Integrating Phylogeny into Studies of C₄ Variation in the Grasses

Pascal-Antoine Christin*, Nicolas Salamin, Elizabeth A. Kellogg, Alberto Vicentini, and Guillaume Besnard

Department of Ecology and Evolution, University of Lausanne, 1015 Lausanne, Switzerland (P.-A.C., N.S., G.B.); Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland (N.S.); Department of Biology, University of Missouri, St. Louis, Missouri 63121 (E.A.K.); and Center for Tropical Forest Science, Smithsonian Tropical Research Institute, Apartado Postal 0843–03092, Panama, Republic of Panama (A.V.)

C₄ photosynthesis consists of morphological and biochemical novelties that create a CO₂ pump that concentrates CO₂ around RuBisCO (Kanai and Edwards, 1999), which decreases photorespiration and the resulting energy waste. Consequently, C₄ photosynthesis provides a competitive advantage in all conditions where photorespiration costs become important, especially at high temperatures and in arid and saline conditions (Sage, 2001). Despite being used by only 3% of extant angiosperm species (Sage, 2004), C₄ plants account for one-fifth of global terrestrial primary production (Ehleringer et al., 2008). The evolution of C₄ grasses dominate most open subtropical environments (Covshoff et al., 2008), leading to important differences in their transcriptomes (Sawers et al., 2007). In addition to the C₄ enzymes, C₄ photosynthesis evolution necessitated rearrangements of chloroplast envelope proteins (Bräutigam et al., 2008). Furthermore, transport of C₄ intermediates between M and BS cells is probably not performed through simple diffusion, which suggests that other, unidentified, mechanisms exist (Sowinski et al., 2008), which may be yet another C₄-specific adaptation.

Many of the enzymes that drive the carbon shuttle in C₄ plants also present in C₃ plants but are involved in other aspects of plant growth and development (Monson, 2003). Tissue-specific regulation of C₄ pathway enzymes appears to have been a crucial step in the evolution of C₄ photosynthesis (Hibberd and Quick, 2002). One aspect of the pathway that remains poorly understood is the genetic components regulating the alteration of leaf anatomy (Kellogg, 1999). The developmental and genetic issues can be addressed with all C₄ species, but the low number of model species used to date limits the generalization of the results.

Grasses have been the focus of much of the recent C₄ research. For example, human-directed improvement of C₃ grass crops, such as rice (Oryza sativa), barley (Hordeum vulgare), and wheat (Triticum aestivum), by introgression of C₄ characteristics is receiving particular attention (Hibberd et al., 2008). Understanding the historical causes of C₄ evolutionary and ecological success is another area of intense research activity (Cerling et al., 1997; Beierling and Osborne, 2006; Osborne and Beierling, 2006; Osborne, 2008). The ecological importance of grasses made this family a natural study system for investigating factors affecting the distribution and success of C₄ plants (Taub, 2000; Carmo-Silva et al., 2007; Cabido et al., 2008; Edwards and Still, 2008). For instance, it has recently been shown that the oldest C₄ origin in grasses is relatively young (approximately 30 million years old), and correlates with a marked decrease of atmospheric CO₂ concentration (Christin et al., 2008; Vicentini et al., 2008). Since atmospheric CO₂ concentration and air

* Corresponding author; e-mail pascal-antoine.christin@unil.ch.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Pascal-Antoine Christin (pascal-antoine.christin@unil.ch), www.plantphysiol.org/cgi/doi/10.1104/pp.108.128553
### Table 1. Characteristics of the C<sub>4</sub> grass lineages

