Supporting Information

Christin et al. 10.1073/pnas.1216777110

SI Materials and Methods

Phylogenetic Inference. A previously published 545-taxa dataset of the grasses based on the plastid markers *rbcL*, *ndhF*, and *trnK-matK* (1) was expanded and used for phylogenetic inference. For species sampled for anatomical cross-sections but not included in the published dataset, the markers *ndhF* and/or *trnK-matK* were either retrieved from GenBank when available or were newly sequenced from extracted genomic DNA with the method and primers described previously (1, 2). These new sequences were aligned to the dataset, excluding the regions that were too variable as described previously (1). The final dataset totaled 604 taxa and was used for phylogenetic inference as implemented in the software Bayesian Evolutionary Analysis by Sampling Trees (BEAST) (3).

The phylogenetic tree was inferred under a general time-reversible substitution model with a gamma-shape parameter and a proportion of invariants (GTR+G+I). *Flagellaria indica* (Flagellariaceae) was used to root the tree, following previous results (4). A relaxed molecular clock was implemented with rates that followed a log-normal distribution. No time constraint was used except for the crown of the BEP-PACMAD clade, which was modeled by a normal distribution with a mean of 51.2 and a SD of 6. The speciation pattern was set to follow a Yule process. The starting values and operators were optimized through repeated analyses. Two final analyses were run each for 20,000,000 generations. A tree was sampled every 5,000 generations after a burn-in period of 5,000,000, which was enough for the analysis to converge, as verified with the software TRACER (5). A consensus tree was then computed, and the tree was pruned to include only the species sampled for cross-sections using the package analyses of phylogenetics and evolution APE (6) in R. The anatomical measurements of *Spartina alterniflora* were assigned to its allopolyploid species *Spartina anglica* in the phylogeny, and those of *Stipagrostis obtusa* were assigned to one of the representative of this monophyletic genus. Two species, for which cross-sections have been measured, were not included in the phylogenetic reconstruction because of a lack of good-quality DNA. These species were used for the description of variables but were excluded from phylogeny-based analyses.

The phylogenetic tree was used to test for the existence of precursors, which might have evolved in some clades and allowed the evolution of C_4 photosynthesis in its descendants. This recently developed approach takes a two-state character (C_3 vs. C_4) and models it as a three-state character (C_3, precursor, C_4), where the transition between C_3 and C_4 is done in two steps, from C_3 to precursor and then from precursor to C_4 (7). In the most complex model currently implemented in r8s (precursor_2), the rate is different for C_3/precursor and for precursor/C_4 transitions, but each transition has an equal rate of reversal. This model was compared with the precursor_1 model (identical rate for C_3/precursor and precursor/C_4 transitions) and the two-state model (different rates for C_3-to-C_4 and C_4-to-C_3 transitions) using Akaike information criteria.

Plant Material and Cross-Sections. Mature leaves were either sampled from plants grown in the greenhouse, collected in the wild, or from dried herbarium specimens. When possible, the second or third leaf below the inflorescence was used. Our sampling was supplemented with 30 slides made available from previous studies (8–10). Leaf blade segments from dried herbarium material were rehydrated in water and 10% (wt/wt) aerosol OT for 48 h. Most samples were stored in 100% (vol/vol) ethanol for at least 2 wk. Segments from the center of leaf blades were then embedded in resin (JB-4; Polysciences), following the manufacturer's instructions. Five-micrometer thick cross-sections of the embedded leaf fragments were cut with a microtome and stained with saturated cresyl violet acetate (CVA). Some samples were fixed in formalin-propionic acid-alcohol (FPA), embedded in paraffin, sectioned at 10 μm, and stained with a safranin O-orange G series (11) as described in (12). All slides were made permanent and are available on request.

Anatomical Measurements. All C_3 grasses possess a double BS, with the outer layer derived from ground meristem to form a "parenchyma sheath," and the internal layer derived from the vascular procambium to form a "mestome sheath" (13). Many C_4 grasses also possess these two BS layers, with one of them specialized in carbon reduction and usually referred to as "carbon reduction" or "Kranz" tissue (14). However, this Kranz tissue can correspond either to the OS or the IS of C_3 grasses (13, 15). In addition, several C_4 grasses possess only one sheath, which in all investigated species derives from procambium, and is, therefore, homologous to the IS (= "mestome") of C_3 grasses (15). The terms "mestome" and "Kranz" are, therefore, functionally oriented and are evolutionarily ambiguous. The two sheaths are consequently referred to as IS and OS throughout this manuscript. The definition of all anatomical characters measured in this study aimed at allowing comparison (through homology) among all of the considered taxa.

