Phylogenomics of C₄ Photosynthesis in Sedges (Cyperaceae): Multiple Appearances and Genetic Convergence

Guillaume Besnard,*1 A. Muthama Muasya,† Flavien Russier,* Eric H. Roalson,‡ Nicolas Salamin,*§ and Pascal-Antoine Christin*

*Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland; †Botany Department, University of Cape Town, Private Bag X3, Rondebosch, South Africa; §School of Biological Sciences and Center for Integrated Biotechnology, Washington State University, Pullman, USA; and §§Swiss Institute of Bioinformatics, Genopode, Lausanne, Switzerland

C₄ photosynthesis is an adaptive trait conferring an advantage in warm and open habitats. It originated multiple times and is currently reported in 18 plant families. It has been recently shown that phosphoenolpyruvate carboxylase (PEPC), a key enzyme of the C₄ pathway, evolved through numerous independent but convergent genetic changes in grasses (Poaceae). To compare the genetics of multiple C₄ origins on a broader scale, we reconstructed the evolutionary history of the C₄ pathway in sedges (Cyperaceae), the second most species-rich C₄ family. A sedge phylogeny based on two plastome genes (rbcL and ndhF) has previously identified six fully C₄ clades. Here, a relaxed molecular clock was used to calibrate this tree and showed that the first C₄ acquisition occurred in this family between 19.6 and 10.1 Ma. According to analyses of PEPC-encoding genes (ppc), at least five distinct C₄ origins are present in sedges. Two C₄ Eleocharis species, which were unrelated in the plastid phylogeny, acquired their C₄-specific PEPC genes from a single source, probably through reticulate evolution or a horizontal transfer event. Acquisitions of C₄ PEPC in sedges have been driven by positive selection on at least 16 codons (3.5% of the studied gene segment). These sites underwent parallel genetic changes across the five sedge C₄ origins. Five of these sites underwent identical changes also in grass and eudicot C₄ lineages, indicating that genetic convergence is most important within families but that identical genetic changes occurred even among distantly related taxa. These lines of evidence give new insights into the constraints that govern molecular evolution.

Introduction

C₄ photosynthesis is a complex adaptation over the classical C₃ pathway (von Caemmerer and Furbank 2003; Sage 2004). It consists of a set of morphological and biochemical modifications that together create a CO₂ pump, which concentrates CO₂ around Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco). This suppresses photorespiration, a phenomenon resulting from the oxygenase activity of Rubisco and leading to CO₂ release and energy loss (Foyer et al. 2009). The C₄ pathway thereby confers an increased photosynthetic efficiency in all conditions promoting high rates of photorespiration, and especially in open and warm habitats (Ehleringer et al. 1997; Sage 2004). The leaf anatomy generally associated with the C₄ syndrome, called Kranz anatomy, is characterized by two rings of cells surrounding the vascular bundles, the mesophyll (M) and bundle sheath (BS) cells. Atmospheric CO₂ reaches M cells via the plant stomata and is converted into HCO₃⁻. This molecule is then fixed on phosphoenolpyruvate by phosphoenolpyruvate carboxylase (PEPC) into oxaloacetate. This four-carbon acid is then transported to the BS cells via a biochemical path that varies among C₄ plants (Kanai and Edwards 1999). Therein, decarboxylating enzymes release CO₂ from the four-carbon compound, which can then enter the Calvin cycle as in C₃ plants through its fixation by Rubisco. Three biochemical C₄ subtypes are recognized based on the decarboxylating enzymes: nicotinamide adenine dinucleotide phosphate–malic enzyme (NADP–ME), nicotinamide ade-
analyzed in a facultative C4 Hydrocharitaceae, ing more C4 systems by enabling multi-scale comparisons (e.g., intra and interfamilies).

Besides grasses, eudicot C4 systems have been the subject of numerous and fruitful investigations on C4 molecular evolution. The genus Flaveria (Asteraceae) in particular has received considerable attention. It contains several C3 and C4 relatives, as well as photosynthetic intermediates representing a gradual increase from C3 to C4 photosynthesis (McKown et al. 2005). The ppc genes and their encoded enzymes have been particularly well characterized in this genus (Bläsing et al. 2000; Svensson et al. 2003; Westhoff and Gowik 2004; Akyildiz et al. 2007; Jacobs et al. 2008). A study was also performed on ppc genes of the genus Alternanthera (Amaranthaceae), which contains C3, C3–C4, and C4 species (Gowik et al. 2006). Genetic information for these C4 eudicots has already allowed a first and rough interfamilial comparison of C4 genetic mutations (Christin et al. 2007). Recently, ppc genes were also analyzed in a facultative C4 Hydrocharitaceae, Hydrilla verticillata (Rao et al. 2008). Unfortunately, the limited number of C4 species and independent lineages in these groups hampered efficient statistical testing of adaptation and other C4-rich systems have to be investigated.