<table>
<thead>
<tr>
<th>No.</th>
<th>Name</th>
<th>Age Estimates</th>
<th>No.</th>
<th>C&lt;sub&gt;4&lt;/sub&gt; Sister</th>
<th>Crops</th>
<th>Habitat</th>
</tr>
</thead>
<tbody>
<tr>
<td>1&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Stipagrostis</td>
<td>15.1 (±4.6) – 7.5 (±3.1)</td>
<td>50</td>
<td>Sartidia&lt;sup&gt;a&lt;/sup&gt;</td>
<td>NADP-ME –</td>
<td>Deserts and semideserts</td>
</tr>
<tr>
<td>2&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Aristida</td>
<td>28.8 (±5.2) – 14.4 (±4.7)</td>
<td>290</td>
<td>Sartidia&lt;sup&gt;a&lt;/sup&gt;</td>
<td>NADP-ME –</td>
<td>Large ecological range</td>
</tr>
<tr>
<td>3&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Core Chloroideae</td>
<td>32.0 (±4.4) – 25.0 (±4.0)</td>
<td>1,410</td>
<td>Merswmuelleria range&lt;sup&gt;a&lt;/sup&gt;</td>
<td>NAD-ME and PCK</td>
<td>Finger millet, teff</td>
</tr>
<tr>
<td>4&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Centropodia</td>
<td>22.0 (±4.6) – 11.3 (±5.5)</td>
<td>4</td>
<td>M. range&lt;sup&gt;a&lt;/sup&gt;</td>
<td>NAD-ME –</td>
<td>Dry open habitats (semideserts)</td>
</tr>
<tr>
<td>5&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Eriachne</td>
<td>11.5 (±3.6) – 6.6 (±2.8)</td>
<td>40</td>
<td>Isachne&lt;sup&gt;a&lt;/sup&gt;</td>
<td>NADP-ME –</td>
<td>Warm open habitats (savannah)</td>
</tr>
<tr>
<td>6&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Arundinelleae</td>
<td>26.4 (±4.4) – 7.9 (±3.4)</td>
<td>95</td>
<td>Centotheceae 2&lt;sup&gt;k&lt;/sup&gt;</td>
<td>NADP-ME –</td>
<td>Large ecological range</td>
</tr>
<tr>
<td>7&lt;sup&gt;cd&lt;/sup&gt;</td>
<td>Panicum/Urochloa/Scataria clade</td>
<td>18.5 (±3.7) – 16.4 (±3.6)</td>
<td>&gt;530</td>
<td>C&lt;sub&gt;3&lt;/sub&gt; Neurachne&lt;sup&gt;a&lt;/sup&gt;</td>
<td>NADP-ME, NAD-ME, and PCK</td>
<td>Foxtail, pearl, and proso millets</td>
</tr>
<tr>
<td>8&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Neurachne munroi</td>
<td>[15.9 (±3.7) – 13.1 (±3.2)]</td>
<td>4.4 (±3.3) – present</td>
<td>1 Neurachne teniuifolia&lt;sup&gt;a&lt;/sup&gt;</td>
<td>NADP-ME –</td>
<td>Dry open habitats (steppes)</td>
</tr>
<tr>
<td>9&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Echinochloa</td>
<td>13.8 (±3.5) – 4.4 (±2.8)</td>
<td>30–40</td>
<td>Paraphylochloa&lt;sup&gt;a&lt;/sup&gt;</td>
<td>NADP-ME –</td>
<td>Warm open habitats</td>
</tr>
<tr>
<td>10&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Alloteropsis</td>
<td>15.3 (±3.5) – present</td>
<td>4–7</td>
<td>Forest shade clade&lt;sup&gt;a&lt;/sup&gt;</td>
<td>NADP-ME and PCK –</td>
<td>Warm open habitats (savannah)</td>
</tr>
<tr>
<td>11&lt;sup&gt;cd&lt;/sup&gt;</td>
<td>Digitaria</td>
<td>21.2 (±3.9) – 8.1 (±3.4)</td>
<td>220</td>
<td>x = 9 Paniceae&lt;sup&gt;a&lt;/sup&gt;</td>
<td>NADP-ME Fonio</td>
<td>Various warm open habitats</td>
</tr>
<tr>
<td>12&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Andropogoneae</td>
<td>[15.9 (±3.7) – 5.4 (±2.1)]</td>
<td>1,085</td>
<td>x = 10 Paniceae&lt;sup&gt;a&lt;/sup&gt;</td>
<td>NADP-ME –</td>
