Leaf shape and size tracks habitat transitions across forest-grassland boundaries in the grass family (Poaceae)

Timothy J. Gallaher[1], Dean C. Adams[2], Lakshmi Attigala[3], Sean V. Burke[4], Joseph M. Craine[5], Melvin R. Duvall[6], Phillip C. Klahs[7], Emma Sherratt[8], William P. Wysocki[9], Lynn G. Clark[10]

[1]Biology, University of Washington, Box 351800, 98195-1800, Seattle, Washington, United States

[2]Ecology, Evolution, and Organismal Biology Department, Iowa State University, 253 Bessey Hall, 50011, Ames, Iowa, United States

[3]Plant Sciences Institute, Plant Sciences Institute, Ames, Iowa, United States

[4]Center for Data Intensive Sciences, University of Chicago, Chicago, Illinois, United States

[5]Jonah Ventures, Manhattan, Kansas, United States

[6]Department of Biological Sciences, Northern Illinois University, DeKalb, Illinois, United States

[7]Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, Iowa, United States

[8]Department of Genetics and Evolution, The University of Adelaide, Adelaide, South Australia, Australia

[9]Center for Data Intensive Sciences, University of Chicago, Chicago, Illinois, United States

[10]Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, Iowa, United States

E-mail: tjgallaher@gmail.com

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/evo.13722.

Abstract: Grass leaf shape is a strong indicator of their habitat with linear leaves predominating in open areas and ovate leaves distinguishing forest-associated grasses. This pattern among extant species suggests that ancestral shifts between forest and open habitats may have coincided with changes in leaf shape or size. We tested relationships between habitat, climate, photosynthetic pathway and leaf shape and size in a phylogenetic framework to evaluate drivers of leaf shape and size variation over the evolutionary history of the family. We also estimated the ancestral habitat of Poaceae and tested whether forest margins served as transitional zones for shifts between forests and grasslands. We found that grass leaf shape is converging towards different shape optima in the forest understory, forest margins and open habitats. Leaf size also varies with habitat. Grasses have smaller leaves in open and drier areas, and in areas with high solar irradiance. Direct transitions between linear and ovate leaves are rare as are direct shifts between forest and open habitats. The most likely ancestral habitat of the family was the forest understory and forest margins along with an intermediate leaf shape served as important transitional habitat and morphology respectively for subsequent shifts across forest-grassland biome boundaries.

Keywords: Anatomical evolution, ecotone, geometric morphometrics, grasslands, macroecology

Introduction

Early researchers had mixed views on the habitat in which the grass family first evolved. Bews (1929) suggested that the forest understory was the birthplace of the grasses based on the idea that bamboos and associated species, which are often found in shady habitats, possess primitive characters while Stebbins (1972) proposed that grasses first evolved in open habitats with subsequent shifts into forest. Clayton & Renvoize (1986) proposed a third hypothesis that grasses may have originated in forest margins.

Subsequent estimations of the ancestral habitat of the first grasses based on a modern phylogenetic understanding of grass evolution have focused on a two-habitat system (open habitat or closed-canopy forest) finding that the earliest grasses likely evolved in forests with major shifts to open habitats occurring approximately around the common ancestor of the BOP (Bambusoideae, Oryzoideae and Pooideae subfamilies) + PACMAD (Panicoideae, Aristidoideae, Chloridoideae, Micraioideae, Arundinoideae and Danthonioideae subfamilies) clades (Osborne and Freckleton 2009) or with multiple independent shifts in both BOP and PACMAD (Bouchenak-Khelladi et al. 2010).

Overlooked in recent biogeographical studies is the forest margin, which represents an ecologically important and globally widespread ecotone habitat (Smith et al. 1997). Forest margin ecotones are transitional zones where relatively abrupt changes in habitat occur over a narrow spatial extent. These habitats often support higher biodiversity as the community tends to be composed of a mix of species that occupy the more homogenous habitats on either side of the ecotone (Kark and van Rensburg 2006). Ecotones may also harbor unique biodiversity (ecotone species) and may serve as an important transitional zone for the evolution of lineages, potentially allowing a lineage to gradually evolve across the ecotone

into new niche space on the other side (Smith et al. 1997; Kark and van Rensburg 2006; Donoghue and Edwards 2014).

Ample evidence in grasses suggests that transitions from forest associated habitats to open habitats were accompanied by distinctive changes in leaf anatomy and morphology (Kellogg 2001; Cayssials & Rodríguez 2013). Forest grasses have significantly wider leaves relative to grassland species (Ryser and Wahl 2001; Cayssials and Rodríguez 2013). Leaf length and width were found to increase with precipitation levels and both measures were found to be greater in C_4 relative to C_3 species (Oyarzabal et al. 2008). However, the relative impact of habitat, climate and photosynthetic pathway on grass leaf shape and size evolution remains unclear. It also remains to be shown whether leaf size has evolved in response to a different set of variables than changes in leaf shape. Finally, it is unknown whether specific phenotypes are associated with forest ecotone habitats.