Photographs of the cross-sections were taken at a magnification of 20× using a Nikon Eclipse E600 compound light microscope (Nikon Instruments) linked to a Nikon DXM1200C digital camera. An area extending from the middle of a vascular bundle to the middle of a distant bundle was measured with ImageJ software (16). When possible, the median vascular bundle was not considered and the measured areas encompassed lateral bundles of different orders. On the selected section, the areas of M, OS, IS, and vascular tissue (total area inside of the IS) were measured together with other characters (Fig. S1). The M was defined as all of the tissue between the epidermis and BS. Structural cells without photosynthetic potential, such as fibers and bulliform cells, were excluded. The OS was defined as the single layer of concentrically arranged cells directly outside the IS. If additional OS cells were present (e.g., BS extension), they were excluded (from both M and OS areas) for comparative purposes. Only cells that were visually differentiated from the surrounding M were considered, and incomplete OS were scored for several taxa. The IS was defined as the single layer of cells that surrounds the vascular tissue (xylem, phloem, and fibers together). In most cases, the IS was complete (i.e., formed a complete ring), but a few exceptions existed, resulting in the measurement of partial IS. When only one BS was present (C_4 taxa only), this was considered as the IS, following previous observations in three independently evolved C_4 taxa (15). It is, however, possible that some C_4 lineages that have not been studied developmentally have a single BS that corresponds to the OS. The measured cross-sectional areas were used to calculate, for each leaf, the proportion of OS compared with M [%OS = OS area/(OS area + M area)] and the proportion of IS compared with M [%IS = IS area/(IS area + M area)]. The center of starch production was identified by the accumulation of chloroplasts, as observed on the cross-sections and described in previous works (8–10), and species were consequently classified as C_4 using the OS (C_4-OS) or C_4 using the IS (C_4-IS) as appropriate. Two species in our sampling are known to use C_2 photosynthesis [also known as C_3-C_4 photosynthesis (17)]

based on the OS (in the genus *Steinchisma*). Because C_2 photosynthesis also represents a derived state over the ancestral C_3 trait, these two species were classified as C_4-OS.

The BSD, defined as the mean distance of M from the outer wall of one BS to the outer wall of an adjacent BS, was also measured on the same sections. In addition, the IVD was measured as the mean distance between centers of veins through the M. In addition, the number of M cells between two adjacent bundles and their longest diameters were measured for a minimum of 15 cells following the shortest path through M. The width of the OS and IS cells was measured for the two most-equatorial cells of each vein along the diameter of the corresponding vein. Leaf thickness was measured from upper epidermis to lower epidermis at the major lateral veins (veins accompanied by fibers and/or separated by bulliform cells). A second leaf thickness (thickness2) was measured from upper epidermis to lower epidermis at the thinnest point between two major veins. A ratio of thicknesses was computed as thickness2 divided by thickness, which estimates whether the leaf is flat (ratio = 1) or thinner between veins than at veins (ratio < 1). The average distance between major veins was also measured (IVDm). All measurements were averaged over the analyzed area and are given in micrometers or square micrometers.

Intraspecific Replicates. To evaluate the intraspecific variability in anatomical traits, measurements were repeated on individuals belonging to the same species but grown in different conditions and/or that originated from different geographic origins (Dataset S1). Replicates were obtained for 15 species, including one outgroup, two C_4, nine C_3 PACMAD, and three C_3 BEP. All measured traits varied within species, which can be attributable to phenotypic plasticity and intraspecific genetic variation. The impact of this variation on the patterns observed in this study was evaluated by linear regressions. For each species, the sample with the largest measured total area was defined as the reference and the sample with the smallest measured area was defined as the duplicate. The R^2 of the linear regression (forcing the intercept to 0) of the duplicates on the references for the different traits ranged from 0.74 (for thickness2) to 0.98 (for OS width), except for the mean vein area ($R^2 = 0.59$). These results indicate that despite intraspecific variation, the trends are consistent among different samples. Consequently, only reference samples were included in comparative analyses.

Comparative Analyses. We used different phylogeny based methods to model the evolution of anatomical characters while correcting for historical factors. All comparative analyses were performed in R using the packages APE (6), GEIGER (18), OUCH (19) and CAPER (20). All variables were log-transformed except for the three ratios (%OS, %IS, and ratio of thicknesses). To differentiate taxonomic and photosynthetic effects, we modeled the variables separately first for the three major clades but excluding C_4 taxa (outgroups, BEP, and PACMAD) and then for all photosynthetic types but in PACMAD only (C_3 PACMAD, C_4-OS, and C_4-IS). The best evolutionary model was selected through hierarchical likelihood ratio tests between a Brownian motion model, an OU model with a single optimum, an OU model with two different optima (best-fit of three possible groupings), and an OU model with three different optima (one per group).

Further analyses were performed on C_3 taxa only to avoid potential biases attributable to different selective regimes and optima in C_3 and C_4 taxa. Correlations between variables were investigated by phylogenetic generalized least squares (GLS) with a lambda correction. Ancestral character states were inferred for all nodes of the tree with a maximum likelihood criterion under a Brownian motion model. The ancestral character states were also estimated for all nodes of the tree that includes both C_3 and C_4 taxa, as well as for the tree that includes all C_3 taxa and all C_4 taxa that possess an OS. To reconstruct the changes that occurred during C_3 to C_4 transitions, the value of the most recent C_3 ancestor of each C_4 lineage was compared with the value at the crown group node of each C_4 lineage. The value of the most recent C_3 ancestor was extracted from the tree that contained C_3 taxa only to avoid values artificially close to C_4 values. Because the node that corresponds to the split of the C_4 lineage was not present in this tree, the average between the nodes that precede and succeed it was considered as representative of the C_3 ancestor. The value for the C_4 crown was extracted from the analysis based on the tree with both C_3 and C_4 taxa. (For %OS, OS area, and OS width, the tree with C_3 taxa and C_4 taxa that possess an OS was used.) For these analyses, the C_4-IS and C_4-OS lineages of *Alloteropsis* were treated as independent C_4 origins, although the most likely scenario involves the existence of unknown intermediate stages (21). The C_4 lineage Tristachyideae was excluded from these analyses because the C_4 type of its ancestor (C_4-IS or C_4-OS) is unknown.

