; Voznesenskaya et al. 2006; Calsa and Figueira 2007; Muhaidat et al. 2007).

So far, PCK-encoding genes (*pck*) have been the subject of only a few studies, and *pck* sequences are available for less than 10 species, most of which are C_4 model species. The exact number of *pck* lineages, as well as their functions, remains largely unknown. This sparse knowledge disables comparative analyses of PCK characteristics among C_3 and C_4 taxa, and the evolutionary history of the PCK C_4-subtype remains largely unknown, at both the phenotype and genotype levels.

In the present study, we isolated *pck* genes of nonmodel C_3 and C_4 grass species. By elucidating the evolutionary relationships between C_4-specific and nonphotosynthetic *pck* genes of grasses, this study aimed to 1) test for the

[1] Present address: Imperial College London, Silwood Park Campus, Ascot, Berkshire, United Kingdom.

Key words: phospho*enol*pyruvate carboxykinase, Poaceae, molecular evolution, multiple origins, C_4 photosynthesis, convergence.

E-mail: pascal-antoine.christin@unil.ch.

occurrence of positive selection during the evolution of C_4-specific *pck*, 2) test for parallel adaptive genetic changes between the independent C_4-PCK lineages in grasses, and 3) map the *pck* C_4-adaptive changes on a species tree to infer the evolutionary history of the PCK phenotype in the grass family. The adaptive significance of the different biochemical subtypes is discussed in the light of independent C_4-PCK inductions detected from *pck* analyses.

Materials and Methods

Isolation of *pck* Genes

GenBank database was screened for genes encoding PCK in grasses. Two divergent sequences, NM_001056130 and NM_001070842, are, respectively, located on chromosomes 3 and 10 of the *Oryza sativa* genome. During this study, only one of these putative duplicates (i.e., NM_001056130) was sampled in other grasses, because its homologous genes encode the C_4 *pck* in *Urochloa panicoides* (Finnegan et al. 1999). Based on these sequences (see supplementary table 1, Supplementary Material online), forward (pck-369-for CAGCGACAGCTCCCTCAAGT) and reverse (pck-1888-rev CGATCTTGTAGCTGGCGAACAC) primers were designed in conserved regions. Fifty-seven grass species were sampled based on a published phylogenetic tree of Poaceae (Christin, Besnard, et al. 2008) to represent the different grass subfamilies and the different C_4 subtypes, as determined from the literature (Gutierrez et al. 1974; Prendergast et al. 1987; Schulze et al. 1996; Sage et al. 1999). PCK-encoding genes were isolated through polymerase chain reaction (PCR) from genomic DNA (gDNA) or complementary DNA (cDNA) isolated from green leaves. The PCR reaction mixture contained approximately 100 ng of DNA template, 5 μl of 10× AccuPrime PCR buffer II, 0.2 mM of each deoxynucleoside triphosphate, 0.2 μM of each primer, 3 mM of $MgSO_4$, 2.5 μl (5% vol) of dimethyl sulfoxide, and 1 unit of a proof-reading *Taq* polymerase (AccuPrime *Taq* DNA Polymerase High Fidelity, Invitrogen, Carlsbad, CA) in a total volume of 50 μl. The samples were incubated for 2 min at 94 °C, followed by 35 cycles consisting of 30 s at 94 °C, 30 s at 57 °C (annealing temperature), and 4 min at 68 °C. The last cycle was followed by a 20-min extension at 68 °C. PCR products were purified using a QIAquick Gel Extraction Kit (Qiagen, Valencia, CA). To separate the different genes (or alleles) putatively amplified, purified PCR products were cloned into the pTZ57R/T vector using the InsT/Aclone PCR Product Cloning Kit (Fermentas, Vilnius, Lithuania) and then PCR amplified with the M13 primers. Up to 20 positive clones were digested using *Taq*I restriction enzyme (Invitrogen). The degree of polymorphism for *Taq*I digestion products was high, allowing an unambiguous distinction of the different *pck* genes. For each species, inserts of each clone presenting a different restriction pattern were sequenced with internal primers using the Big Dye 3.1 Terminator cycle sequencing kit (Applied Biosystems, Foster City, CA), following the provider instructions, and separated on an ABI Prism 3100 genetic analyzer (Applied Biosystems). A segment of 2,400–2,800 bp, including about 1,500 coding bp (ca., 75% of the total coding sequence) and up to 10 introns, was sequenced. All sequences were deposited in the European Molecular Biology Laboratory (EMBL) database (accession numbers indicated in supplementary table 1, Supplementary Material online).

Isolation of PCK Transcript Segments from Green Leaves

The PCK isoforms involved in C_4 photosynthesis are highly expressed in leaves of species belonging to the PCK subtype (Edwards et al. 1971; Finnegan et al. 1999; Voznesenskaya et al. 2006). To identify the most transcribed PCK-encoding genes in green leaves, cDNAs from three C_4-PCK grasses (*Urochloa maxima*, *Sporobolus africanus*, and *Spartina maritima*) and one NADP–ME species known to express PCK at a relatively high level (*Digitaria sanguinalis*; Gutierrez et al. 1974) were screened. In addition, cDNA of a species previously described as NAD–ME for which two different *pck* genes were isolated (*Enteropogon prieurii*; see Results) was also analyzed. Fresh leaves from plants grown in greenhouses were sampled at noon, and total mRNAs were isolated using the NucleoSpin RNA Plant kit (Macherey-Nagel, Düren, Germany). Double-stranded cDNAs were then obtained with the ProtoScript First Strand cDNA Synthesis Kit (New England BioLabs, Beverly, MA). A segment of PCK gene transcripts was then PCR amplified using the following primers: GTACGCGGGAGAGATGAAGAAGG (pck-927-for) and CTGATGAACTGGTACATGGTCTG (pck-1441-rev). These primers were designed in regions highly conserved in all sequences previously isolated (see above) to amplify a fragment of 491 bp whatever the gene. However, some primer sites were not conserved in all PCK sequences. A nucleotide never present in grass *pck* was then chosen when designing primers. These precautions strongly reduce the probability of a PCR amplification bias. PCR was carried as described above except that $MgSO_4$ concentration was increased to 4 mM, annealing temperature was 54 °C, and extension time was reduced to 1 min at 68 °C. PCR products were purified using a QIAquick PCR Purification Kit (Qiagen), and both strands were directly sequenced with primers used for the PCR reaction.