The sedge family (Cyperaceae) is the second most important C4 family, with approximately 1,500 C4 species (more than 20% of all C4 plants; Sage 2004). However, C4 sedges have been subject to limited investigations compared with other C4 systems. The different C4 species of Cyperaceae present anatomical and biochemical variations suggesting that their C4 traits are not always homologous (Bruhl et al. 1987; Ueno and Koyama 1987; Ueno et al. 1989; Bruhl and Perry 1995; Soros and Dengler 2001). Four anatomical types have been recognized, which are generally accompanied by a given C4 biochemical subtype (Soros and Dengler 2001). The rhynchosporoid (Rhynchospora), chlorocyperoid (Cyperaceae), and fimbristyloid (Abildgaardiaceae) anatomical types are associated with the NADP–ME subtype, whereas species presenting the eleocharoid (Eleocharis) anatomical type (and one anomalous fimbristyloid species, Eleocharis vivipara; Bruhl et al. 1987; Ueno et al. 1988) use the NAD–ME biochemical type. Phylogenetic investigations suggest at least four independent C4 origins in Cyperaceae (Soros and Bruhl 2000; Bruhl and Wilson 2007), but further considerations are needed, especially for clades containing C3 and C4 species and whose phylogeny is not completely resolved (such as genera Eleocharis and Rhynchospora, or Tribe Abildgaardiaceae; Roalson and Friar 2000; Ghamkhari et al. 2007; Thomas et al. 2009). Investigations of C4 genetics in sedges are even more sparse and concerned exclusively E. vivipara (Agarie et al. 2002), one of a small group of species capable of employing either C3 or C4 photosynthesis depending on environmental conditions (Ueno et al. 1988; Ueno 2001, 2004; Edwards et al. 2004; Murphy et al. 2007).

The present study used species relationships deduced from plastid markers (Christin, Salamin, et al. 2008) to estimate the dates of the C4 origins in this family. Analyses of nuclear ppc genes enabled testing for independent acquisitions of the C4 pathway in the different C4 lineages deduced from plastid markers. This phylogenetic framework was used here to shed light on the genetics of convergence, and specifically to 1) test for a role of adaptive genetic evolution in the acquisition of C4-specific PEPCs, 2) detect the C4-adaptive mutations and compare them among the different C4 lineages of sedges, and 3) compare these changes with those that occurred in C4 grasses and C4 eudicots. These multi-scale comparisons revealed strong infrafamilial genetic convergence and even angiosperm-wide convergence of some genetic adaptations.

Materials and Methods

Plant Material and Molecular Dating of the Plastid DNA Phylogenetic Tree

The 104 species of Cyperaceae (including 31 C4 species) analyzed with plastid DNA markers by Christin, Salamin, et al. (2008; supplementary table 1) were chosen to represent main tribes of Cyperaceae and all known C4 lineages and their C3 relatives (Bruhl and Wilson 2007; Muasya et al. 2009). A subsample of 63 species (including 18 C4 species) was characterized in the present study for ppc gene variation (see below). This subsample represents all of the main tribes of Cyperaceae, and all C4 lineages detected with plastid markers as well their C3 relatives (supplementary table 1, Supplementary Material online).

The phylogenetic tree of Poales previously inferred from two plastid markers, rbcL and ndhF, using Bayesian methods (Christin, Salamin, et al. 2008) contains 334 commelinid species and four Asparagales that were used as outgroup. In this tree, C4 sedges cluster in six fully C4 lineages. This topology was used for Bayesian molecular dating, using the multidivtime software (Kishino et al. 2001; Thorne and Kishino 2002). The analysis was run as described by Christin, Besnard, et al. (2008), except that the upper bound set at 60 My on the stem node (divergence of the clade from its sister group) on the BEP–PACMAD clade of grasses was removed as suggested by Vicentini et al. (2008). This constraint had no significant effect on the age estimates (supplementary table 2, Supplementary Material online).