C_3/C_4 transitions were modeled by taking into account a potential effect of anatomical traits. In the case of significant anatomical enablers, the rate of C_4 evolution would be very small below (or above) a given value of the anatomical character and increase once the character becomes larger (or smaller). This was modeled by collapsing to 10^{-5} all branches of the phylogeny for which the reconstructed value of the anatomical trait was below (or above) a given value. The transformed topology was used to model C_3/C_4 with a two-parameter model, and the threshold for the change of probabilities was optimized from the data under maximum likelihood as implemented in the software BayesTraits (22). The model that assumes an increase of transition rates after a given value of an anatomical character can be compared with the model without branch transformation through a likelihood ratio test with one degree of freedom. To circumvent the issue of strong differences between C_3 and C_4 values, the models were optimized on a phylogeny that contained only C_3 taxa and C_3 species were typed as "sister to C_4-OS" or "sister to C_4-IS," which was then used as a proxy for C_4 origins. Because C_3/C_4 transitions happened on branches separating C_4 lineages and their C_3 sister groups, asking whether a given anatomical trait increases the probability of becoming an actual C_4 group or sister to a C_4 group is equivalent. C_3 species not sister to C_4 lineages were given a score of 0, whereas the others were given a score of 1. The C_4-IS groups Andropogoneae and Anthephorinae are both sister to groups that encompass other C_4 lineages, as well as their respective C_3 sister groups (1). All C_3 species in these two groups were consequently given a score of 1 (sister to C_4-IS). The C_3 taxa sister to C_4-IS lineages nested within each of these groups were given a score of 2 to model additional transitions. The trait was modeled as a multistate character where transition rates q01 = q12, q21 = q10, q02 = q20 = 0, forcing the transitions to occur on branches connecting C_3 ancestors and C_4-IS descendants.

1. Grass Phylogeny Working Group II (2012) New grass phylogeny resolves deep evolutionary relationships and discovers C_4 origins. *New Phytol* 193(2):304–312.
2. Taylor SH, et al. (2012) Photosynthetic pathway and ecological adaptation explain stomatal trait diversity amongst grasses. *New Phytol* 193(2):387–396.
3. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7:214.
4. Givnish TJ, et al. (2010) Assembling the tree of the Monocotyledons: Plastome sequence phylogeny and evolution of Poales. *Ann Mo Bot Gard* 97(4):584–616.
5. Rambaut A, Drummond AJ (2007) Tracer v1.4. Available at http://beast.bio.ed.ac.uk/Tracer. Accessed June 17, 2010.
6. Paradis E, Claude J, Strimmer K (2004) APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* 20(2):289–290.

7. Marazzi B, et al. (2012) Locating evolutionary precursors on a phylogenetic tree. *Evolution* 66(12):3918–3930.
8. Renvoize SA (1985) A survey of leaf-blade anatomy in grasses V. The bamboo allies. *Kew Bull* 40(3):509–535.
9. Renvoize SA (1986) A survey of leaf-blade anatomy in grasses VIII. *Arundinoideae*. *Kew Bull* 41(2):323–338.
10. Renvoize SA (1987) A survey of leaf-blade anatomy in grasses XI. *Paniceae*. *Kew Bull* 42(3):739–768.
11. Sharman BC (1943) Tannic acid and iron alum with safranin and orange G in studies of the shoot apex. *Stain Technol* 18(3):105–111.
12. Columbus JT (1998) Morphology and leaf blade anatomy suggest a close relationship between *Bouteloua aristidoides* and *B*. (*Chondrosium*) *eriopoda* (Gramineae: Chloridoideae). *Syst Bot* 23(4):467–478.
13. Brown WV (1975) Variations in anatomy, associations, and origins of Kranz tissue. *Am J Bot* 62(4):395–402.
14. Dengler NG, Dengler RE, Donnelly PM, Hattersley PW (1994) Quantitative leaf anatomy of C_3 and C_4 grasses (Poaceae): Bundle sheath and mesophyll surface area relationships. *Ann Bot (Lond)* 73(3):241–255.
15. Dengler NG, Dengler RE, Hatterlsey PW (1985) Differing ontogenetic origins of PCR ("Kranz") sheaths in leaf blades of C_4 grasses (Poaceae). *Am J Bot* 72(2):284–302.
16. Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* 9(7):671–675.
17. Sage RF, Sage TL, Kocacinar F (2012) Photorespiration and the evolution of C_4 photosynthesis. *Annu Rev Plant Biol* 63:19–47.
18. Harmon LJ, Weir JT, Brock CD, Glor RE, Challenger W (2008) GEIGER: Investigating evolutionary radiations. *Bioinformatics* 24(1):129–131.
19. Butler MA, King AA (2004) Phylogenetic comparative analysis: A modeling approach for adaptive evolution. *Am Nat* 164(6):683–695.
20. Orme CDL, et al. (2012) Caper: Comparative analyses of phylogenetics and evolution R. Package version 0.5. Available at http://cran.r-project.org/web/packages/caper/index.html. Accessed August 14, 2012.
21. Christin PA, et al. (2012) Adaptive evolution of C_4 photosynthesis through recurrent lateral gene transfer. *Curr Biol* 22(5):445–449.
22. Pagel M, Meade A, Barker D (2004) Bayesian estimation of ancestral character states on phylogenies. *Syst Biol* 53(5):673–684.