DNA Sequence Analyses

Exons of *pck* genes isolated from gDNA were identified by homology with exons of *Zoysia japonica* (AB199899) and cDNAs, following the GT-AG rule. A Bayesian phylogenetic tree was constructed using MrBayes 3.1 (Ronquist and Huelsenbeck 2003) on the coding sequences of *pck* genes isolated from both gDNA and cDNA. The best-fit model, chosen by hierarchical likelihood ratio tests, was the general time reversible substitution model with a gamma shape parameter and a proportion of invariant sites (GTR + G + I). All model parameters were optimized independently for first, second, and third positions of codons. Two analyses, each of four chains, were run for 10,000,000 generations. A tree was sampled each 1,000 generations after a burn-in period of 3,000,000.

Coding sequences can be phylogenetically misleading due to adaptive evolution (Christin et al. 2007). A phylogenetic tree was thus also inferred from introns and third positions of codons to check for congruence with the coding-sequence topology. Considering only *pck* genes isolated from gDNA, sequences with both introns and exons were aligned using ClustalW (Thompson et al. 1994) with

gap opening and gap extension penalties set to 15 and 6.66, respectively, for both pairwise and multiple alignments. All exons were then removed from this data set. Intron alignment was visually checked but was not manually edited to avoid subjectivity. The phylogenetic tree was inferred simultaneously from introns and third positions with the same analysis parameters than for the coding sequences. The best-fit model for the introns was the GTR + G + I model. All parameters were optimized separately for introns and third positions.

Positive-Selection Tests

In order to test for adaptive changes linked to the evolution of C_4-specific PCKs, three codon models implemented in the software codeml from the PAML package version 4 (Yang 2007) were used. Model M1a allows omega (d_N/d_S ratio) to vary among codons of the coding sequences but not among branches of the phylogenetic tree. In this null model, codons evolve under either purifying selection (omega < 1) or relaxed selection (omega $= 1$). The alternative model A (Zhang et al. 2005) allows omega to vary both among codons and branches of the phylogenetic tree. This model, in addition to sites that evolve under purifying or neutral selection in the whole phylogenetic tree, allows some sites to evolve under purifying or neutral selection in the whole tree except in foreground branches where they evolve under positive selection (omega > 1). The third model A′ is identical to model A except that positive selection in foreground branches is replaced by relaxed selection (omega $= 1$). The foreground branches have to be defined a priori. In this study, basal branches of the four *pck* gene lineages shown to be involved in C_4 photosynthesis (see Results; fig. 1, in thick black) were used as foreground branches. These models were optimized using both the topology inferred from the coding sequences and the topology inferred from introns and third positions of codons. Models M1a and A, as well as models A′ and A, are nested, enabling comparisons through likelihood ratio test.

Results

DNA Sequence Analyses

Sixty-four PCK-encoding sequences were isolated from gDNA of 50 grass species. In addition, nine sequences were obtained from cDNA. Together with the 10 sequences taken from GenBank, the data set included 83 sequences. The phylogenetic tree inferred by Bayesian inference based on *pck*-coding sequences (fig. 1) was largely congruent with a phylogenetic tree inferred from plastid markers for the same species (Christin, Besnard, et al. 2008). One of the rice (*O. sativa*) *pck* genes (taken from GenBank) was placed as sister to all other grasses, suggesting a gene duplication preceding the grass diversification. All the subfamilies were monophyletic except the Arundinoideae (represented by *Arundo donax* and *Phragmites australis*), but these two species were positioned in a portion of the tree that lacked support (fig. 1). In the Chloridoideae, two different *pck* gene lineages were isolated, pointing to subfamily specific–gene duplication. One of the duplicates (hereafter called *pck-B*, fig. 1) was isolated from only five species spread into the subfamily, whereas the other duplicate was obtained from all sampled Chloridoideae species. Another gene duplication apparently occurred in the *Spartina* genus, producing the *pck-C* gene (fig. 1). Inside the Panicoideae subfamily, the position of *Tristachya* was not congruent with the plastid phylogenetic tree, its *pck* sequence being placed with sequences of the *Panicum* genus (fig. 1). In Panicoideae, the presence of two divergent gene lineages in three different C_4 clades supports three other independent gene duplications. A first one occurred at the base of the *Digitaria* genus (leading to *pck-D*, fig. 1) and one before the divergence of the monophyletic *Melinis–Brachiaria–Urochloa* clade (gene duplication leading to *pck-E*, fig. 1). Finally, genes from three distinct *Paspalum* species were sister to a second gene isolated from one of these species (fig. 1).

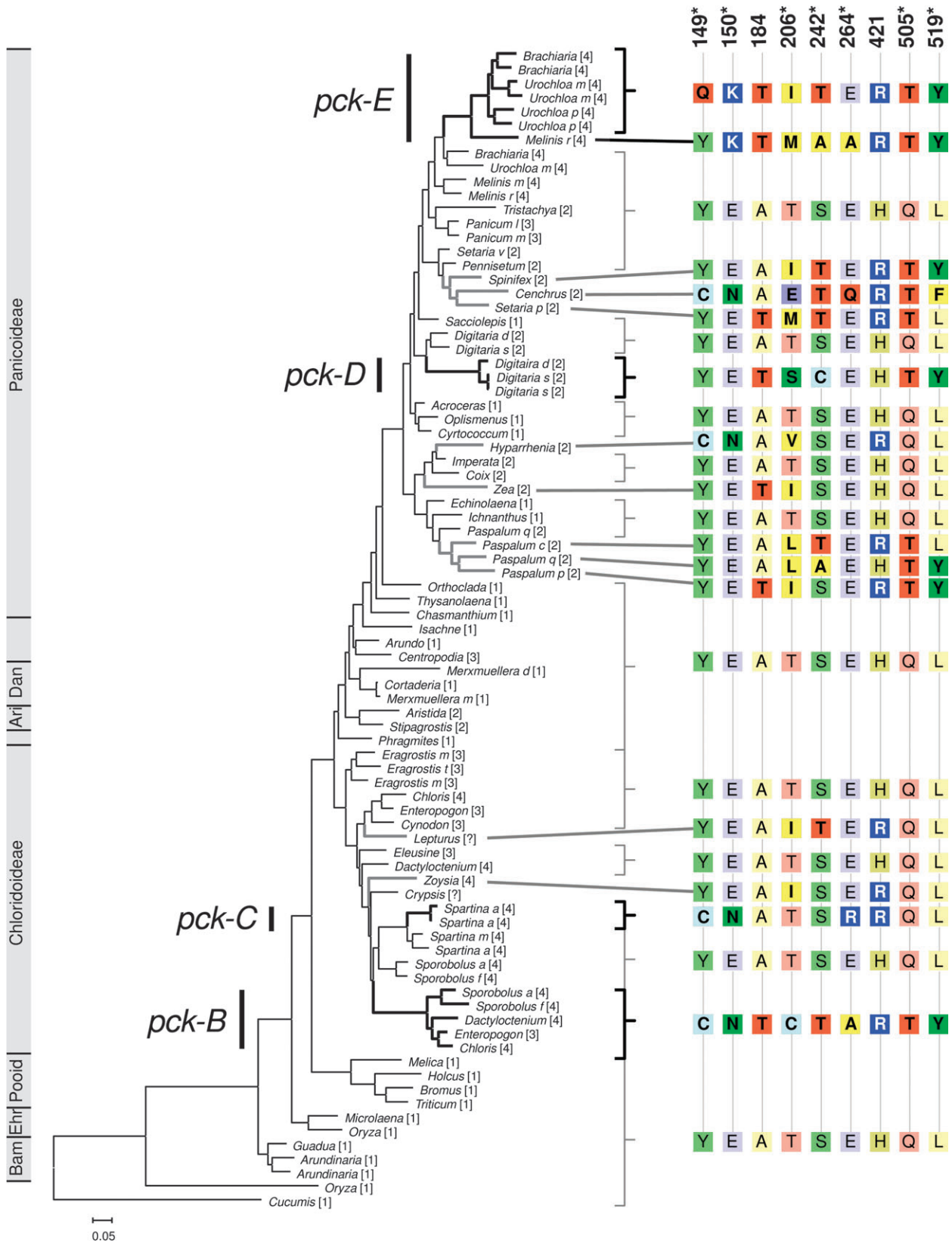
The topology inferred from introns and third positions (fig. 2) was similar to the one inferred from the coding sequences, except for two notable exceptions: Firstly, the sequence from *Tristachya* was not placed with *Panicum* species but basal to other Panicoideae, in congruence with the plastid markers (Christin, Besnard, et al. 2008). Secondly, the *pck-B* genes did not branch inside the other Chloridoideae sequences but as a sister group, which supports the occurrence of a gene-duplication event before or at the beginning of the diversification of this subfamily.