<td>Maize, sorghum, sugarcane</td>
</tr>
<tr>
<td>13&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Paspalum clade</td>
<td>[24.3 (±4.9) – 19.1 (±4.5)]</td>
<td>14.1 (±3.4) – 8.5 (±3.1)</td>
<td>&gt;345 Streptoch sos asperifolia&lt;sup&gt;a&lt;/sup&gt;</td>
<td>NADP-ME Kodo millet</td>
<td>Warm open habitats (savannah)</td>
</tr>
<tr>
<td>13&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Ophiochloa clade</td>
<td>[11.7 (±3.3) – present]</td>
<td>115</td>
<td>S. asperifolia&lt;sup&gt;a&lt;/sup&gt;</td>
<td>NADP-ME –</td>
<td>Large ecological range</td>
</tr>
<tr>
<td>14&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Anthaenantia&lt;sup&gt;a&lt;/sup&gt;</td>
<td>14.3 (±3.5) – present</td>
<td>1</td>
<td>Steinchisma clade&lt;sup&gt;a&lt;/sup&gt;</td>
<td>NADP-ME –</td>
<td>Warm open habitats (savannah)</td>
</tr>
<tr>
<td>15&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Streptochos ramosa</td>
<td>[15 (±3.7) – present]</td>
<td>15.5 (±3.5) – present</td>
<td>1 Cyphonanthus&lt;sup&gt;i&lt;/sup&gt;</td>
<td>NADP-ME –</td>
<td>Warm open habitats (savannah)</td>
</tr>
<tr>
<td>16&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Panicum prionitis clade</td>
<td>[16.3 (±3.7) – present]</td>
<td>10.4 (±2.9) – 6.3 (±2.7)</td>
<td>&gt;5 Arthropon lanceolatus&lt;sup&gt;k&lt;/sup&gt;</td>
<td>NADP-ME –</td>
<td>Warm open habitats (savannah)</td>
</tr>
<tr>
<td>17&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Mesosetum clade</td>
<td>[11.2 (±2.9) – present]</td>
<td>12.3 (±3.2) – 11.3 (±3.0)</td>
<td>40 Homolepis&lt;sup&gt;k&lt;/sup&gt;</td>
<td>NADP-ME –</td>
<td>Warm open habitats (savannah)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Christin et al. (2008).  
<sup>b</sup>Independent origin confirmed by PEPC analyses (Christin et al., 2007).  
<sup>c</sup>Independent origin based on putative species relationships only.  
<sup>d</sup>Phylogeny from Vicentini et al. (2008) found Digitaria and the main x = 9 Paniceae C<sub>4</sub> clade clustered together, suggesting a single C<sub>4</sub> origin.  
<sup>e</sup>Previously named Leptocoryphium lanatum.  
<sup>f</sup>Christin et al. (2008) and Vicentini et al. (2008) into square brackets, ages are given in millions of years.  
<sup>g</sup>Duvall et al. (2007).  
<sup>h</sup>Sánchez-Ken et al. (2007).  
<sup>i</sup>Morrone et al. (2008).  
<sup>j</sup>C<sub>3</sub> subspecies of A. semialata could represent a reversion from C<sub>4</sub> to C<sub>3</sub> (Ibrahim et al., 2009).  
<sup>k</sup>Vicentini et al. (2008).  
<sup>l</sup>Morrone et al. (2007).  
<sup>m</sup>Excluding foders.  
<sup>n</sup>Watson and Dallwitz (1992).
temperature both affect C₄ plant success, the current changes in global climate will potentially trigger important perturbations in major ecosystems, and could affect the performance of extensively cultivated tropical cereals. Therefore, a complete understanding of C₄ ecology and physiology is necessary for conservation biology and agriculture to face future climate changes (Sage and Kubien, 2003; Ainsworth et al., 2008).