Isolation of PEPC-Encoding Genes

Only one ppc gene sequence of Cyperaceae (a cDNA of C4 ppc from E. vivipara; EMBL no AB085948) was available in public databases. Polymerase chain reaction (PCR) isolation of the whole ppc genes was not easily feasible on genomic DNA due to its relatively large number of base pairs (bp). We first isolated sequences of ppc genes transcribed in leaves of a C3 species, Cyperus eragrostis. Fresh leaves from a plant grown in a greenhouse were sampled at noon and total mRNAs were isolated using the NucleoSpin RNA Plant kit (Macherey-Nagel). Double-stranded complementary DNAs were then obtained with the ProtoScript First Strand cDNA Synthesis Kit.
(BioLabs). Two primers able to PCR amplify a long ppc cDNA segment in monocots (~2,000 bp) were developed: ppc-850F (5′CAG TTC TCT TCY TGG ATG GG3′) and ppc-2872R (5′GCR GCR ATR CCC TTC ATG GT3′). The PCR reaction mixture contained ~100 ng of cDNA template, 5 µl of 10× AccuPrime PCR Buffer, 0.2 µM of each primer, 3 mM of MgSO4, and 1 unit of proof-reading Taq polymerase (AccuPrime Taq DNA Polymerase High Fidelity, Invitrogen) in a total volume of 50 µl. This mixture was incubated at 94 °C for 2 min, followed by 36 cycles consisting of 45 s at 94 °C, 45 s at 53 °C, and 2 min 30 s at 68 °C. The last cycle was followed by 20 min at 68 °C. PCR products were purified with the QIAquick Gel Extraction Kit (Qiagen). The purified fragments were cloned into the pTZ57R/T vector with the InstAclone PCR Product Cloning Kit (Fermentas). Positive clones were identified according to the preliminary assessment of sedge ppc diversity, all genes encoding C4 PEPc were identified by the presence of a serine at position 780 (according to maize C4 PEPc; CAA33317), which characterizes all C4 PEPcs and is necessary for the C4 function (Bläsing et al. 2000; Christin et al. 2007).

According to the preliminary assessment of sedge ppc diversity, all genes encoding C4 PEPc in this family belong to the same gene lineage (ppc-I; see Results). We thus focused on this gene lineage for the analysis of the subsample of 63 Cyperaceae species (supplementary table 1, Supplementary Material online). A gene segment covering exons 8–10 and carrying major C4 amino acids (Bläsing et al. 2000; Christin et al. 2007) was characterized. Two introns are generally present in this ppc gene segment (Christin et al. 2007; supplementary fig. 1, Supplementary Material online). Based on ppc sequences isolated from Cy. eragrostis (EMBL no FM208065 and FM208066) and E. vivipara (AB085948), one first primer pair was designed: ppc-1336F (5′TTT GGT CTC TCT YTT GTG CGT C3′) and ppc-2727R (5′GGT SGG GTC CCT GAT YCT TTT G3′). These primers allow amplification of a ppc segment with 1,369 bp of coding sequence (supplementary fig. 1, Supplementary Material online). In a 50-µl volume, the PCR mixture contained ~100 ng of genomic DNA, 5 µl of 10× AccuPrime PCR Buffer II, 0.2 µM of each primer, 3 mmol of MgSO4, 2.5 µl of dimethyl sulfoxide and 1 unit of Taq DNA polymerase (AccuPrime, Invitrogen). This mixture was incubated at 94 °C for 2 min, followed by 36 cycles consisting of 45 s at 94 °C, 45 s at 50 °C, and 2 min at 68 °C. The last cycle was followed by 20 min at 68 °C. PCR product purifications were done as described in the previous section. Twenty-nine species were successfully characterized with these primers (supplementary table 1, Supplementary Material online). In other cases, PCR failed (probably because DNAs were not of sufficiently high quality for amplification of DNA segments up to 2,000 bp) and thus a second primer pair was designed to amplify a shorter ppc segment (with 1,201 bp of coding sequence; supplementary fig. 1, Supplementary Material online): ppc-1426F (5′GGG TCM TAC CGT GAG TGG TC3′) and ppc-2647R (5′CTT TGY TTT ATG TAG GGA TCT CC3′). The PCR conditions used were exactly the same as those previously described. For sequencing, five different primers were used to directly sequence the ppc segments: ppc-2032F (5′GAG CAG TCR TTT GGT GAG GAG C3′), ppc-2209R (5′GGR GGA AAT ACT CAA CAA AGC G3′), ppc-2246F (5′TGG AGT ATG GCC GYA TGA ACA T3′), ppc-2411R (5′CAT STG RAT GGT GCC CAC ATC CT3′), and ppc-2600F (5′CMA AGA AKC TYC TTC TTC AGG AGG3′). Sequences obtained with these different primers overlap with one another. The sequencing protocol previously described was used (Christin et al. 2007). Every DNA sequencing chromatogram was checked and edited. In some species, double-peak or unreadable chromatograms were obtained due to the presence of multiple alleles or gene paralogs. To separate the different sequences, the purified fragments were cloned following the protocol described for cDNA. The distinct clones (identified by their cDNA restriction profile) were then sequenced.