Isolation of *pck* Transcripts from Green Leaves

A phylogenetic tree was inferred from coding sequences (as previously described) to position the *pck* gene segment sequenced from cDNA (Supplementary fig. 2, Supplementary Material online). The PCK transcript isolated from *U. maxima* green leaves clustered with *pck-E* genes from this species, a cluster that includes *U. panicoides* genes, previously named PCK1 and PCK2 but now forming a monophyletic group, which are the ones involved in C_4 photosynthesis according to Finnegan et al. (1999). The three Chloridoideae *pck* transcripts from *Chloris*, *Enteropogon*, and *Sporobolus* unambiguously corresponded to the *pck-B* lineage (fig. 1) pointing to an involvement of this gene lineage in C_4 photosynthesis of these species. *S. maritima* transcript unambiguously clustered with one of the genes isolated from *Spartina anglica* gDNA (*pck-C*) and its corresponding transcript, which was also isolated from leaf (E12730; Imaizumi N, Osugi R, and Samejima M unpublished data). The transcript from the second *Spartina* gene was not detected in green leaves, and this gene cannot thus be considered as involved in C_4 photosynthesis. Finally, the transcript isolated from *D. sanguinalis* corresponded to one of the *Digitaria pck* clusters (called *pck-D*, fig. 1). These cDNA analyses thus revealed four distinct C_4-related *pck* lineages (fig. 1, in thick black), whose basal branches were set as the foreground branches in positive-selection analyses.

Positive-Selection Tests

The model allowing positive selection on branches leading to C_4 *pck* was significantly better than the null models, with both the topology inferred on coding sequences (M1a vs. A: chi-squared $= 199.6$, df $= 2$, P value < 0.0001;


149* 150* 184 206* 242* 264* 421 505* 519*
pck-E
pck-D
pck-C
pck-B
Panicoideae
Chloridoideae
Pooid
Ehr
Ari
Dan
Bam
Brachiaria [4]
Brachiaria [4]
Urochloa m [4]
Urochloa m [4]
Urochloa p [4]
Urochloa p [4]
Melinis r [4]
Brachiaria [4]
Urochloa m [4]
Melinis m [4]
Melinis r [4]
Tristachya [2]
Panicum l [3]
Panicum m [3]
Setaria v [2]
Pennisetum [2]
Spinifex [2]
Cenchrus [2]
Setaria p [2]
Sacciolepis [1]
Digitaria d [2]
Digitaria s [2]
Digitaira d [2]
Digitaria s [2]
Digitaria s [2]
Acroceras [1]
Oplismenus [1]
Cyrtococcum [1]
Hyparrhenia [2]
Imperata [2]
Coix [2]
Zea [2]
Echinolaena [1]
Ichnanthus [1]
Paspalum q [2]
Paspalum c [2]
Paspalum q [2]
Paspalum p [2]
Orthoclada [1]
Thysanolaena [1]
Chasmanthium [1]
Isachne [1]
Arundo [1]
Centropodia [3]
Merxmuellera d [1]
Cortaderia [1]
Merxmuellera m [1]
Aristida [2]
Stipagrostis [2]
Phragmites [1]
Eragrostis m [3]
Eragrostis t [3]
Eragrostis m [3]
Chloris [4]
Enteropogon [3]
Cynodon [3]
Lepturus [?]
Eleusine [3]
Dactyloctenium [4]
Zoysia [4]
Crypsis [?]
Spartina a [4]
Spartina a [4]
Spartina m [4]
Spartina a [4]
Sporobolus a [4]
Sporobolus f [4]
Sporobolus a [4]
Sporobolus f [4]
Dactyloctenium [4]
Enteropogon [3]
Chloris [4]
Melica [1]
Holcus [1]
Bromus [1]
Triticum [1]
Microlaena [1]
Oryza [1]
Guadua [1]
Arundinaria [1]
Arundinaria [1]
Oryza [1]
Cucumis [1]
0.05