To address these outstanding questions, we investigated the evolution of leaf shape and size across the family with relation to precipitation, solar irradiance, temperature, photosynthetic pathway and forest association (hereafter referred to as habitat). We examined phylogenetic correlations among variables and tested for evidence of selection and convergent evolution of leaf shape and size in understory, forest margin and open habitats. Finally, we examined the ancestral habitat reconstruction of the grass family to infer whether forest margins have played an important role for evolutionary transitions across the forest-grassland boundary.

Materials and Methods

Phylogeny estimation and temporal calibration

Two phylogenies were generated for analyses: A complete-plastome-only phylogeny for purposes of estimating ancestral node ages and a larger phylogeny combining plastomes with available sequences of the NADH subunit F (*ndhF*) gene and the *trnL-trnF* intergenic spacer to produce a topology which included representatives from as many genera as available sequences would allow. In both cases only one representative per genus was used.

First, we constructed the plastome-only phylogeny using 261 complete or near-complete plastome sequences with *Joinvillea* set as the outgroup. Of these, 223 were acquired from GenBank and 38 were sequenced for this study (Supplementary Table S1 & S2). Library generation and sequencing methods followed those of Burke et al. (2016b). Sequences, with inverted repeat 'a' removed, were globally aligned with MAFFT v.7.310 (Katoh et al. 2002). Alignment columns with > 10% gaps were stripped resulting in an alignment of 104,716 base pairs in length of which 49,307 sites were variable. The best fit model of evolution estimated using jmodeltest v.2.1.6, was GTR+G+I and the sample size corrected Akaike information criterion (AICc) model weight for this model was 1.0. Maximum likelihood (ML) phylogeny estimation was carried out using RAxML v.8.2.9 (Stamatakis 2014). The best scoring tree estimated from 100 parsimony-based starting trees was annotated with the results from 1000 bootstrap replicates using the GTRCAT model. Bayesian phylogenetic inference and time-calibration was performed using BEAST v.2.4.2 (Bouckaert et al., 2014) using the GTR+G+I model with a chain length of 500,000,000 sampled every 5000 steps. *Joinvillea* was constrained as the outgroup. A lognormal relaxed clock model and seven fossil-based calibration points were used to calibrate the phylogeny. The fossils included isolated phytoliths assigned to the Chloridoideae and Bambusoideae

(Stromberg, 2005), silicified epidermis and phytoliths assigned to Oryzeae (Prasad et al. 2011) and to the root of the family (Wu et al. 2017), anthoecia assigned to an internal clade in the Stipeae tribe (Thomasson 1985) and inflorescence/spikelet compression fossils assigned to the genus *Leersia* (Walther, 1974; Akhmetiev et al., 2009), the bistigmatic clade (which includes the BOP and PACMAD clades plus their sister lineage the Puelioideae) (Crepet and Feldman 1991), and the root of the family (Poinar 2003, 2011) (Supplementary Table S3). A uniform prior was set for each clade to which the fossil was assigned with the minimum age set at the most recent time estimate of the associated fossil and a maximum age set to 200 Ma which is much older than the oldest described angiosperm fossils (Doyle et al. 2008; Iles et al. 2015). The root calibration used a gamma prior with the alpha and beta parameters each set to 4.

Three separate calibrations were run. The first used all seven fossils. The second and third were made under an assumption that grass silica short cell phytoliths, microscopic silica bodies that are formed in specialized epidermal cells in grasses and which take on distinctive shapes, cannot be reliably assigned to clades within the family but that they are diagnostic only at the family level. Calibration set 2 used the 98+ Ma cuticle and phytoliths recovered from hadrosaur teeth in China (Wu et al. 2017) and possible grass leaf fragments and spikelets recovered from Burmese amber (Poinar 2003) to calibrate the root node and calibration set 3 used the 65+ Ma cuticle and phytoliths recovered from dinosaur coprolites as the root calibration (Prasad et al. 2011).

Log files were evaluated with Tracer v.1.5 (Rambaut & Drummond 2009) to determine the burn-in percentage and to ensure that effective sample sizes > 200 had been reached for all estimated parameters. Following a burn-in of 25%, posterior probability (PP) support was annotated onto the maximum clade credibility tree.

The larger phylogeny included 581 taxa including *Flagellaria*, *Joinvillea* and *Ecdeiocolea* and 578 grass species representing all Poaceae tribes and subtribes and approximately 75% of all grass genera. An uncalibrated Bayesian phylogeny estimate was done using the same strategy as with the plastome-only analysis. *Flagellaria* was set as the outgroup taxon based on the well supported relationships between families in the Poales found in recent phylogenetic studies (Bouchenak-Khelladi et al. 2014; Barrett et al. 2016). The 95% confidence intervals (CI's) of age estimates for 33 nodes from the plastome-only phylogeny were assigned to the corresponding nodes in the 581-taxon phylogeny and the larger tree was calibrated using relaxed penalized likelihood implemented by the 'chronos' function in R package ape v5.2 (Paradis et al. 2004). The larger phylogeny was used for all subsequent analyses with tips dropped as necessary to accommodate each dataset. All phylogenetic analyses were done on the CIPRES Science Gateway (Miller et al. 2010). Trees and alignments were deposited in TreeBASE (www.treebase.org).