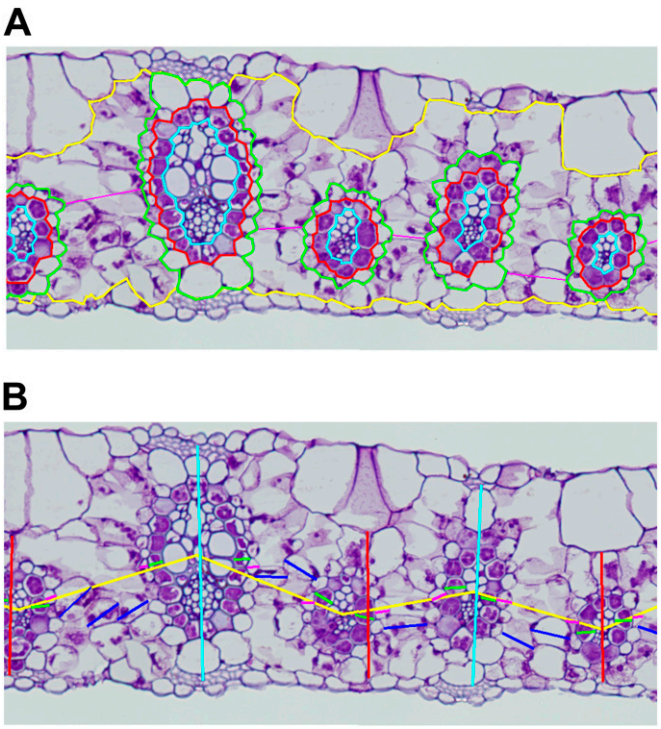

Fig. S1. Representation of the measured areas and lengths. The measured characters are shown on a leaf section of the C_4-IS *Alloteropsis semialata*. (*A*) BSD (purple lines) and main areas (yellow, M; green, OS; red, IS; light blue, veins). (*B*) Other lengths (yellow, IVD; light blue, thickness; red, thickness2; purple, OS size; green, IS size; blue, M size).

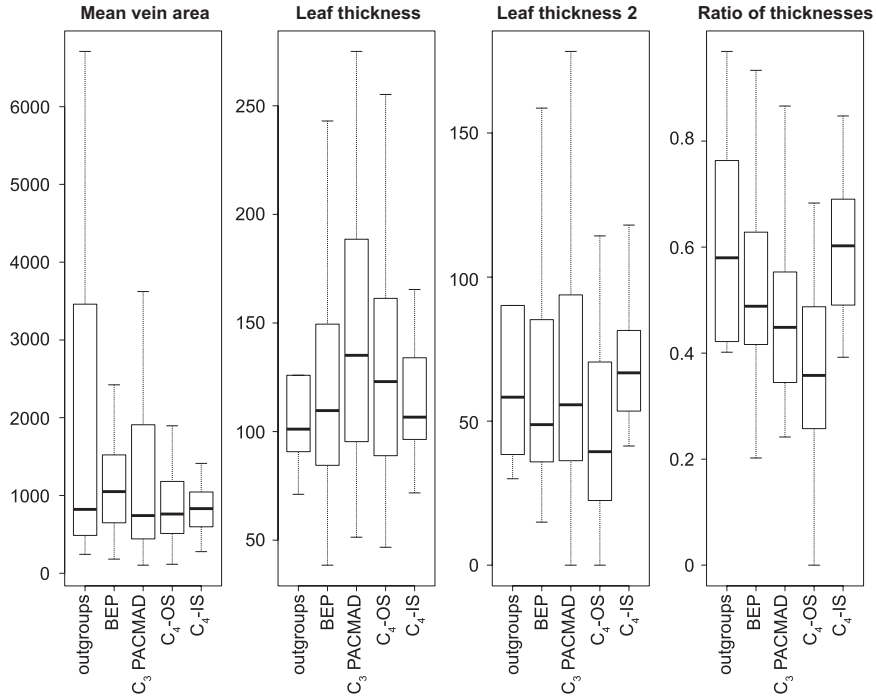

Fig. S2. Distribution of trait values among clades and photosynthetic types. Boxes indicate the 25th and 75th percentiles, and whiskers indicate the extreme values still within 1.5 interquartile range of the upper and lower quartiles, respectively.