A′ vs. A: chi-squared = 14.5, df = 1, P value < 0.0005) and the topology inferred from introns and third positions (M1a vs. A: chi-squared = 132.4, df = 2, P value < 0.0001; A′ vs. A: chi-squared = 6.7, df = 1, P value < 0.01). The omega for sites under positive selection was optimized to 2.37 and 1.84 for each topology, respectively. Seven codons had a posterior probability, as estimated by the Bayes empirical Bayes analysis (Yang et al. 2005), of having evolved under positive selection in foreground branches greater than 0.99 with both topologies and two additional sites had a probability greater than 0.95 with both topologies (fig. 1; supplementary table 2, Supplementary Material online). Some of the changes observed on these sites are shared between distant C_4 *pck* lineages (e.g., positions 184, 421, 505, and 519; fig. 1). On the other side, some of these sites mutated from a common ancestral amino acid to distinct residues in the independent C_4 *pck* groups (e.g., positions 149, 150, and 206; fig. 1). The presence of codons for putative C_4-adaptive amino acids on *Z. japonica* gene suggests a role of this gene in the C_4 pathway, as previously hypothesized (Nomura et al. 2005). In addition to being shared between C_4-PCK gene lineages, some of the codons identified under positive selection underwent parallel changes in other grass C_4 groups. For instance, residues at positions 149 and 150 (numbered following *Z. japonica* complete sequence; AB199899) are a tyrosine and a glutamic acid, respectively, in all C_3 species. They are mutated to a cysteine and an asparagine in C_4 *pck* of Chloridoideae (*pck-B*) and *Spartina* but also in *Cenchrus* and *Hyparrhenia* (fig. 1). Other C_4-PCK changes also occurred in C_4 species belonging to genera *Lepturus*, *Paspalum*, *Setaria*, *Spinifex*, and *Zea* (fig. 1; supplementary table 2, Supplementary Material online).

Discussion

Genetic Convergence of the C_4 Isoforms

Using nonspecific primers to amplify *pck* transcripts, the direct sequencing of *pck* transcripts allowed the identification of the most highly expressed decarboxylation gene in green leaves of three grass species of the PCK C_4 subtype and one with a significant C_4-related PCK activity (*D. sanguinalis*). In all cases, the dominant gene was the only one detectable in chromatograms. Therefore, these genes likely encode proteins involved in the C_4 pathway of these species. The putative C_4 gene in *S. africanus* is part of a clade formed mainly of species belonging to the PCK subtype (except *E. prieurii*, but see below), suggesting that genes of this cluster are all involved in C_4 photosynthesis. In this study, four distinct C_4-*pck* gene lineages (named *pck-B*, *C*, *D*, and *E*) were identified in grasses, and their evolution was driven by positive selection, as attested by the codon analyses. This suggests that adaptive amino acid changes modified the catalytic properties of the enzyme to optimize it for the C_4 pathway, as was observed for other C_4 enzymes (Svensson et al. 2003; Rondeau et al. 2005; Gowik et al. 2006; Christin et al. 2007; Estavillo et al. 2007). However, the PCK enzymes have not been really characterized at the kinetic level, and depicting the kinetic consequences of the identified amino acid changes would necessitate further biochemical investigations. Mutations of some amino acids could also account for the phosphorylation differences observed between C_3 plants and C_4 grasses (Walker and Leegood 1996; Walker et al. 1997), although this remains speculative.

At the positively selected codons, the same amino acids were recurrently mutated in the different C_4-PCK groups, in several cases to an identical residue (fig. 1). This convergent molecular evolution suggests that changes required to fulfill the C_4-specific function were achieved through recurrent alterations of the same sites, emphasizing the limited number of ways by which evolution can shape proteins (Weinreich et al. 2006). This adds to evidence of convergent genetic evolution acquired from other enzymes of the C_4 pathway (Bläsing et al. 2000; Christin et al. 2007; Christin, Salamin, et al. 2008) and from enzymes not related to C_4 photosynthesis (reviewed in Wood et al. 2005).

The recurrent mutations of the same codons in distant C_4-PCK lineages should lead to amino acids diagnostic of C_4 *pck* genes, similarly to the "serine 780" of PEPC (Bläsing et al. 2000; Svensson et al. 2003; Christin et al. 2007). For instance, the presence of C_4-PCK adaptive amino acids in *Lepturus*, a genus whose C_4-subtype was not determined (Sage et al. 1999), suggests that it belongs to the PCK subtype or, alternatively, that it could use this secondary pathway to complement the NAD–ME decarboxylating system. Several of the mutations driven by adaptive evolution in branches leading to C_4-PCK are also found in C_4 grasses belonging to other C_4 subtypes but are absent from C_3-species (fig. 1, supplementary table 2, Supplementary Material online). However, the classification of C_4 plants in biochemical subtypes is not unequivocal. In most cases, the decarboxylating type was only deduced from anatomical features (Sage et al. 1999), and it has been shown that anatomical clues can be misleading (Voznesenskaya et al. 2005; Muhaidat et al. 2007). Thus, the presence of C_4-PCK residues in *pck* of *E. prieurii*, a species previously described as NAD–ME type based on its anatomy (Schulze

FIG. 1.—Phylogenetic tree inferred from PCK-coding sequences. *Cucumis* PCK sequence was used to root the tree. Grass subfamilies are indicated on the left. Numbers in square brackets after species names indicate photosynthetic types and subtypes. [1]: C_3, [2]: C_4 NADP–ME, [3]: C_4 NAD–ME, and [4]: C_4-PCK. Black bars on the left of the tree show duplicated genes, which putatively underwent divergent evolution (see Discussion). Branches leading to C_4-PCK genes as deduced from transcripts analyses are indicated in thick black. The basal branches of each thick black group were used as foreground branches in the positive-selection tests. Amino acids at the positions evolving under positive selection (numbered following *Zoysia japonica* complete coding sequence; AB199899) in foreground branches are shown on the right. For each group of sequences, the most abundant residue is indicated (see supplementary table 2 for details, Supplementary Material online). Asterisks mean that residue posterior probability of being under positive selection was greater than 0.99, whereas the others codons had a posterior probability greater than 0.95. Amino acids indicative of C_4 *pck* are in bold, for readability purposes. Branches leading to C_4-PCK genes as deduced from the amino acid sequences are in thick gray. Support values can be found in supplementary fig. 1, Supplementary Material online. Abbreviations: Bam = Bambusoideae; Ehr = Ehrhartoideae; Pooid = Pooideae; Ari = Aristidoideae; and Dan = Danthonioideae.