Analyses of PEPC-Encoding Genes

Sequences obtained from genomic DNA were aligned with cDNAs, and introns were identified following the GT–AG rule. Phylogenetic trees were then reconstructed from coding sequences using Bayesian inference as implemented in MrBayes 3.1 (Ronquist and Huelsenbeck 2003). The best-fit substitution model, determined through hierarchical likelihood tests, was the general time reversible (GTR) 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 parallel 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 generations. A first phylogeny was inferred from coding sequences of sedge genes isolated with monocot ppc primers and PEPC-encoding genes from other plant families (supplementary table 4, Supplementary Material online). A second phylogenetic tree was then inferred only from coding sequences of gene lineage ppc-I (supplementary table 1, Supplementary Material online).

To confirm the phylogenetic pattern observed for ppc-I genes of the Eleocharis clade (see Results), their two introns were manually aligned and used to infer a phylogenetic tree, using the analysis parameters described above. The best-fit model for these introns was a GTR + G.

Positive Selection Tests

Three codon models were optimized on coding sequences of ppc-I from sedges, using the topology previously obtained. The software codeml, implemented in PAML 4
FIG. 1.—Calibrated Cyperaceae phylogenetic tree. The part of the phylogenetic tree corresponding to sedges and Juncaceae is presented. Branch lengths are proportional to estimated times, in million years before present (Ma). Branches leading to C₄ clades are in bold. C₄ lineages are numbered according to Christin, Salamin, et al. (2008). Estimates of C₄ evolutionary tempo can be found in table 1. Taxon names were abbreviated to the genus plus the first letter(s) of the species (when several species of the same genus were analyzed).
(Yang 2007), was used. These models evaluate the selective pressures driving the evolution of genes via the nonsynonymous versus synonymous mutations rate ratio ($d_{N}/d_{S}$; omega). An omega smaller than 1 indicates purifying selection, whereas an omega equal to 1 suggests relaxed selection. An omega value significantly greater than 1 demonstrates a proportion of fixed nonsynonymous mutations that is greater than expected by chance, which is usually interpreted as evidence for positive selection. The first model M1a is a site model allowing omega to vary among codons, being either smaller than 1 or equal to 1 (Yang et al. 2000). The alternative model A is a branch-site model that allows the omega ratio to vary both among codons and among branches of the phylogenetic tree (Yang and Nielsen 2002). In this model, sites are attributed to four different classes; either purifying selection in the whole tree, relaxed selection in the whole tree, purifying selection but positive selection in foreground branches, or relaxed selection but positive selection in foreground branches. The last model A’ is identical to model A except that positive selection is replaced by relaxed selection (Zhang et al. 2005). Both models M1a and A’ are nested in model A, enabling comparisons through likelihood ratio tests. The first test compares models M1a and A, whereas the second test, which compares models A’ and A, is more conservative (Zhang et al. 2005). In models A and A’, the foreground branches on which positive selection is expected have to be defined a priori. In this study, they were set simultaneously to all branches leading to a group in which all members encode C$_4$ PEPC with a serine at position 780.

The Bayes empirical Bayes procedure (Yang et al. 2005) implemented in codeml calculates the posterior probability of each codon to have evolved under positive selection in foreground branches. Codons for which this probability was greater than 0.999 were considered as having evolved under positive selection.

**Results**

**Molecular Dating**

The calibrated Cyperaceae phylogenetic tree is presented in figure 1. For taxa already present in our previous calibrated tree (Christin, Besnard, et al. 2008), divergence time values were fully congruent with the prior estimates (supplementary table 2, Supplementary Material online). Consequently, only the part of the tree corresponding to Cyperaceae is presented here. The divergence of sedges from Juncaceae (stem node) was estimated at 87.3 ($\pm$ 7.9) Ma. The first divergence of two sedges (crown node) was at 75.1 ($\pm$ 7.7) Ma. The age estimates for the sedge C$_4$ lineages are presented in table 1. Following our analyses, the first C$_4$ appearance occurred between 19.6 ($\pm$ 4.9) and 10.1 ($\pm$ 3.6) ago, in the Bulbostylis clade (table 1; fig. 1).