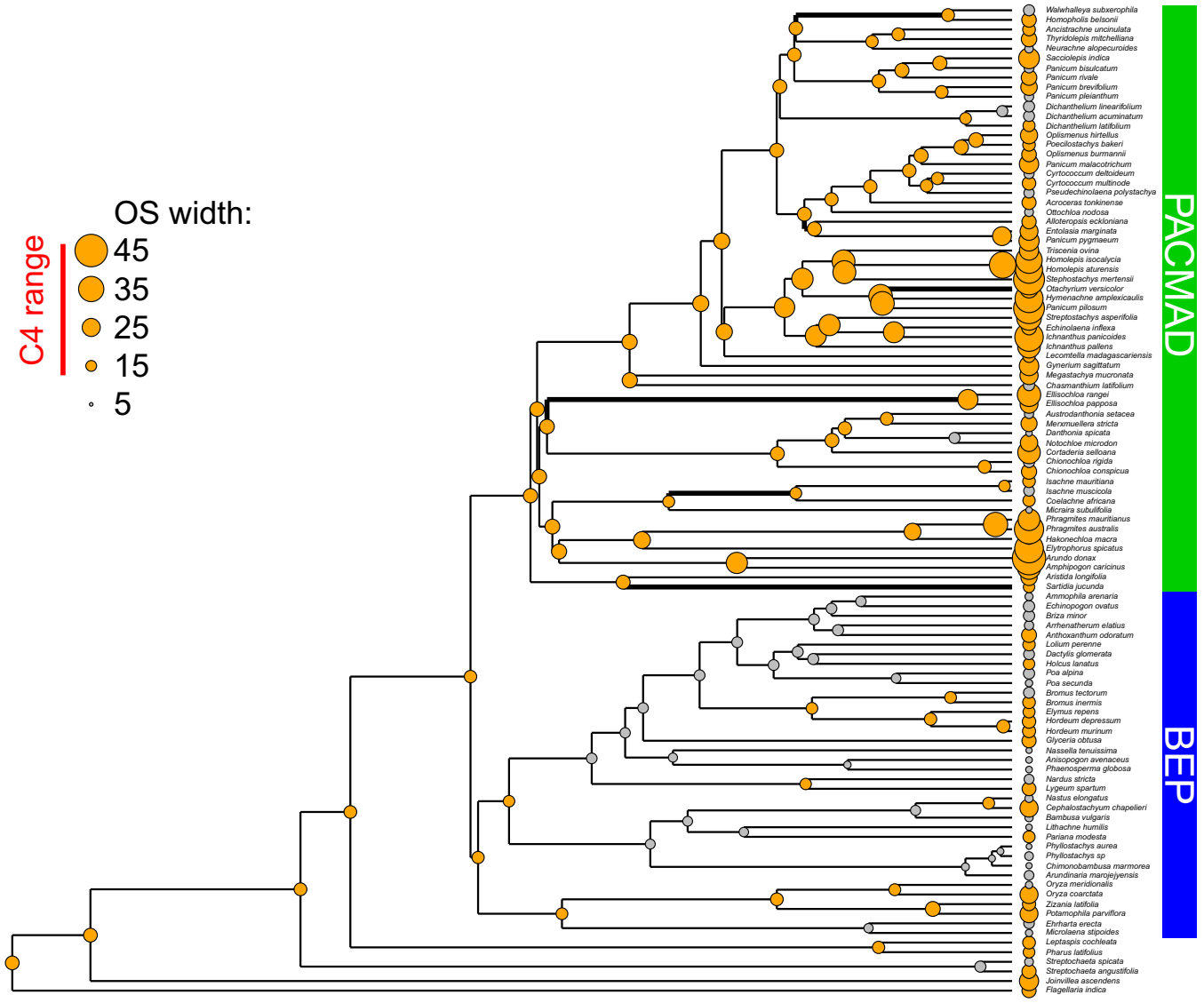

Fig. S3. Mapping of measured and reconstructed values of OS width on the phylogeny. The measured and inferred values of OS width are mapped on a calibrated phylogeny of the C_3 grasses included in this study. The diameter of the dots is proportional to OS width values. The values above the threshold that promotes C_4-OS evolution are in orange. Branches where the C_3 to C_4-OS transitions occurred are thicker.