Climate and habitat

Occurrence data for all grass species was obtained from the Global Biodiversity Information Facility (GBIF; www.gbif.org). We excluded duplicate records, those not represented by a voucher and those with other geo-referencing issues. Spatial climate data (Bioclim) represented by 35 variables in four categories, precipitation (mm), temperature (°C), solar irradiance (W m-2) and soil moisture availability index [0 (dry) – 1 (wet)], which is calculated from weekly rainfall and pan evaporation data, was downloaded from www.climond.org (Hijmans et al. 2005; Kriticos et al. 2014). For each of the 35 climate variables, means were calculated from all filtered occurrence records for each species and species averages were used to calculate genus average.

Mean annual precipitation (MAP) is associated with and often used as a proxy for the relative cover of forest or grasslands. Closed canopy forests are much more likely when MAP exceeds 2000 mm while grasslands typically dominate when MAP falls below 1000 mm (Staver et al. 2011). Between these two values either grasslands or forest can exist as one of two alternate states. Ancestral MAP was compared with the discretely coded ancestral estimation of habitat.

The habitat for each genus was classified as forest understory (full/obligate-shade in tropical or temperate forests), forest margins or gaps, and/or open (full sun grasslands, savanna, open woodlands and most wetland taxa) based on published information and personal observations by the authors (See Supplementary Table S5). Genera were coded as occurring in a habitat type if one or more species was reported to occur in that habitat. For genera associated with more than one habitat type, equal probabilities were used for each habitat.

Leaf morphometrics

Leaf sampling was carried out to maximize taxonomic breadth and leaves were selected from available material at the ISC herbarium supplemented with leaf images from online digitized specimens. Whole mature leaves were removed from herbarium specimens and mounted with clear tape to a new herbarium sheet for digitizing. We did not use flag leaves and avoided other leaves close to an infloresence. In many cases, the leaves were re-wetted with Pohl's solution (Pohl 1965) to unfold and straighten them. Mounted leaves were imaged, and the outline of individual leaves digitized. Altogether 792 leaves were prepared in this manner and an additional 177 leaves were digitized directly from herbarium specimen images. A total of 969 leaf outlines were digitized representing 206 species (Supplementary Table S6).

Leaf blade outlines (auricles, ligule and leaf sheaths excluded) were characterized by landmark-based geometric morphometrics. Digital images were made binary and the outlines extracted as XY coordinates using *ImageJ* v1.51 (Rasband 2012). A zip file of leaf images used in this study is available from the first author. Fixed landmarks were set, one at the center of the base of the blade and the other at the leaf apex and 100 evenly spaced sliding semi-landmarks were set around the margins to characterize the two-dimensional shape. Further leaf digitization and geometric morphometrics were performed in R 3.4.0 (R Core Team 2017) using *geomorph* v3.0.3 (Adams and Otárola-Castillo 2013). Procrustes alignment was performed on the set of leaves for each species allowing semi-landmarks to slide along their tangent direction while minimizing Procrustes distance. Species averages were calculated, and a final Procrustes alignment was performed on the set of average outlines. A principal component analysis (PCA) was done on the aligned coordinates to visualize the shape variation in a low dimensional space. Measures of leaf size (length, width and area) were taken for each leaf image and the mean and standard error calculated. In addition, leaf width and length data from the Grassbase database (Clayton et al. 2016) was taken for each species and the species and genus averages were computed. Width, length and area were log-transformed prior to statistical analysis. In addition to the leaf outline, the width-length (W:L) ratio was calculated as a further proxy for leaf shape for comparability with other studies and for use in analyses requiring one-dimensional data. Leaf shape was also coded as discrete data. A leaf width:length ratio < 0.09 was coded as linear, between 0.09 and 0.15 as lanceolate and > 0.15 as ovate. Photosynthesis type (C_3 or C_4) and subtype (C_3, NADP-ME, NAD-ME or PCK) was recorded from the literature (Supplementary Table S5).

Statistical analyses

Analyses were done in the R statistical environment v.3.4.0. The log transformed means and standard errors of leaf length, width, area and the W:L ratio were tested for phylogenetic signal by computing Pagel's Lambda, and Blomberg's Kappa with 1000 simulations using the R package phytools v.0.6 (Revell 2012). Phylogenetic signal for the multivariate leaf outline data was estimated using a generalization of the K-statistic for multidimensional variables (Adams 2014a) implemented in geomorph.

Ancestral state estimations (ASE) of discrete characters including habitat and photosynthesis type were performed using the 'raydisc' function in the R package *corHMM* (Beaulieu et al. 2017), which allows for each terminal taxon to be coded with multiple states. We also used the 'ace' function in *ape* (Paradis et al. 2004) and 'fitdiscrete' in *geiger* (Pennell et al. 2014). The equal rates (ER), symmetric rates between pairs (SYM) and all rates different (ARD) models were evaluated for ASE of discrete. The log likelihood of nested models was compared using a chi-squared test. The best model was used to simulate 1000 ancestral state histories using 'simmap' in *phytools* (Revell 2018). ML methods implemented in the *ape* and *phytools* R packages were used to estimate ancestral values for continuous traits.