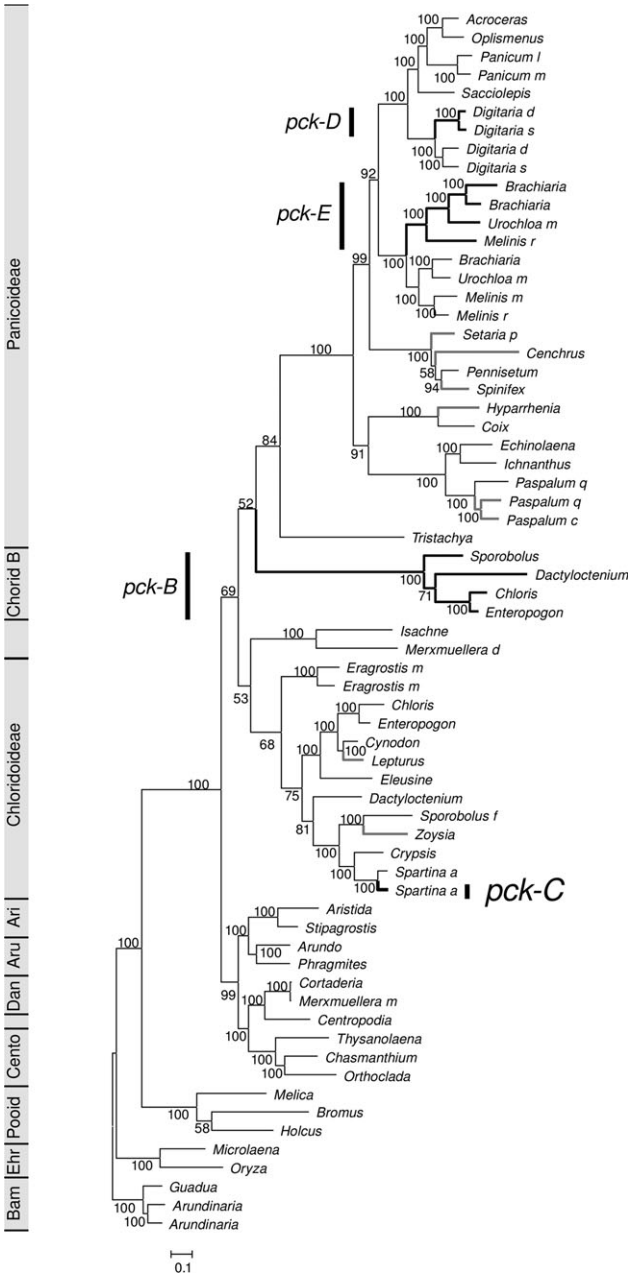


FIG. 2.—Phylogenetic tree inferred from PCK introns and third positions of codons. This tree is rooted on the Bambusoideae subfamily. Grass subfamilies are indicated on the left. Vertical black bars show duplicated genes and their names are given. Branches leading to C_4-PCK genes as deduced from cDNA analyses are indicated in thick black. The most basal branches of each thick black group were used as foreground branches in the positive-selection tests. Branches leading to C_4-PCK genes as deduced from the amino acids sequences are in thick gray. Bayesian support values are indicated near branches. Abbreviations: Bam = Bambusoideae; Ehr = Ehrhartoideae; Pooid = Pooideae; Cento = Centotecoideae; Dan = Danthonioideae; Aru = Arundinoideae; Ari = Aristidoideae; Chorid B = Chloridoideae (gene *pck-B*).

et al. 1996), could be due to a mismatch between morphological and biochemical characters. Moreover, the subtypes are attributed following the most expressed decarboxylating enzymes, but in many cases, multiple decarboxylating enzymes were significantly expressed in the same species (Gutierrez et al. 1974; Prendergast et al. 1987; Lin et al. 1993; Ueno and Sentoku 2006; Voznesenskaya et al. 2006; Muhaidat et al. 2007). The different C_4 decarboxylation pathways apparently co-occur in plants (Muhaidat et al. 2007), a NAD–ME shuttle completing the PCK pathway (Burnell and Hatch 1988; Kanai and Edwards 1999; Voznesenskaya et al. 2006) and a C_4-PCK carbon acquisition contributing to the NADP–ME cycle of some Panicoideae species (Wingler et al. 1999; Voznesenskaya et al. 2006). All species that share some of the putative C_4-PCK adaptive amino acids changes and that were biochemically analyzed exhibited significant leaf expression of PCK in addition to a predominant NADP–ME activity (*Cenchrus echinatus*, *Paspalum conjugatum*, and *Zea mays* [fig. 1, supplementary table 1, Supplementary Material online]). On the contrary, C_4 species in which leaf PCK activity was not detectable do not present any of the C_4-PCK adaptive changes (*Cynodon dactylon*, *Eleusine indica*, *Imperata cylindrica*, *Panicum miliaceum*, and *Stipagrostis pennata* [fig. 1, supplementary table 1, Supplementary Material online]), supporting the association between the amino acid changes highlighted in this study and a C_4-related PCK activity. Thus, *pck* genes with C_4-signature amino acids likely also contribute to the C_4-photosynthetic pathway and should be considered as C_4 *pck*. Consequently, at least six supplementary C_4-PCK gene lineages (fig. 1, in thick gray) add to the four groups detected through transcripts analyses (fig. 1, in thick black).

Number of C_4-PCK Appearances

This study thus identified 10 putative C_4-PCK lineages in the grass family. In five of them, the appearance of C_4-*pck* genes was linked to a duplication (twice in Chloridoideae and three times in Panicoideae), confirming the importance of this evolutionary process in the recruitment of C_4 enzymes from non-C_4 ones (Monson 2003). In Chloridoideae, the presence of *pck-B* in species spread within the subfamily (fig. 1) suggests that this gene appeared following a gene duplication event prior to the diversification of the subfamily. This is supported by the *pck-B* cluster being excluded from the other Chloridoideae genes in the topology inferred from introns and third positions (fig. 2). All five *pck-B* genes isolated in this study exhibit the same amino acids on almost all positively selected sites (supplementary table 2, Supplementary Material online), suggesting that the acquisition of the C_4 function occurred only once in their common ancestor. The C_4-PCK subtype thus seems to have evolved during the early diversification of the Chloridoideae subfamily, even if contradicting information between topologies inferred from coding sequences and from introns and third positions (figs. 1 and 2) does not allow a precise statement of when this event occurred. The C_4-PCK function persisted in some of the Chloridoideae lineages but was probably recurrently lost in many Chloridoideae species (fig. 3). The *pck-B* gene was not detectable in chromatograms obtained through direct sequencing of PCR products amplified from gDNA with pck-927-for and pck-1441-rev primers from the Chloridoideae genera *Cynodon*, *Eleusine*, *Spartina*, and *Zoysia* (data not shown),