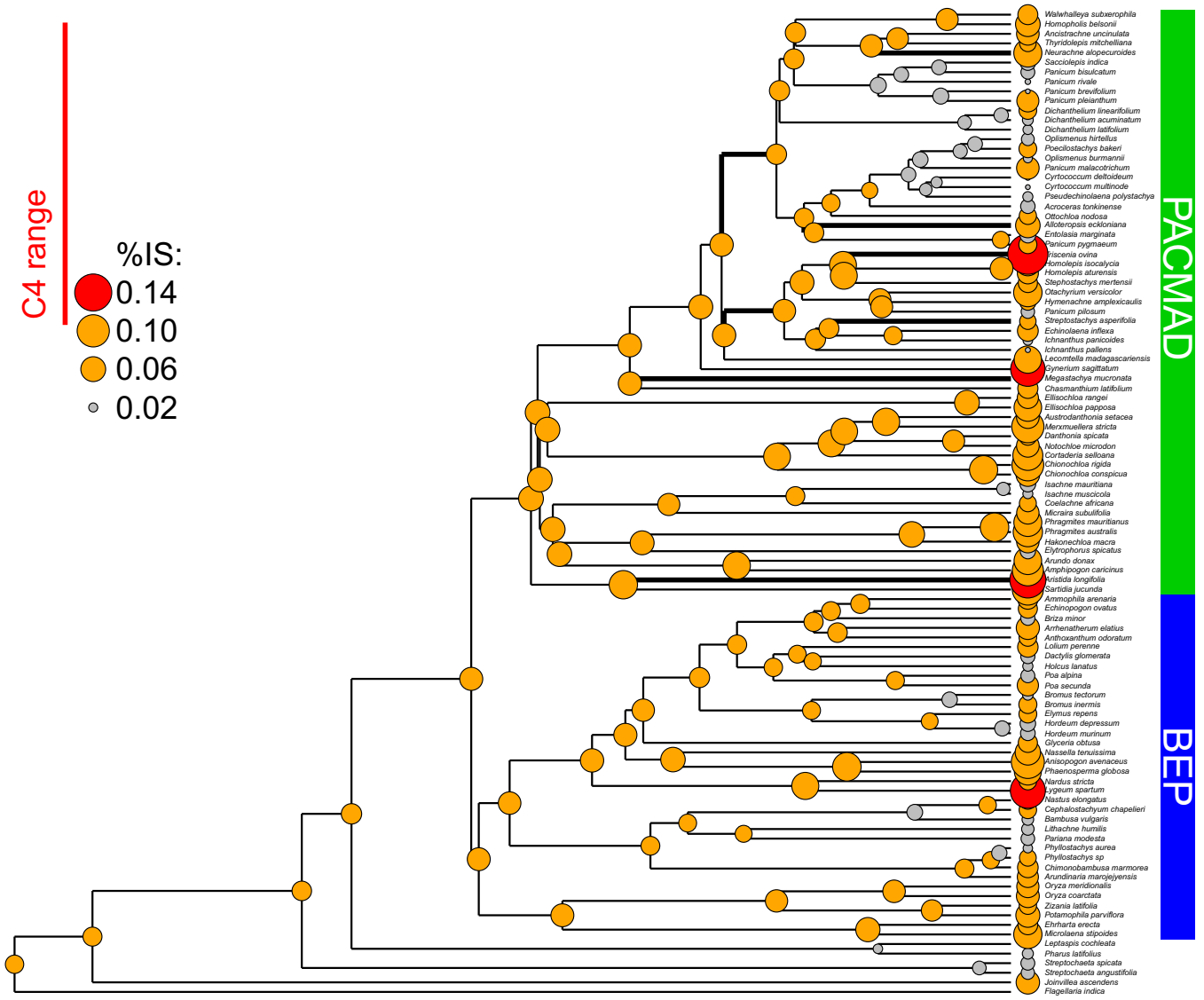

Fig. S4. Mapping of measured and reconstructed values of %IS on the phylogeny. The measured and inferred values of %IS are mapped on a calibrated phylogeny of the C_3 grasses included in this study. The diameter of the dots is proportional to %IS values. The values in the C_4 range are indicated in red, and those outside the C_4 range but above the threshold that promotes C_4-IS evolution are in orange. Branches where the C_3- to C_4-IS transitions occurred are thicker.