Correlations between habitat, photosynthetic type, and 35 climate variables were assessed as explanatory variables for leaf shape, and size using several methods. Phylogenetic generalized least squares (PGLS) analysis was done in the R package *caper* v.0.5.2 (Orme, 2013) for models utilizing one-dimensional variables and in *geomorph* with 1000 simulations to assess statistical significance when high-dimensional shape data was used as the response variable (Adams, 2014b). Environmental variables were classified into six categories: habitat (3 models), photosynthetic type, temperature (11 variables), precipitation (8), solar irradiance

(8) and moisture index (8). When more than one variable or model per category was significant after Bonferroni correction, the predictor with the highest sample-size-corrected Akaike information criterion (AICc) score was selected for multivariate analysis. Adjusted R^2 values and AICc were used to compare models that differ in the numbers of predictor variables. Phylogenetic *post-hoc* tests for differences in means between groups were performed using the 'phylANOVA' function in *phytools* with 1000 simulations (Revell 2012). Three habitat models were evaluated including the three-category model described above (understory, margins, open) and two two-category models: (2a) with open habitat and forest margin samples combined and (2b) with understory and forest margin samples combined.

We also used phylogenetically independent contrasts (PIC) implemented in the R package *ape* (Paradis et al. 2004) and the threshold model of character covariance implemented in *phytools* (Revell 2012) to test for significant correlations among the leaf shape and size characters and between those characters and each habitat pair. For the threshold method, each model was run for 10,000,000 generations sampled every 1000 steps with a 20% burnin.

To assess convergent evolution for leaf shape (outline) we adopted the approach used by Stayton (2015) and implemented in the R package *convevol* v1.0, which estimates a value (C_1) for the strength of convergent evolution by calculating the quotient of the mean distance between potentially convergent species over the maximal distance of their shared ancestors in morphospace.

Finally, to evaluate evidence of selection on leaf shape and size in different habitats, we evaluated the relative support for Brownian motion (BM) and Ornstein-Uhlenbeck (OU) models of evolution and tested for significant differences in estimated value optima (θ),

strength of selection (α) and rates of evolution (σ^2) using the R package *OUwie* v1.5 (Beaulieu and O'Meara 2016). To set adaptive regimes on the phylogeny, the ancestral habitat state with the highest probability at each node from the three-category model for 581 taxa was calculated following 1000 simulations. The tree was reduced to include only the 206 taxa for which measured data was available. For each trait, we used the species mean along with the standard error to account for intraspecific variation. We dropped the estimate of the root parameter from all except the BMS model and this resulted in more stable parameter estimates. The analysis of shape used the W:L ratio since the associated methods are not yet well developed for the multidimensional data that describes the overall leaf shape. Five models were assessed for each character including two single optima models: BM which assumes $\alpha = 0$ and OU1 which assumes $\alpha > 0$, and three models that allow for different parameter regime for different groups of taxa: BMS (multiple rate BM), OUM (different θ's between groups, α and σ^2 held constant across groups), and OUMV (different θ's and σ^2 between groups, α held constant across groups). Each analysis was assessed to ensure that all eigenvalues were positive and that all estimated paraameters were reasonable. Following parameter estimation with 500 parametric bootstrap replicates, the mean and 95% CI of each estimated parameter were calculated as a weighted averaged from all models with an AICc weight > 1%. The analysis was repeated using a phylogenetic tree scaled to a root height =1 in order to produce an estimate of α comparable across studies and we calculated -log(α) (Ives and Garland Jr 2009; Cooper et al. 2016) to aid in interpreting the magnitude of α.

For all statistical hypothesis tests, the null hypothesis was rejected when $P < 0.05$ for the entire model. For sequential exploratory tests of multiple predictors for each dependent variable, the Bonferroni method was used to correct for the high number of statistical tests. In multiple regression, the model was considered a better fit to the data if the overall model was

significant, each variable remained significant and the AICc indicated a better fit than alternative models.

RESULTS

Phylogeny estimation and temporal calibration

The plastome-based phylogeny estimate was well supported with only minor incongruences between the ML and Bayesian inference (BI) trees (Supplementary Fig. S1 and S2). Subfamilial and tribal relationships agree with other recent plastome-based studies (Cotton et al. 2015; Wysocki et al. 2015; Burke et al. 2016a; Teisher et al. 2017; Saarela et al. 2018). Support < 0.99 posterior probability and < 99% bootstrap was found for 20 and 60 nodes respectively including relatively low support for the node supporting Panicoideae as the sister group to the remaining subfamilies of the PACMAD clade. (see Saarela et al. (2018) for a discussion of the low support values observed at this node.) CI's for the node age estimates were wide (Supplementary Fig. S1, Supplementary Table S4). Calibration set 2 and 3 provided younger ages than set 1. Under set 1 and 2, Poaceae split from its sister lineage in the Early Cretaceous (131.39 [110.56-155.94] and 117.18 [102.58-137.89] respectively) while the BOP and PACMAD (together referred to as the core grasses) began to diversify in the Early-Late Cretaceous (97.97 [81.84-118.42] and 83.26 [67.71-99.48]). Set 3 produced the youngest ages with an estimated origin in the Early-late late Cretaceous (99.1 [76.73-120.85]) with core grass diversification in the Paleocene to Late Cretaceous (71.45 [56.5-88.42]) (Supplementary Fig. S1, Supplementary Table S4). The 581-taxon phylogeny (Supplementary Fig. S3) also had a well-supported backbone and clades were well supported to the tribal level although many of the relationships among genera within tribes did not have strong posterior probability support.