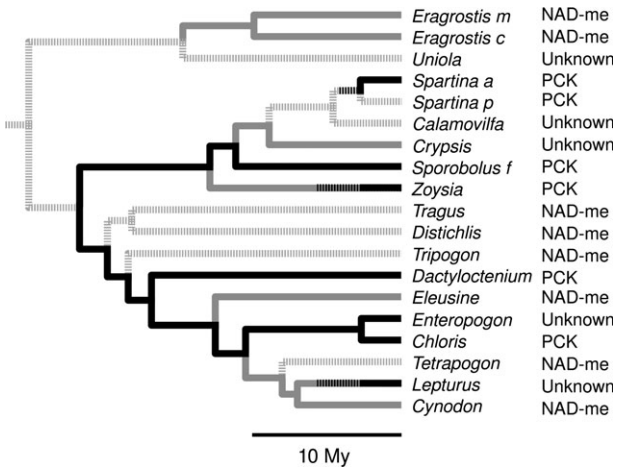


FIG. 3.—Evolutionary scenario of C4-PCK subtype in Chloridoideae as deduced from PCK genes. The subtree for the Chloridoideae subfamily was extracted from a monocot calibrated phylogenetic tree deduced from two plastid markers (Christin, Besnard, et al. 2008). Branch lengths are proportional to time, in millions of years. The putative use of C_4-PCK, as extrapolated from *pck* genes analyses, is indicated by black branches. Gray branches indicate the biochemical NAD–ME subtype without a C_4-PCK enzyme, whereas the state on dashed gray branches was not determinable with the species sample of the present study. The previously postulated C_4 subtypes are indicated on the right. On the branches leading to *Zoysia*, *Spartina*, and *Lepturus*, the changes of color represent secondary PCK subtype reacquisitions following losses, as attested by the C_4-PCK of these species having evolved from another duplicate than those of *Sporobolus*, *Dactyloctenium*, *Enteropogon*, and *Chloris* (figs. 1 and 2).

suggesting that it is not present in the genome of these species. The C_4 function of PCK was secondarily reacquired at least three times independently in *Spartina*, *Zoysia*, and *Lepturus* (figs. 1–3). This evolutionary scenario would not have been detectable based on a species tree as commonly inferred from presumably neutral genes. This highlights the importance of studying genes directly involved in the function when inferring the evolutionary history of a given phenotype (Christin, Besnard, et al. 2008; Goldberg and Igić 2008).

The number of independent acquisitions of C_4-specific PCK amino acids in the Panicoideae subfamily could be greater than the six presented in figure 1. Indeed, *Spinifex*, *Cenchrus*, and *Setaria* PCK genes all have putatively C_4-adaptive amino acids, but only two of the eight sites with putative C_4-adaptive changes are shared between these three species (fig. 1), suggesting independent optimizations of these C_4 characters. The same observation stands for sequences from the *Paspalum* genus. In addition, the Panicoideae sampling is not exhaustive and several Panicoideae with a demonstrated PCK activity (e.g., *Echinochloa* [Voznesenskaya et al. 2006] and *Alloteropsis* [Ueno and Sentoku 2006]) were not included. These species belong to Panicoideae C_4 lineages that were not included in the current study (Christin, Besnard, et al. 2008). It can thus be expected that their *pck* genes also evolved C_4 characteristics independently, which would increase the number of C_4-PCK lineages. These observations point to the acquisition of C_4-PCK-specific characteristics through a high number of parallel changes on *pck* genes in distant grass lineages.

Evolutionary Significance of the Biochemical C_4 Subtypes

This study revealed many independent acquisitions of C_4-specific PCK genes. Many of these events unequivocally occurred long after the initial evolution of C_4 photosynthesis, as deduced from both plastid and PEPC genes (Christin et al. 2007; Christin, Besnard, et al. 2008). In Chloridoideae, the evolution of C_4 photosynthesis was likely followed by several switches between PCK and NAD–ME subtypes (fig. 3), and in Panicoideae, the C_4-PCK function recurrently appeared in species using the NADP–ME subtype (fig. 1). The overlap between the different C_4 subtypes could have facilitated these recurrent switches. Nevertheless, the selective pressures responsible for these switches had to be significant to recurrently produce the high number of adaptive changes observed on *pck* genes. Despite difficulty in calculations of its energy requirement, the C_4-PCK pathway seems to require less ATP per CO_2 fixed than the two other C_4 subtypes (Kanai and Edwards 1999). This low-energy requirement could confer an advantage to the PCK subtype (Muhaidat et al. 2007). The adaptive values of the other C_4 subtypes are not yet fully understood. For instance, NAD–ME species have a better resistance to drought than NADP–ME species (Fravolini et al. 2002; Sato and Kubota 2004; Carmo-Silva et al. 2007; Ghannoum 2009). The status of PCK species is less obvious, PCK Chloridoideae (*Zoysia japonica*, *Bouteloua curtipendula*) being largely less drought resistant than NAD–ME species (*C. dactylon*, *Eragrostis lehmanniana*; Fravolini et al. 2002; Carmo-Silva et al. 2007), whereas in Panicoideae, PCK (*U. maxima*) and NAD–ME (*P. miliaceum*) taxa exhibit similar hydrological stress resistance (Sato and Kubota 2004). These discrepancies confirm the fact that the drought tolerance has a strong subfamilial component (Taub 2000; Cabido et al. 2008). However, the greater drought tolerance of NAD–ME Chloridoideae compared with PCK Chloridoideae (Fravolini et al. 2002; Carmo-Silva et al. 2007) suggests that the NAD–ME/PCK subtype switches that occurred in the Chloridoideae subfamily could have been driven by changes of environment xericity. This hypothesis would necessitate further investigations, but the numerous subtype switches supported by the *pck* gene analyses clearly point to different adaptive values of PCK and NAD–ME pathways, whatever they are.

Conclusions

The phylogenetic approach adopted here enabled us to identify codons that likely underwent adaptive changes to fulfill the C_4-specific function of some PCK gene lineages. Based on the phylogenetic relationships of C_4-PCK genes with their nonphotosynthetic relatives, it was possible to postulate an evolutionary scenario involving a single appearance of C_4-PCK in Chloridoideae followed by numerous losses and secondary acquisitions (fig. 3). It was also shown that C_4-specific PCKs recurrently appeared in different NADP–ME C_4 lineages of the Panicoideae subfamily, emphasizing the overlap between the different C_4 subtypes. These decarboxylating enzyme switches raise interesting

questions about the adaptive values of the different subtypes. Unfortunately, in physiological analyses generally only a few species in each subtype have been examined, and these have been considered representative of the subtype (Fravolini et al. 2002; Sato and Kubota 2004; Carmo-Silva et al. 2007). Because of the convergent nature of C_4 photosynthesis, it is possible to compare species of the same subtype but which evolved this subtype independently. This study of *pck* genes identified such species as well as others that changed from the PCK to the NAD–ME subtype. This knowledge of the PCK type evolutionary history, together with the high amount of phylogenetic data available for the grass family (Giussani et al. 2001; Hilu and Alice 2001; Christin, Besnard, et al. 2008; Vicentini et al. 2008), should be taken into account when designing experiments involving the different C_4 biochemical subtypes. A phylogeny oriented–species sampling could strongly decrease the risk of spurious correlations due to shared evolutionary history (Taub 2000; Edwards et al. 2007; Christin et al. 2009). Merging physiology, ecology, species phylogeny, and evolutionary genetics would lead to a complete and unbiased picture of constraints and selective pressures that led to the evolution of distinct C_4 subtypes.