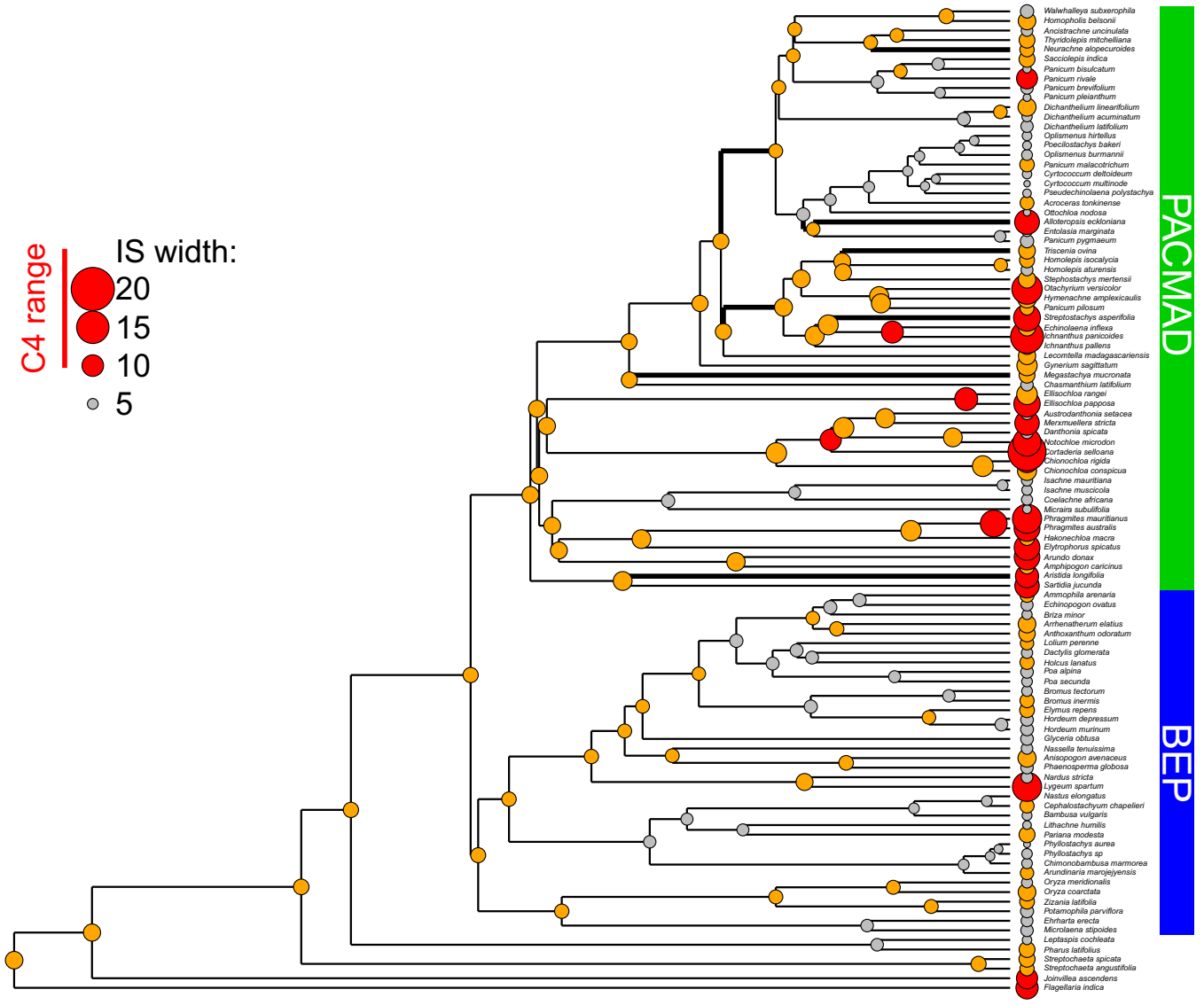

Fig. S5. Mapping of measured and reconstructed values of IS width on the phylogeny. The measured and inferred values of IS width are mapped on a calibrated phylogeny of the C_3 grasses included in this study. The diameter of the dots is proportional to IS width values. The values in the C_4 range are indicated in red, and those outside the C_4 range but above the threshold that promotes C_4-IS evolution are in orange. Branches where the C_3 to C_4-IS transitions occurred are thicker.