Comparative analyses offer an attractive approach for both the study of genetic determinants of C₄ photosynthesis (Christin et al., 2007) and the identification of attributes associated with it (Edwards et al., 2007; Edwards and Still, 2008). Such an approach requires comparing several independent origins of C₄ plants to determine characteristics that are shared among them. Indeed, if two C₄ species inherited the C₄ trait from their common ancestor, they do not represent independent replicates. Ideally, comparative studies should consist of distinct C₃ clades, known to represent distinct origins of the C₄ pathway, as well as C₃ sister groups to each of the C₄ lineages. For this approach to work, species relationships have to be assessed by phylogenetic analyses, rendering the phylogenetic framework of systematic botany useful to evolutionary and physiological investigations.

C₄ EVOLUTIONARY LINEAGES IN GRASSES

The grass family is composed of approximately 10,000 species, of which about 45% are C₄ (Sage, 2004). Grass taxonomy recognizes between 12 and 13 main subfamilies but all C₄ grasses belong to the PACMAD clade (Fig. 1; Duvall et al., 2007; or PACC-MAD, Sánchez-Ken et al., 2007). Both the distribution of C₄ grasses in distinct taxonomic groups and the high variability of their C₄ syndrome led to the inference of multiple origins of the C₄ pathway in this family (Sinha and Kellogg, 1996; Kellogg, 2001). Phylogenetic analyses of the subfamily Panicoideae further suggested that C₄ photosynthesis appeared several times independently, although a single appearance followed by multiple reversions could not be excluded (Giussani et al., 2001; Duvall et al., 2003; Vicentini et al., 2008). The ancestral state reconstructions adopted in these studies are strongly dependent on species sampling and rely on statistical methods whose assumptions can produce different results. In addition, the transition rate from C₃ to C₄ could also change through time (Vicentini et al., 2008), for instance as a function of atmospheric CO₂ levels (Christin et al., 2008) or after the acquisition of preadaptations to C₄ photosynthesis (Sage, 2001). Finally, inferences of characters that affect the rates of speciation or extinction can yield erroneous conclusions if not carefully considered (Goldberg and Igic, 2008).

Some studies have thus focused on the evolutionary dynamics of specific key enzymes involved in the C₄ pathway, in particular phosphoenolpyruvate carboxylase (PEPC). The use of PEPC for the atmospheric CO₂ fixation is one of the rare characteristics common to all C₄ plants (Sinha and Kellogg, 1996; Sage, 2004), and its recruitment is an important step in the integration and optimization of C₄ biochemistry (Svensson et al., 2003) and can be considered as a critical event in the evolution into a C₄ plant. The presence of a Ser at position 780 of PEPC (numbered based on the maize sequence) is required for C₄ function (Svensson et al., 2003) and was accompanied by many other recurrent adaptive amino acid changes (Christin et al., 2007) that left reliable C₄-specific genetic signatures. Because changes along a DNA sequence are amenable to statistical modeling, they can easily be traced on a PEPC phylogenetic tree. This technique was used to identify the grass lineages that likely evolved the C₄ trait independently (Table I; Christin et al., 2007, 2008).

C₄ MODEL SPECIES IN GRASSES

The grasses contain few examples of closely related C₃/C₄ pairs, and those that exist are not easily accessible. *Alloteropsis semialata* contains a C₃ and a C₄ subspecies, which are closely related (Ibrahim et al., 2009) but differ in chromosome number (Liebenberg and Fossey, 2001) and so are presumably intersterile. A recent phylogenetic study suggested that C₃ subspecies of *A. semialata* could represent an evolutionary reversion from C₄ to C₃ photosynthesis (Ibrahim et al., 2009). The genus *Neurachne* includes both C₃ and C₄ species (Moore and Edwards, 1989); these are native to Australia and grow in relatively inaccessible parts of the continent and have not, to our knowledge, been cultivated. The C₃/C₄ intermediate *Steinchisma hians* (formerly *Panicum milioides*) is sister to a group of C₃ species, and has been crossed with them (Brown et al., 1985). *Steinchisma* as currently circumscribed is mainly South American.