**PEPC Evolutionary History**

Using monocot ppc primers, 18 gene sequences were initially isolated from seven species (Ca. pendula, Ch. dodii, Cy. eragrostis, Cy. nipponicus, F. littoralis, P. sanguinolentus, and S. holoschoenus). Phylogenetic reconstructions showed that these genes belonged to five different gene lineages (named ppc-1, ppc-2, ppc-3, ppc-4, and ppc-5; fig. 2). Lineage ppc-1 (isolated from six species) is apparently sister to ppc-aL1 lineages of grasses. Closely related lineages ppc-2 (isolated only from Cy. nipponicus), ppc-3 (isolated only from S. holoschoenus), and ppc-4 (isolated from six different species) are positioned as sister to grass lineages ppc-aR, ppc-B1, and ppc-B2 (fig. 2). The fifth lineage, named ppc-5, was isolated only from F. littoralis, and is related to grass ppc-aL2 (fig. 2). In the phylogenetic tree, this gene was positioned inside grass ppc-aL2, which likely results from a bias due to the absence of other sedge ppc-5 and the small length of the isolated fragment (854 bp of coding sequence). Genes encoding C$_4$-PEPCs (with Ser780) were identified in E. vivipara (AB085948; Agarie et al. 2002), F. littoralis and P. sanguinolentus and were only detected in lineage ppc-1.

We then focused on the ppc-1 gene lineage for which 78 sequences were isolated from 63 sedge species (supplementary table 1, Supplementary Material online). The species relationships deduced from the phylogenetic tree inferred from coding sequences of this gene (fig. 3) were almost perfectly congruent with those deduced from plastid markers (fig. 1). The presence of two distinct ppc-1 gene clusters in C$_4$ Fimbristylis indicates that gene duplication has occurred before the divergence of this clade. In Eleocharis, up to three different genes were isolated from the same individual (in E. limosa). Inside this genus, species relationships deduced from ppc-1 genes do not concord with those inferred from plastid markers or ribosomal DNA internal transcribed spacer (ITS) sequences (Roalson and Friar 2000). The topology retrieved from ppc-1 introns was identical to that inferred from coding sequences (supplementary fig. 2, Supplementary Material online). The fine-scale phylogenetic incongruence between plastid, ITS, and ppc-1 genes points to a complex evolutionary history in Eleocharis.

At least one C$_4$ ppc-1 sequence (i.e., encoding PEPC with Ser 780) was isolated from all C$_4$ species analyzed. Five distinct C$_4$ ppc gene lineages were identified in the phylogenetic tree (fig. 3). They display a faster rate of nonsynonymous substitutions compared to non-C$_4$ ppc-1, as attested by the branch lengths inferred from amino acid sequences (supplementary fig. 3, Supplementary Material online). The ppc-1 gene isolated from Kyllinga underwent a large number of synonymous mutations, which led to a very long branch in the phylogeny inferred from nucleotides (fig. 3) but not in the amino acid tree (supplementary fig. 3, Supplementary Material online) indicating an increased evolutionary rate in this

<table>
<thead>
<tr>
<th>Lineage No</th>
<th>Clade Name</th>
<th>Stem Group Age</th>
<th>Crown Group Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>C$_4$ Rhynchospora</td>
<td>7.4 ($\pm$ 2.8)</td>
<td>4.2 ($\pm$ 2.2)</td>
</tr>
<tr>
<td>19</td>
<td>C$_4$ Fimbristylis</td>
<td>12.3 ($\pm$ 3.8)</td>
<td>5.8 ($\pm$ 2.6)</td>
</tr>
<tr>
<td>20</td>
<td>Bulbostylis clade</td>
<td>19.6 ($\pm$ 4.9)</td>
<td>10.1 ($\pm$ 3.6)</td>
</tr>
<tr>
<td>21</td>
<td>Eleocharis baldwini</td>
<td>10.5 ($\pm$ 3.2)</td>
<td>NA</td>
</tr>
<tr>
<td>22</td>
<td>Eleocharis vivipara</td>
<td>4.4 ($\pm$ 2.1)</td>
<td>NA</td>
</tr>
<tr>
<td>23</td>
<td>C$_4$ Cyperaceae</td>
<td>10.9 ($\pm$ 3.4)</td>
<td>8.4 ($\pm$ 2.9)</td>
</tr>
</tbody>
</table>
species. The five C₄ ppc groups correspond to the C₄ lineages inferred from plastid markers, except for Eleocharis baldwinii and E. vivipara, which formed two distinct lineages in the plastid tree (fig. 1) but strongly clustered together in the ppc-1 phylogeny (fig. 3).