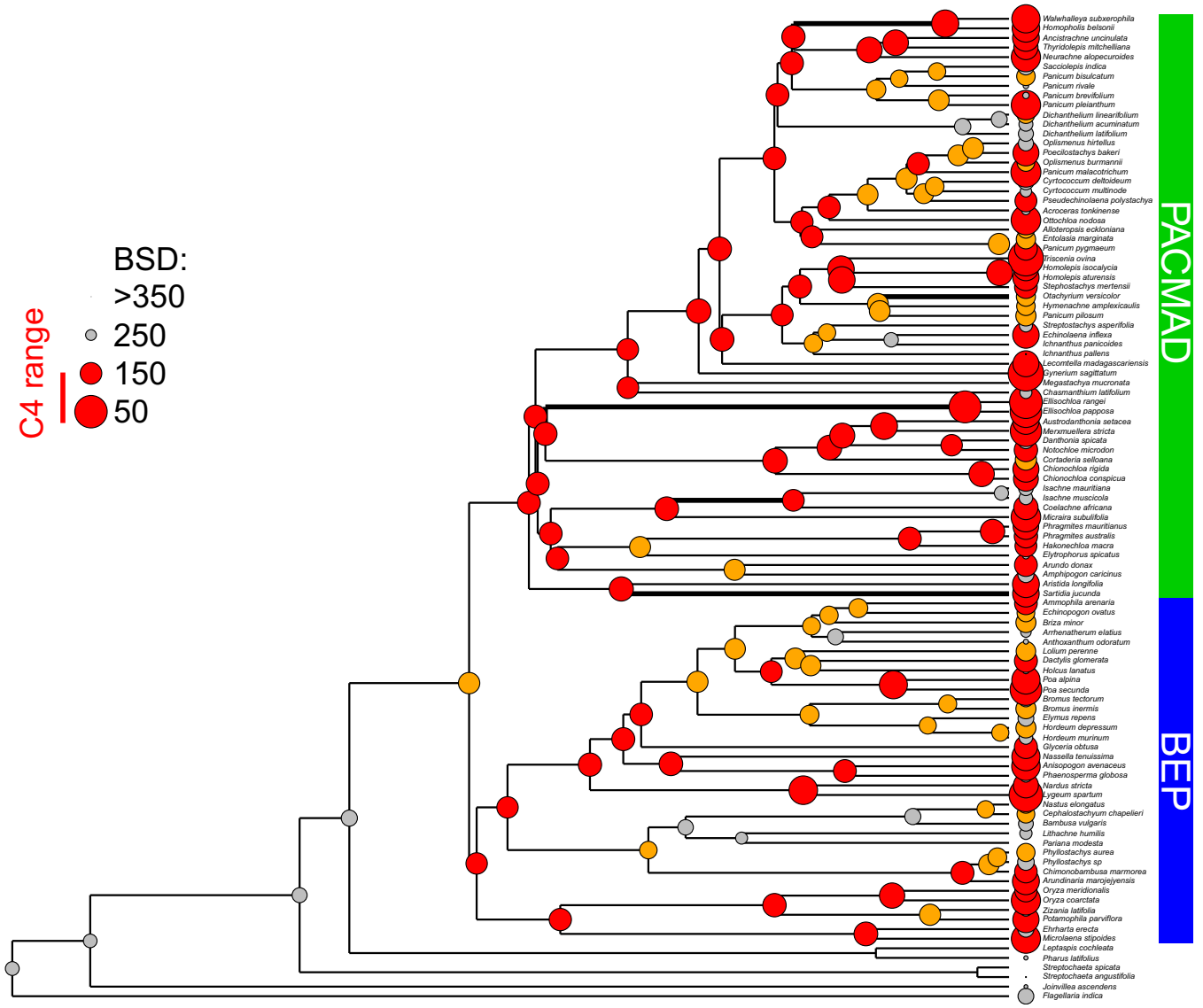

Fig. S6. Mapping of measured and reconstructed values of BSD on the phylogeny. The measured and inferred values of BSD are mapped on a calibrated phylogeny of the C_3 grasses included in this study. The diameter of the dots is inversely proportional to BSD values. The values in the C_4 range are indicated in red, and those outside the C_4 range but below the threshold that promotes C_4-OS evolution are in orange. Branches where the C_3- to C_4-OS transitions occurred are thicker.

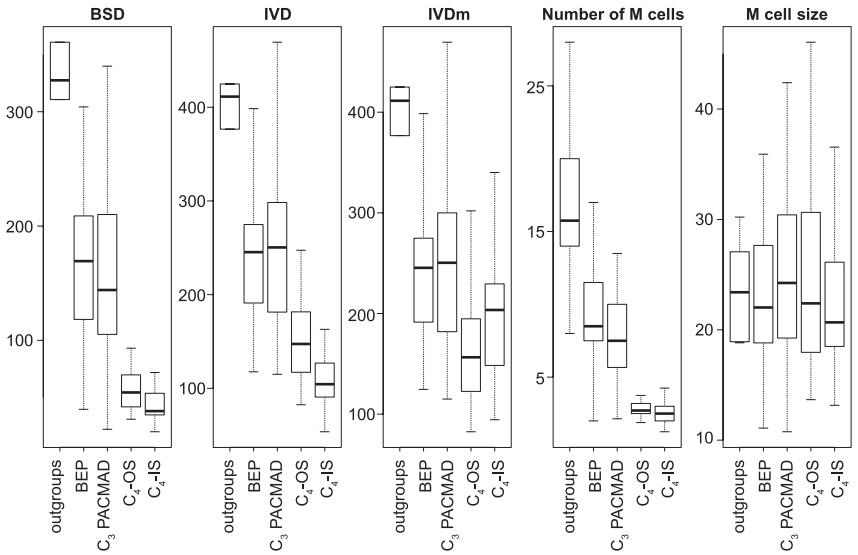

Fig. S7. Distribution of trait linked to BSD among clades and photosynthetic types. The distribution of five variables is summarized by boxplots. Boxes indicate the 25th and 75th percentiles and whiskers the extreme values still within 1.5 interquartile range of the upper and lower quartiles, respectively. BSD, inter-BS distance; IVDm, distance between major veins. Number of M cells is between two adjacent BSs.

Fig. S8. Detail of a portion of *Homolepis* leaf. This picture shows several vascular bundles of the C_3 *Homolepis aturensis*. The centripetal position of chloroplasts in the outer BS is usually characteristic of C_4 species.

Dataset S1. Sample details and anatomical measurements

Dataset S1

For each species, the measured and extrapolated values are indicated. Phylogeny indicates whether the species was included in the phylogeny. The voucher accession is indicated with the herbarium where the voucher is deposited in parentheses. SisterC4.OS: if the species belongs to a clade sister to a C_4-OS group, its value is 1. SisterC4.IS: if the species belongs to a clade sister to a C_4-IS group, its value is 1 (2 for species sister to a C_4-IS group and nested in a clade sister to another C_4-IS group). Duplicate: the value is 0 if there was only one sample per species, 1 if the sample was considered as the reference, and 2 if it was considered as the replicate. area.bundle, total of OS+IS+vein; area.total, total measured area (M+OS+IS+vein); B&T, seeds acquired from B and T World Seeds (Aigues Vives, France); measured_width, total width (connecting the veins) of the measured area; nb.M, mean number of M cells between two adjacent bundles; nb.veins, number of veins in the section used for measurements; nb.veins.major, number of major veins in the section used for measurements; USDA, seeds acquired from the US Department of Agriculture, Agricultural Research Service (Griffin, GA).