Habitat, climate and photosynthesis pathway

The ML estimate found the forest understory as the most probable ancestral habitat for the common ancestors of Poaceae, the spikelet clade, Anomochlooideae, Pharoideae and Puelioideae. In each case the forest margin was the second most probable state (Table 1, Fig. 1a). The highest probability for the habitat of the ancestor of the bistigmatic clade, core grasses, BOP clade, Oryzoideae, and Bambusoideae were forest margins while the crown node of the Pooideae was equivocal between open habitat and forest margins. We estimated between 12 and 41 transitions to open habitat over the phylogeny and all of these from forest margins. This result underestimates the true number of transitions given the incomplete phylogeny used in the analyses. Of these, the earliest and most significant transitions to open habitat occurred within the Pooideae, at the crown of the PACMAD clade and in the Oryzeae. Subsequent shifts from open habitats to forest associated habitats occurred in several subfamilies but most notably in the Panicoideae. The ARD model for habitat transitions had the highest likelihood. The confidence interval for the transition rate and estimated number of transitions from open habitat to the understory overlapped with zero as did the rate from understory to open habitat (Table 2). The highest transition rates, with positive 95% CI and positive range estimates for the number of transitions, were from open habitats to forest margins and from forest margins to the understory. Intermediate transition rates, with a CI overlapping with 0 (although the number of transitions did not overlap with 0), were found from from the understory to forest margins and from forest margins to open habitat.

The inferred MAP for the origin of the family was 1413-1881 mm per year (Table 1, Fig.2b). Shifts into habitats with rainfall below 1000 mm occurred along the backbone of both ACMAD (the clade consisting of Aristidoideae, Chloridoideae, Arundinoideae, Micrairoideae, Arundinoideae and Danthonioideae) and Pooideae in the Paleocene-Eocene

and within the Paniceae (Panicoideae). The Anomochlooideae, Oryzoideae, Bambusoideae and Panicoideae were inferred to have evolved in habitats with precipitation between 1300 and 1800 mm while the Pharoideae and Puelioideae evolved in habitats with precipitation over 1800 mm. There was a significant relationship between MAP and habitat using PGLS (R^2=0.18, P<<0.0001).

One thousand simulations of the ancestral state estimation of photosynthesis type inferred on average 18.6 shifts from C_3 to C_4 photosynthesis with no reverse shifts (Supplementary Table S11).

Leaf shape and size

Leaf shape.—The first principal component (PC1) of 206 aligned composite leaf outlines explained over 95% of the shape variance and comparison of the warp grids at the extreme values along this axis indicates that much of the variation in PC1 is related to leaf linearity (Fig. 3a,b). There were clear phylogenetic patterns and significant phylogenetic signal in leaf outline (Fig. 1b, Fig. 4, Table 4). Shifts to more linear leaves occurred in Oryzeae, Pooideae, and PACMAD clades. Estimates of the transition rates between discretely coded leaf shapes found a rate of zero for direct transitions between linear and ovate leaves in either direction (Table 3). All other transition rates were positive.

Habitat was strongly and significantly related to leaf shape (Table 4,5, Fig. 5, Supplementary Table S7). AICc values indicate that the regression model with three habitat categories was a better fit for the data than either of the two category habitat models (Supplementary Table S7). Grass leaf shapes (as a complete outline and W:L ratio) are significantly different between open habitat, forest margins and forest understory species. Under the threshold model there were significant correlations between the W:L ratio and each pair of habitats with mean R^2 values of 0.51 between open habitat and forest margins and

0.46 between forest margins and forest understory (Supplementary Table S8). There were no significant differences in shape between C_3 and C_4 taxa after Bonferroni correction (Table 4,5, compare Fig.1b and 2a, Fig. 5). Leaf length and width were significantly associated with overall shape and the leaf shape outline was strongly correlated with the W:L ratio. Leaf shape was not significantly correlated with leaf area (Table 5, Supplementary Table S9). The W:L ratio was weakly correlated with several precipitation and moisture availability variables; however, these were not significant after Bonferroni correction and R^2 values were low (Supplementary Table S7).

There was significant convergent evolution towards linear leaves in open habitats, and lanceolate leaves in forest margins. Leaves of extant open habitat species are 51.6% (C_1=0.516, P<0.001) more similar than the maximal morphological differences among their ancestors. This strong pattern of evolutionary convergence is seen visually in the principal components biplot, where species found in open habitats occupy a small, and distinct, region of morphospace (Fig. 3a). Forest margin species show an average of 34% convergence (P<0.001), a pattern also confirmed visually in morphospace (Fig. 3a). By contrast, morphological convergence in understory habitats was not statistically significant. The best evolutionary model for the W:L ratio was a multiple regime Ornstein-Uhlenbeck-based Hansen model (Supplementary Table S10). The W:L ratio is evolving towards separate optima in all three habitat classes with non-overlapping 95% CI's (Table 6). The scaled -log(α) was −2.27 indicating a moderate to high strength of pull towards the optima.