Supplementary Material

Supplementary figure 1 (Phylogenetic tree inferred from PCK-coding sequences), supplementary figure 2 (Phylogenetic placement of direct sequencing of cDNA PCR products), supplementary table 1 (Accession numbers of *pck* genes obtained in this study), and supplementary table 2 (Codons under positive selection in branches leading to C_4 *pck*) are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

This work was funded by Swiss NSF grant 3100AO-105886. We thank Emanuela Samaritani and Zuzana Khodlová for their help in the laboratory. Nicole Galland, Rowan F. Sage, and two anonymous reviewers made useful comments on earlier versions of the manuscript.

Literature Cited

Bläsing OE, Westhoff P, Svensson P. 2000. Evolution of C_4 phospho*enol*pyruvate 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.

Burnell JN, Hatch MD. 1988. Photosynthesis in phopsho*enol*-pyruvate carboxykinase-type C_4 plants: photosynthetic activities of isolated bundle sheath cells form *Urochloa panicoides*. Arch Biochem Biophys. 260:177–186.

Cabido M, Pons E, Cantero JJ, Lewis JP, Anton A. 2008. Photosynthetic pathway variation among C_4 grasses along a precipitation gradient in Argentina. J Biogeogr. 35:131–140.

Calsa T, Figueira A. 2007. Serial analysis of gene expression in sugarcane (*Saccharum* spp.) leaves revealed alternative C_4 metabolism and putative antisense transcripts. Plant Mol Biol. 63:745–762.

Carmo-Silva AE, Soares AS, Marques da Silva J, Bernardes da Silva A, Keys AJ, Arrabaça MC. 2007. Photosynthetic responses of three C_4 grasses of different metabolic subtypes to water deficit. Funct Plant Biol. 34:204–213.

Christin PA, Besnard G, Samaritani E, Duvall MR, Hodkinson TR, Savolainen V, Salamin N. 2008. Oligocene CO_2 decline promoted C_4 photosynthesis in grasses. Curr Biol. 18:37–43.

Christin PA, Salamin N, Kellogg EA, Vicentini A, Besnard G. 2009. Integrating phylogeny into studies of C_4 variation in the grasses. Plant Physiol. 149: doi:10.1104/pp.108.128553.

Christin PA, Salamin N, Muasya AM, Roalson EH, Russier F, Besnard G. 2008. Evolutionary switch and genetic convergence on *rbcL* following the evolution of C_4 photosynthesis. Mol Biol Evol. 25:2361–2368.

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

Delgado-Alvarado A, Walker RP, Leegood RC. 2007. Phospho*enol*pyruvate carboxykinase in developing pea seeds is associated with tissues involved in solute transport and is nitrogen-responsive. Plant Cell Environ. 30:225–235.

Dengler NG, Nelson T. 1999. Leaf structure and development in C_4 plants. In: Sage RF, Monson RK, editors. C_4 plant biology. San Diego (CA): Academic Press. p. 133–172.

Edwards EJ, Still CJ, Donoghue MJ. 2007. The relevance of phylogeny to studies of global change. Trends Ecol Evol. 22:243–249.

Edwards GE, Franceschi VR, Voznesenskaya EV. 2004. Single-cell C_4 photosynthesis versus the dual-cell (Kranz) paradigm. Annu Rev Plant Biol. 55:173–196.

Edwards GE, Kanai R, Black CC Jr. 1971. Phospho*enol*pyruvate carboxykinase in leaves of certain plants which fix CO_2 by the C_4-decarboxylic acid cycle of photosynthesis. Biochem Biophys Res Commun. 45:278–285.

Estavillo GM, Rao SK, Reiskind JB, Bowes G. 2007. Characterization of the NADP malic enyzme gene family in the facultative, single-cell C_4 monocot *Hydrilla verticillata*. Photosynth Res. 94:43–57.

Famiani F, Cultrera NGM, Battistelli A, Casulli V, Proietti P, Standardi A, Chen ZH, Leegood RC, Walker RP. 2005. Phospho*enol*pyruvate carboxykinase and its potential role in the catabolism of organic acids in the flesh of soft fruit during ripening. J Exp Bot. 56:2959–2969.

Finnegan PM, Suzuki S, Ludwig M, Burnell JN. 1999. Phospho*enol*pyruvate carboxykinase in the C_4 monocot *Urochloa panicoides* is encoded by four differentially expressed genes. Plant Physiol. 120:1033–1041.

Fravolini A, Williams DG, Thompson TL. 2002. Carbon isotope discrimination and bundle sheath leakiness in three C_4 subtypes grown under variable nitrogen, water and atmospheric CO_2 supply. J Exp Bot. 53:2261–2269.

Ghannoum O. 2009. C_4 photosynthesis and water stress. Ann Bot. doi:10.1093/aob/mcn093.

Giussani LM, Cota-Sánchez JH, Zuloaga FO, Kellogg EA. 2001. A molecular phylogeny of the grass subfamily Panicoideae (Poaceae) shows multiple origins of C_4 photosynthesis. Am J Bot. 88:1993–2012.

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

Gowik U, Engelmann S, Bläsing OE, Raghavendra AS, Westhoff P. 2006. Evolution of C_4 phospho*enol*pyruvate 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. 223:359–368.