Positive Selection

The model A, implementing positive selection in branches leading to sedge C₄ ppc, was significantly better than both null models (models M1a vs. A; chi-squared = 379.6, df = 2, P value < 0.0001; models A’ vs. A; chi-squared = 64.18, df = 1, P value < 0.0001). In model A, 31 sites were assigned to positive selection with a posterior probability greater than 0.95. By setting the threshold to 0.99, the number of codons under positive selection fell to 16. The amino acids encoded by these codons are shown for both C₄ and non-C₄ ppc-1 groups (fig. 3). Many parallel genetic changes are observed between the five C₄ ppc lineages on these codons. In addition to the alanine to serine transition at position 780 (numbered according to Zea mays sequence, CAA33317), the amino acid at position 540 changed five times independently from a proline to a threonine, the one at position 665 from a histidine to a asparagine and other positions underwent similar or identical changes between two and four times independently (fig. 3). Only five positions (572, 665, 733, 761, and 780) were significantly detected as evolving under positive selection during the evolution of C₄-specific PEPC in both sedges (fig. 3) and grasses (Christin et al. 2007). The amino acid changes at these positions were the same in both families and even in some eudicots (table 2; Christin et al. 2007). In addition, some C₄-adaptive changes detected in PEPC genes of either grasses or sedges were also observed in some lineages of...
Fig. 3—Evolutionary history of sedge PEPC encoding genes. This phylogenetic tree was obtained through Bayesian inference on coding sequences of ppc-1 genes from sedges. It was rooted on *Chrysitrix dodii* sequence (EMBL no FM208000). Bayesian support values are indicated near branches. Branches leading to C₄ ppc are in bold. For each C₄ and non-C₄ ppc clade, the most abundant amino acid for each codon under positive selection (numbered according to *Zea mays* sequence, CAA33317) is indicated on the right. C₄-specific amino acids are brightened and in bold. Taxon names were abbreviated to the genus plus the first letter(s) of the species (when several species of the same genus were analyzed).


**Table 2 Adaptive Changes Shared by Sedge and Grass C₄ ppc**

<table>
<thead>
<tr>
<th>Codon (No)</th>
<th>Group</th>
<th>N</th>
<th>572</th>
<th>665</th>
<th>733</th>
<th>761</th>
<th>780</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grasses</td>
<td>8</td>
<td>E</td>
<td>Q</td>
<td>7</td>
<td>H</td>
<td>N</td>
<td>6</td>
</tr>
<tr>
<td>Sedges</td>
<td>5</td>
<td>E</td>
<td>Q</td>
<td>4</td>
<td>H</td>
<td>N</td>
<td>5</td>
</tr>
<tr>
<td>Eudicots</td>
<td>5</td>
<td>E</td>
<td>Q</td>
<td>1</td>
<td>H</td>
<td>N</td>
<td>3</td>
</tr>
</tbody>
</table>

**NOTE.—** For the five positions (numbered according to Zea mays accession CAA33317) that were detected as evolving under positive selection in both sedges (this study) and grasses (Christin et al. 2007), the number of times C₄-linked amino acid mutations occurred is indicated, for sedges, grasses, and eudicots. The number of independent lineages (N), for which C₄ PEPC is available, is indicated for each group.

the other family, although they were not significantly attributed to positive selection in this second family (supplementary table 5, Supplementary Material online).

**Discussion**

Number and Tempo of C₄ Origins in Sedges

Six distinct C₄ lineages were present in the phylogenetic tree inferred from plastid markers (fig. 1). Analyses of *ppc*-1, which encodes the key enzyme of the C₄ pathway, unequivocally showed that this gene acquired C₄ attributes at least five times independently (fig. 3), confirming the multiple origins of the C₄ pathway in Cyperaceae. Intriguingly, E. baldwinii and E. vivipara, two C₄ species that are clearly distinguished based on plastid DNA (fig. 1), ITS (Roalson and Friar 2000), as well as phenotypic C₄ characteristics (Ueno et al. 1988; Soros and Dengler 2001), were strongly supported as monophyletic considering their C₄ ppc genes, which are very similar (fig. 3). This incongruence between ppc genes and other phylogenetic markers has been suggested to be due to phylogenetic bias caused by adaptive evolution (Roalson 2007) as seen for grass ppc genes (Christin et al. 2007). However, even when considering only introns, the C₄ ppc genes of E. vivipara and E. baldwinii clustered together with very high support (supplementary fig. 2, Supplementary Material online). These two unrelated taxa seem to have acquired their C₄ ppc from the same source. The very low divergence between these ppc genes (fig. 3; supplementary fig. 2, Supplementary Material online) rules out an acquisition of C₄ ppc during the early diversification of the Eleocharis genus followed by recurrent losses in C₃ species. It is thus likely that one of the two lineages evolved a C₄-specific ppc gene that was recently transmitted to the other taxa (probably E. vivipara; Roalson forthcoming), either through horizontal gene transfer or hybridization. C₄ evolution in these two Eleocharis lineages is thus not completely independent. Polyloidization is apparently very frequent in the genus Eleocharis (Yano et al. 2004; da Silva et al. 2008; Roalson 2008a) and an acquisition of C₄ genes through allopolyploidization between C₃ and C₄ parents is also likely. Further investigations are required to solve this issue. Species sampling of Eleocharis and number of molecular markers should especially be increased.