Leaf size.— There was no discernable phylogenetic pattern in leaf length variation (Fig. S5). Width and area generally decreased along the backbone of the phylogeny leading to the Oryzoideae, Pooideae, and PACMAD clades and increased at the crown nodes of the Anomochlooideae, Pharoideae, Bambusoideae and Panicoideae (Fig. S5).

Leaf length and width were positively correlated with each other (Supplementary Table S9, Fig. S4). There was no significant difference in leaf length, width or area between species utilizing the C_3 or C_4 pathways or among C_4 subtypes (Table 4, Fig. 5, Supplementary Table S7). Estimated leaf length optima overlapped among habitat classes and the overall -log α was −1.52 indicating a moderate level of pull towards the optima. Length, width and area were significantly and positively associated with higher moisture and precipitation and negatively associated with levels of solar irradiance (Fig. 6, Supplementary Table S7). Measures of leaf size were not significantly related to air temperature (Supplementary Table S7).

A larger width and area were associated with forest habitats whereas length was not significantly associated with habitat (Supplementary Table S7, S8, Fig. 5). In multiple regression with habitat, mean annual moisture index and highest weekly radiation each remained a significant predictor of leaf width and area (Table 5). Habitat with highest weekly solar radiation resulted in the best predictive model and explained approximately 25% and 32% of the variation in width and area respectively after phylogenetic correction (Table 5, Fig. 6). Leaf width and area are evolving towards separate optima in different habitats although the 95% CI's for width overlap between open areas and forest margins and the intervals for area overlap between forest margin and understory species (Table 6, Supplementary Table S10). The overall estimate -log α values were −3.23 for leaf width and −5.27 for leaf area.

DISCUSSION

Drivers of leaf shape and size evolution in the grasses

Leaf shape in grasses is strongly associated with habitat and this relationship holds when accounting for phylogenetic non-independence of samples. Across the family, grass leaves have converged towards optimal shapes in each habitat with linear leaves in open habitats, lanceolate leaves in forest margins and ovate leaves in the forest understory. Leaf shape variation in the grasses is not strongly related to other variables that we assessed such as temperature, precipitation, solar radiation levels or photosynthetic pathway. Grass leaf shape evolution shows a strikingly similar pattern to the estimated evolutionary history of habitat both in terms of inferred ancestral states and in the estimated transition rates between states (compare Fig. 1a &1b). In each case, direct transitions between extremes (shape or habitats) occur at zero or near zero rates (Table 2,3). This shared pattern suggests that leaf shape has tracked habitat shifts throughout the history of the family. Shifts between linear leaves and ovate leaves or between forest understory and open habitat are more likely to take place through an intermediate (lanceolate) shape and through intermediate habitat such as the forest margin.

Variation in leaf size, particularly width and area, follows a similar pattern to shape with wider leaves with greater surface area in forested habitat and narrower leaves with smaller surface areas in open habitat (Table 4). However, leaf length variation does not vary with habitat type, therefore changes in leaf shape in the grasses are primarily due to habitat-shift associated changes in leaf width absent a corresponding change in leaf length. One possible explanation for this relationship is that an increased leaf width in shadier habitats will allow for a larger area for light interception in an otherwise light limiting environment without expending additional resources on producing non-photosynthetic tissues. Adding

length or many more linear leaves requires the production of additional non-photosynthetic vascular tissue while additional width can be gained by increasing the photosynthetic region between veins in few leaves without additional investment in non-photosynthetic tissue. In contrast, open habitats favor linear leaves. Linear leaves have smaller boundary layers which allows for a higher rate of convective heat loss to the environment preventing high internal leaf temperatures which can lead to damage to enzymes and tissues, reducing photosynthetic efficiency (Givnish 1979). Leaves in open habitats are also likely to be subject to higher levels of mechanical stress due to wind and a linear shape may allow a leaf to twist and bend without damaging tissues (Niklas 1999). Linear leaves can minimize self-shading and allows for the production of a greater number of leaves per plant potentially limiting the per-leaf investment and risk. Intermediate leaves, generally lanceolate in shape, in the forest margins may reflect the partially shaded but still somewhat exposed conditions within the ecotone.

In addition to the effect of habitat, grass leaf size is also related to a wider range of environmental variables than leaf shape (see Table S7). Concerted increases in leaf width and length, without significant changes in overall shape, occur in response to greater moisture availability and lower levels of solar irradiance in all habitat types. However, in contrast with studies from other taxonomic groups, air temperature does not appear to be an important determinant of leaf shape or size for grasses (Yates et al. 2010; Chitwood et al. 2012; Wright et al. 2017).