Historically much of the work on C₄ grasses focused on the genus *Panicum* because it appeared to have species with all possible photosynthetic pathways. Unfortunately, this genus was an assemblage of unrelated species (Aliscioni et al., 2003) whose taxonomy is being completely redefined (Morrone et al., 2007, 2008; Sede et al., 2008). The name *Panicum* should be restricted to a set of species that are all C₄ with the subtype using the NAD-malic enzyme (NAD-ME), including switchgrass (*Panicum virgatum*). C₃ species of *Panicum* are not closely related to true *Panicum* (Aliscioni et al., 2003).

Future C₄ research should consider additional C₄ species systems since including other independent lineages would increase the power of comparative analyses. In particular, *Aristida* and *Stipagrostis*, as well as the subfamily Chloridoideae, represent interesting C₄ lineages. These groups are ecologically important (Table I) and strongly differ from the Panicoideae C₄ species in terms of ecological attributes, such as aridity tolerance (Taub, 2000; Sato and Kubota, 2004; Carmo-Silva et al., 2007). They are species rich and widely
INTEGRATING PHYSIOLOGICAL STUDIES IN A PHYLOGENETIC CONTEXT

Understanding C₄-specific growth, survival, and reproductive success, as well as the environmental conditions that influence these traits, is of prime ecological, agricultural, and evolutionary importance. Assessment of plant physiological traits, such as photosynthetic activity and efficiency, is time consuming, especially when performed under a range of environmental conditions. Therefore, physiological studies typically only consider a limited number of species. Unfortunately, due to the strong variations of the C₄ pathway (Sinha and Kellogg, 1996), all C₄ plants are far from being equivalent. Species sampling for physiological investigations is crucial to ensure the generalization of conclusions. As noted above, taxa that inherited their C₄ trait from a common ancestor do not represent independent replicates. Their common ancestry can potentially lead to spurious correlations, which in turn can entangle characteristics due to the C₄ trait and those resulting from a close phylogenetic relationship (Taub, 2000). A sound phylogenetic framework showed that a low carbonic anhydrase activity, previously attributed to C₄ grasses (Gillon and Yakir, 2001), characterizes the whole PACMAD clade.
and is not linked to the C₄ trait (Edwards et al., 2007). Thanks to its highly convergent nature, the C₄ trait is present in numerous natural replicates. Species sampling for C₄ physiological studies can take advantage of this by comparing species from independent C₄ lineages, as well as each C₄ clade with its C₃ sister group (Table I). Therefore, species relationships deduced from molecular markers should serve as a guide for species sampling.

As a C₄ study system, the grass family allows combining physiological, ecological, genomic, and evolutionary approaches, which are all necessary for a complete understanding of C₄ photosynthesis. Integration of the wide knowledge we are gaining about C₄ grasses to reach a full picture requires incorporation of evolutionary history by using phylogenetic information. Important efforts have led to a reasonably well-resolved phylogenetic tree for the grass family (e.g., Grass Phylogeny Working Group, 2001; Aliscioni et al., 2003; Duvall et al., 2007; Christin et al., 2008; Vicentini et al., 2008) but conflicts between plastid and nuclear markers (Fig. 1) still need to be resolved. Recent analyses of C₄ genes have identified grass lineages that evolved the C₄ pathway independently (Christin et al., 2007, 2008). These correspond to more than 15 independent replicates (Fig. 1), enabling wide-scale comparative studies to sort general attributes of C₄ plants as well as particular ones. By taking advantage of the convergent nature of C₄ photosynthesis, multidisciplinary studies in the grasses could bring a complete view of the selective pressures and genetic mechanisms responsible for the evolution of C₄ photosynthesis and the factors that control the current distribution and success of C₄ plants. C₄ photosynthesis in grasses could become a model of macroevolution process when completely elucidated, from the selective pressures to the genetic mechanisms that led to its appearances.

Received August 28, 2008; accepted October 30, 2008; published January 7, 2009.

LITERATURE CITED

proteomes in maize mesophyll and bundle sheath cells. Mol Cell Proteomics 7: 1609–1638