In the family Cyperaceae, the first C₄ evolution (Bulboystis) has likely occurred between 19.6 (± 4.9) and 10.1 (± 3.6) Ma, but the other C₄ appearances all arose during the last 12 My (table 1), making C₄ sedges generally younger than their grass counterparts (Christin, Besnard, et al. 2008). This raises questions about the drivers of these evolutionary events (Roalson 2008b). Indeed, if the Oligocene CO₂ decline created the general condition for C₄ photosynthesis to be advantageous in some environments (Christin, Besnard, et al. 2008; Roalson 2008b), other factors, such as heat, drought, fire, and other disturbances, could have locally selected for the C₄ pathway (Osborne 2008). Increasing perturbations of the midlatitude ecosystems throughout the Miocene have thus probably contributed to the success of C₄ sedges, as for C₄ grasses (Keeley and Rundell 2005; Beerling and Osborne 2006; Osborne 2008). However, C₄ sedges generally occupy wetter habitats than C₃ sedges within the fire-adapted vegetation (Linder and Rudall 2005), and they probably evolved their C₄ systems in a wetland context (Li et al. 1999; Stock et al. 2004). Photorespiration is generally reduced in such conditions because water availability allows stomatal aperture, which helps in maintaining significant intracellular CO₂ concentrations. It is thus possible that ecosystem-scale feedbacks between fire and vegetation have lately affected wetter grasslands, partially explaining why C₄ sedges have generally appeared later than C₃ grasses.

In Cyperaceae, however, a few C₃ and C₄ taxa occupy seasonally dry habitats, such as shallow soils of inselbergs in arid tropical savannah (Porembiski and Barthlott 2000; Proctor and Pence 2002; Proctor and Tuba 2002). Some taxa are desiccation tolerant (e.g., pokikoholyd) having the ability to evade drought, by rapid equilibration of the plant’s water content to that of the surrounding environment, and yet retain the ability to resume active metabolism when conditions are favorable. Desiccation tolerance has evolved independently in sedge tribes Trilepideae (C₃; at least 5 species of 17), Cariceae (C₅; 1 species), Abildgaardiae (C₄; 2 species), and Cyperaeae (C₄; 3 species). Nevertheless, this adaptive trait is a common characteristic only in the Trilepideae (e.g., Coleochloa and Microdrocoides), which have a crown age of 30.8 (± 7.5) My in our phylogenetic analysis (fig. 1). The few C₄ sedge taxa that have colonized inselbergs (e.g., Kyllinga alba, Bulbostylis leucoatrichya) may have acquired the desiccation-tolerance trait recently. The C₄ pathway, which raises the water-use efficiency compared with the C₃ type, has also probably contributed to such evolutionary transitions to arid habitats in these lineages as recently suggested in grasses (Osborne and Freckleton 2009).

Constraints in C₄ PEPC Evolution

In sedges, the five C₄ *ppc*-1 lineages have clearly undergone a very high number of nonsynonymous substitutions (supplementary fig. 3, Supplementary Material online). Positive selection analyses showed that this was due to important adaptive changes along the protein sequence, as in grasses (Christin et al. 2007). The evolution of C₄-specific enzymes likely involves not only changes of regulatory sequences to acquire the light-dependent expression pattern in M cells (Akyildiz et al. 2007), but also kinetic modifications (Dong et al. 1998; Blässing et al. 2000; Gowik et al. 2006; Rao...
et al. 2008) that were likely achieved through the adaptive amino acid changes identified on C₄ ppc-1.