Habitat shifts in the grass family

This study suggests that forest margin ecotone habitats may have played an important role in the diversification history of the grasses, partially consistent with the hypothesis of Clayton & Renvoize (1986). Although our ASE favors the understory for the common ancestor of the grasses, a forest ecotone habitat cannot be ruled out and the forest margin

habitat increases in probability along much of the subsequent backbone of the phylogeny leading to the core grasses (Fig. 1). These results are generally in agreement with previous analyses based on a two state (closed or open) reconstruction (Osborne and Freckleton 2009; Bouchenak-Khelladi et al. 2010) but they highlight the important role for the forest margin and suggest that this ecotone may have been a required transitional habitat for shifts between the forest understory and open habitats.

Evidence, mainly from studies of animals, indicates that ecotone habitats such as forest gaps and edges can promote rapid evolution, the evolution of genetic novelty and the generation of new species (Moritz et al. 2000; Schilthuizen 2000; Schneider et al. 2014). The forest margin ecotone is physically positioned between forest understory and open habitats allowing for more regular dispersal into those habitats and the ecotone has similar climate and intermediate levels of sun and exposure allowing a space for lineages to adapt gradually to new habitat (Donoghue and Edwards 2014).

The early major independent shifts from the forest margins to open habitat occurred in the PACMAD and Pooideae (Fig 1a) and under calibration set 1 (Table S4) these occured in the Cretaceous or Paleocene long before the first potential evidence for grass-dominated ecosystems in the Eocene or the eventual dominance of savannas and grasslands beginning in the Miocene (Jacobs et al. 1999; Edwards et al. 2010; Strömberg 2011). The alternate calibration scenarios each produce younger estimates but only calibration set 3 is consistent with a scenario in which the evolution of grasslands closely followed or coincided with the shift of grasses away from forest-associated habitats.

The shifts to open habitat correspond with reductions in MAP below approximately 1500 mm (Table 1) and MAP estimates quickly drop below 1000 in the Pooideae and the ACMAD lineage but increase to more mesic levels in the Panicoideae. If the ancestors of the

Panicoideae evolved in wetter conditions, this may in part explain the generally longer and wider leaves inferred for ancestral members of that subfamily relative to many in the sister ACMAD lineage. In addition, examination of the ancestral values of the width:length ratio in ancestral Panicoideae shows a still linear but somewhat more intermediate shape than at the base of ACMAD or Pooideae. The larger and slightly more lanceolate ancestral leaves in the Panicoideae may have facilitated later shifts back into forest margins and understory habitats contributing to the habitat breadth and species diversity that characterize the Panicoideae today. A similar situation may have occurred in the Bambusoideae, which retained an ancestral intermediate leaf shape, possibly allowing for subsequent shifts back into the forest understory (Olyreae, some Bambuseae) or into open habitats (some Bambuseae and Arundinarieae).

Leaf shape, size and photosynthetic pathway

Although there was no significant difference in leaf size or shape among photosynthetic pathways in contrast with previous findings (Oyarzabal et al. 2008), there may still be non linear relationships between leaf morphology and photosynthetic pathway. For example, although the width and shape of leaves of C_4 species overlap with C_3 species, the leaves of C_4 taxa exhibit a narrow range of variation towards the linear and narrow-width range of C_3 species (see Fig. 5). It has been shown that increased vein density is a necessary precondition for C_4 photosynthesis (Christin et al. 2013) and evolving a more linear shape or a narrower width is likely to result in increased vein densities. Linear leaves, and by extension the open habitats in which linear leaves evolved, may therefore represent a necessary precondition for evolving C_4 photosynthesis in grasses. To test this hypothesis, additional work is needed to examine allometric relationships between leaf shape, size and

vascular architecture to evaluate indirect relationships between leaf shape and size and photosynthetic pathway.

Edwards and Smith (2010) found that 18 of 20 transitions from C_3 to C_4 photosynthesis were accompanied by reductions in MAP, suggestive of shifts from closed forests to open woodlands or savanna. Our MAP estimates, albeit with lower species-level sampling density, shows a similar overall pattern for precipitation, but the habitat ASE in our study indicates that grasses shifted into open habitats long before the evolution of C_4 photosynthesis. Each of the 18-24 shifts from C_3 to C_4 photosynthesis in our dataset took place in open habitats. These results suggest that climatic shifts towards more arid conditions, independent of habitat shifts from forest to open habitat, were the proximate drivers of the evolution of C_4 photosynthesis in some grass lineages although nitrogen limitation in wetter habitats or other factors may have also played an important role for promoting C_4 photosynthesis in some lineages (Liu and Osborne 2015; Rao and Dixon 2016).

The C_4 lineages in the Chloridoideae and Paniceae likely evolved under arid conditions whereas those of the Micrairoideae and most other C_4 lineages in the Panicoideae likely evolved in more mesic to wet conditions. Since these different climate conditions may have had an important impact on leaf size it will be interesting to assess whether different ancestral leaf sizes played any role in the subsequent evolution of C_4 anatomy.