Gutierrez M, Edwards GE, Brown WV. 1976. PEP carboxykinase containing species in the *Brachiaria* group of the subfamily Panicoideae. Biochem Syst Ecol. 4:47–49.
Gutierrez M, Gracen VE, Edwards GE. 1974. Biochemical and cytological relationships in C_4 plants. Planta. 119:279–300.
Hilu KW, Alice LA. 2001. A phylogeny of Chloridoideae (Poaceae) based on *matK* sequences. Syst Bot. 26:386–405.
Kadereit G, Borsch T, Weising K, Freitag H. 2003. Phylogeny of Amaranthaceae and Chenopodiaceae and the evolution of C_4 photosynthesis. Int J Plant Sci. 164:959–986.
Kanai R, Edwards GE. 1999. The biochemistry of C_4 photosynthesis. In: Sage RF, Monson RK, editors. C_4 plant biology. San Diego (CA): Academic Press. p. 49–87.
Kellogg EA. 1999. Phylogenetics aspects of the evolution of C_4 photosynthesis. In: Sage RF, Monson RK, editors. C_4 plant biology. San Diego (CA): Academic Press. p. 411–444.
Leegood RC, ap Rees T. 1978. Phospho*enol*pyruvate carboxykinase and gluconeogenesis in cotyledons of *Cucurbita pepo*. Biochim Biophys Acta. 524:207–218.
Leegood RC, Walker RP. 2003. Regulation and roles of phospho*enol*pyruvate carboxykinase in plants. Arch Biochem Biophys. 414:204–210.
Lin CH, Tai YS, Liu DJ, Ku MSB. 1993. Photosynthetic mechanisms of weeds in Taiwan. Aust J Plant Physiol. 20:757–769.
Monson RK. 2003. Gene duplication, neofunctionalization, and the evolution of C_4 photosynthesis. Int J Plant Sci. 164:S43–S54.
Muhaidat R, Sage RF, Dengler NG. 2007. Diversity of Kranz anatomy and biochemistry in C_4 eudicots. Am J Bot. 94:362–381.
Nomura M, Higuchi T, Ishida Y, Ohta S, Komari T, Imaizumi N, Miyao-Tokutomi M, Matsuoka M, Tajima S. 2005. Differential expression patterns of C_4 bundle sheath expression genes in rice, a C_3 plant. Plant Cell Physiol. 46:754–761.
Prendergast HDV, Hattersley PW, Stone NE. 1987. New structural/biochemical associations if leaf blades of C_4 grasses (Poaceae). Aust J Plant Physiol. 14:403–420.
Rondeau P, Rouch C, Besnard G. 2005. NADP-malate dehydrogenase gene evolution in Andropogoneae (Poaceae): gene duplication followed by sub-functionalization. Ann Bot. 96:1307–1314.
Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 19:1572–1574.
Sage RF. 2004. The evolution of C_4 photosynthesis. New Phytol. 161:341–370.
Sage RF, Li M, Monson RK. 1999. The taxonomic distribution of C_4 photosynthesis. In: Sage RF, Monson RK, editors. C_4 plant biology. San Diego (CA): Academic Press. p. 551–584.
Sato A, Kubota F. 2004. Specific difference in photorespiration activity in C_4 subtype plants and its relationship with drought tolerance of leaf photosynthesis. J Fac Agr Kyushu Univ. 49:25–32.
Schulze ED, Ellis R, Shulze W, Trimborn P. 1996. Diversity, metabolic types and $\delta^{13}C$ carbon isotope ratios in the grass flora of Namibia in relation to growth form, precipitation and habitat conditions. Oecologia. 106:352–369.
Sinha NR, Kellogg EA. 1996. Parallelism and diversity in multiple origins of C_4 photosynthesis in the grass family. Am J Bot. 83:1458–1470.
Soros CL, Dengler NG. 2001. Ontogenic deviation and cell differentiation in photosynthetic tissues of C_3 and C_4 Cyperaceae. Am J Bot. 88:992–1005.
Svensson P, Bläsing OE, Westhoff P. 2003. Evolution of C_4 phospho*enol*pyruvate carboxylase. Arch Biochem Biophys. 414:180–188.
Taub DR. 2000. Climate and the US distribution of C_4 grass subfamilies and decarboxylation variants of C_4 photosynthesis. Am J Bot. 87:1211–1215.
Thompson JD, Higgins DJ, Gibson TJ. 1994. ClustalW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and matrix choice. Nucleic Acids Res. 22:4673–4680.
Ueno O, Sentoku N. 2006. Comparison of leaf structure and photosynthetic characteristics of C_3 and C_4 *Alloteropsis semialata* subspecies. Plant Cell Environ. 29:257–268.
Vicentini A, Barber JC, Aliscioni SS, Giussani LM, Kellogg EA. 2008. The age of the grasses and clusters of origins of C_4 photosynthesis. Glob Change Biol. doi:10.1111/j.1365–2486.2008.01688.x.
von Caemmerer S, Furbank RT. 2003. The C_4 pathway: an efficient CO_2 pump. Photosynth Res. 77:191–207.
Voznesenskaya EV, Chuong SDX, Kiirats O, Franceschi VR, Edwards GE. 2005. Evidence that C_4 species in genus *Stipagrostis*, family Poaceae, are NADP-malic enzyme subtype with nonclassical type of Kranz anatomy (Stipagrostoid). Plant Sci. 168:731–739.
Voznesenskaya EV, Franceschi VR, Chuong SDX, Edwards GE. 2006. Functional characterization of phospho*enol*pyruvate carboxykinase-type C_4 leaf anatomy: immuno-, cytochemical and ultrastructure analyses. Ann Bot. 98:77–91.
Voznesenskaya EV, Franceschi VR, Kiirats O, Freitag H, Edwards GE. 2001. Kranz anatomy is not essential for terrestrial C_4 plant photosynthesis. Nature. 414:543–546.
Walker RP, Acheson RM, Técsi LI, Leegood RC. 1997. Phospho*enol*pyruvate carboxykinase in C_4 plants: its role and regulation. Aust J Plant Physiol. 24:459–468.
Walker RP, Leegood RC. 1996. Phosphorylation of phospho*enol*pyruvate carboxykinase in plants. Studies in plants with C_4 photosynthesis and Crassulacean acid metabolism and in germinating seeds. Biochem J. 317:653–658.
Weinreich DM, Delaney NF, DePristo MA, Hartl DL. 2006. Darwinian evolution can follow only very few mutational paths to fitter proteins. Science. 312:111–114.
Wingler A, Walker RP, Chen ZH, Leegood RC. 1999. Phospho*enol*pyruvate carboxylase in involved in the decarboxylation of aspartate in the bundle sheath of maize. Plant Physiol. 120:539–545.
Wood TE, Burke JM, Rieseberg LH. 2005. Parallel genotypic adaptation: when evolution repeats itself. Genetica. 123:157–170.
Yang ZH. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24:1586–1591.
Yang ZH, Wong WSW, Nielsen R. 2005. Bayes empirical Bayes inference of amino acids sites under positive selection. Mol Biol Evol. 22:1107–1118.
Zhang JZ, Nielsen R, Yang ZH. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol Biol Evol. 22:2472–2479.