The ppc lineages that evolved the C₄ function are clearly not randomly distributed in the phylogeny because only one ppc lineage of five in sedges became recurrently involved for the C₄ pathway. Similarly, of the six ppc gene lineages that exist in grasses, the ppc-B2 lineage developed a C₄ function at least eight times independently (Christin et al. 2007). Interestingly, gene lineages ppc-I and ppc-B2 are clearly not orthologous (fig. 2). Recurrent recruitments of the same gene lineages in Cyperaceae or Poaceae suggest that they have predispositions for the C₄ function. For instance, a non-C₄ gene allowing the acquisition of a light-induced and cell-specific expression through simple genetic changes (as shown in Flaveria; Akyildiz et al. 2007) is likely to be preferentially selected for the C₄ cycle. The possibility of easily altering the enzyme expression patterns likely constrained the origins of the C₄ PEPC to some ppc lineages. Once the C₄ transcription was acquired, kinetic properties had to be adapted to the new substrate and product concentrations present in the C₄ photosynthetic cells (Dong et al. 1998; Bläsing et al. 2000). The different C₄ sedge groups evolved a C₄ PEPC starting from very similar genes, and the high similarity of amino acid sequences of non-C₄ ppc-I (supplementary fig. 3, Supplementary Material online) suggests that, before C₄ evolution, genes belonging to this lineage had conserved functions and kinetic properties. This probably accounts for the recurrence of many changes on ppc-I in the five C₄ sedge lineages. The number of amino acid changes possible along a sequence is extremely large, but many of them would be detrimental because they would reduce or modify enzymatic activity. Out of the small proportion of amino acid changes that would be advantageous after the recruitment of a ppc gene in the C₄ pathway, many will be of reach. A new optimum is accessible only through step mutations that have to be all advantageous or at least neutral, which limits the number of possible evolutionary paths (Weinreich et al. 2006). The molecular convergence during sedge C₄-PEPC evolution probably results from both a limited number of new optima and the genetic background constraining the coverage of the protein space (Hodin 2000).

Intra and Interfamilial Genetic Convergence

Of the 16 codons with a high probability of having evolved under positive selection in branches leading to C₄ ppc of sedges, only five were also detected as C₄-adaptive in grasses (Christin et al. 2007). Despite showing a strong genetic convergence within sedges and grasses, respectively, C₄-specificities of PEPC strongly vary among families. This is probably due to the strong divergence of the non-C₄ genes that were recruited for the C₄ pathway in each of the two families (fig. 2). Grass ppc genes diverged from each other between 10 and 30 My before acquiring a C₄ function (Christin, Besnard, et al. 2008) and sedge ppc genes less than 40 My (fig. 1). On the other side, C₄ recruitments of ppc occurred more than 90 My after the divergence of grasses and sedges (Christin, Besnard, et al. 2008). Non-C₄ ppc genes of grasses and sedges are only distantly related and are putatively responsible for different functions, with other associated kinetics. Therefore, it is likely that the path to the C₄ optimum varies depending on the starting point. The same explanation probably accounts for the nontransferability of most grass C₄-adaptive amino acid changes to eudicot systems. A common starting point strongly constrained the evolutionary path to C₄-optimized enzymes in closely related plant lineages, whereas the important divergence of non-C₄ genes among major C₄ clades opened the road toward other C₄ optima.

Despite family-specific adaptive changes to fulfill the C₄ function, it is worth noting that five codons underwent similar changes in sedges, grasses, and eudicots (table 2). In addition to the alanine at position 780 that mutated to a serine at least five times in sedges, eight times in grasses, and five times in eudicots, the histidine at position 665 mutated to an asparagine at least five, six, and three times in the three groups, respectively. Recurrent changes in many C₄ groups are observed at these five positions (table 2). Whatever the gene recruited in the C₄ pathway, these five amino acids were often mutated to an identical residue. They are thus indicative of the very strongly convergent genetic evolution linked to the acquisition of C₄-specific PEPCs, and of the repeatability of some evolutionary processes, at the genetic level and across very broad taxonomic scales. These five sites are also potential universal determinants of the C₄ function and could be used to transform ppc genes of any plant to a C₄-specific enzyme, opening promising opportunities for the engineering of the C₄ pathway in C₃ plants (Hibberd et al. 2008). In the next decade, phylogenomic analyses of C₄ evolution should be extended to other families and additional enzymes toward a comprehensive understanding of the genetic mechanisms linked to the numerous C₄ origins. In the coming genomic era, multi-level C₄ comparative studies could position C₄ photosynthesis as a case study for molecular evolution.

Supplementary Material

Supplementary tables 1–5 and supplementary figures 1–3 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

This work was funded by Swiss NSF grant 3100A0-105886 to G.B. It partially results from a Switzerland–South Africa joint meeting, funded by the Swiss NSF grant IB32A0-112804. P. Danton, F. Dessimoz, N. Fumeaux, B. Gehrke, F. Jacquemoud, and O. Ueno provided very useful samples.

Literature Cited


Christin PA, Salamin N, Savolainen V, Duvall MR, Besnard G.

Bruhl JJ, Wilson KL. 2007. Towards a comprehensive survey of


David Irwin, Associate Editor

Accepted May 5, 2009