Phylogenetic calibration

Phylogenetic calibration set 1 produced an older age estimate than analyses which assign fossil phytoliths only to the root node of the family. These results are similar to studies which used the 65+ Ma putative Orzyeae cuticle and phytoliths (Vicentini et al. 2008; Bouchenak-Khelladi et al. 2010; Prasad et al. 2011; Christin et al. 2014; Jones et al. 2014). The late Cretaceous grass cuticle and phytoliths assigned to Oryzeae have been controversial

due to their placement in a relatively deeply nested clade results in a phylogeny much older than previous estimates (Prasad et al. 2011; Christin et al. 2014; Kellogg 2015). However, there is growing evidence to suggest that the placement may be accurate. The transversely oriented "scooped bilobate" phytoliths surrounded by a thickened silicified cuticle with a regular pattern of triangular costal papillae which make up the rails of the "ladder-like structure" described by Yamanaka et al. (2009) and the shape of the cells between the costal phytoliths are all diagnostic of members of the Oryzeae and this combination of characters is not found elsewhere within the family (First author, personal observations). Additional fossil discoveries including possible spikelets and leaves in Burmese amber (Poinar 2003, 2011; Poinar et al. 2015) and grass phytoliths and cuticle of unknown intrafamiliar placement from China dating to the Albian (100-113 Ma) (calibration set 2) lend additional support for a much older age of the family.

Literature Cited

Adams, D. C. 2014a. A generalized K statistic for estimating phylogenetic signal from shape and other high-dimensional multivariate data. Syst. Biol. 63:685–697.

Adams, D. C. 2014b. A method for assessing phylogenetic least squares models for shape and other high-dimensional multivariate data. Evolution (N. Y). 68:2675–2688.

Adams, D. C., and E. Otárola-Castillo. 2013. Geomorph: an R Package for the collection and analysis of geometric morphometric shape data. Methods Ecol. Evol. 4:393–399.

Akhmetiev, M., H. Walther, and Z. Kvacek. 2009. Mid-latitude Paleogene floras of Eurasia bound to volcanic settings and palaeoclimatic events - experience obtained from the east of Russia (Sikhote-Alin') and Central Europe (Bohemian Massif). Acta Musei Natl. Pragae, Ser. B - Hist. Nat. 65:61–129.

Barrett, C. F., W. J. Baker, J. R. Comer, J. G. Conran, S. C. Lahmeyer, J. H. Leebens-Mack, J. Li, G. S. Lim, D. R. Mayfield-Jones, L. Perez, J. Medina, J. C. Pires, C. Santos, D. W.

Stevenson, W. B. Zomlefer, and J. I. Davis. 2016. Plastid genomes reveal support for deep phylogenetic relationships and extensive rate variation among palms and other commelinid monocots. New Phytol. 209:855–870.

Beaulieu, J. M., and B. C. O'Meara. 2016. OUwie: analysis of evolutionary rates in an OU framework.

Bews, J. W. 1929. The World's Grasses. Longmans, Green and Co., London.

Bouchenak-Khelladi, Y., A. M. Muasya, and H. P. Linder. 2014. A revised evolutionary history of Poales: Origins and diversification. Bot. J. Linn. Soc. 175:4–16.

Bouchenak-Khelladi, Y., G. A. Verboom, V. Savolainen, and T. R. Hodkinson. 2010. Biogeography of the grasses (Poaceae): a phylogenetic approach to reveal evolutionary history in geographical space and geological time. Bot. J. Linn. Soc. 162:543–557.

Bouckaert, R., J. Heled, D. Kuhnert, T. Vaughan, C. Wu, D. Xie, M. A. Suchard, A. Rambaut, and A. J. Drummond. 2014. BEAST 2 : A software platform for Bayesian evolutionary analysis. PLOS Comput. Biol. 10:1–6.

Burke, S. V., C.-S. Lin, W. P. Wysocki, L. G. Clark, and M. R. Duvall. 2016a. Phylogenomics and plastome evolution of tropical forest grasses (Leptaspis, Streptochaeta: Poaceae). Front. Plant Sci. 7:1–12.

Burke, S. V., W. P. Wysocki, F. O. Zuloaga, J. M. Craine, J. C. Pires, P. P. Edger, D. Mayfield-Jones, L. G. Clark, S. A. Kelchner, and M. R. Duvall. 2016b. Evolutionary relationships in Panicoid grasses based on plastome phylogenomics (Panicoideae ; Poaceae). BMC Plant Biol. 16:1–11.

Cayssials, V., and C. Rodríguez. 2013. Functional traits of grasses growing in open and shaded habitats. Evol. Ecol. 27:393–407.

Chitwood, D. H., L. R. Headland, D. L. Filiault, R. Kumar, J. M. Jiménez-Gómez, A. V. Schrager, D. S. Park, J. Peng, N. R. Sinha, and J. N. Maloof. 2012. Native environment modulates leaf size and response to simulated foliar shade across wild tomato species. PLoS One 7:1–15.

Christin, P.-A., C. P. Osborne, D. S. Chatelet, J. T. Columbus, G. Besnard, T. R. Hodkinson, L. M. Garrison, M. S. Vorontsova, and E. J. Edwards. 2013. Anatomical enablers and the evolution of C_4 photosynthesis in grasses. Proc. Natl. Acad. Sci. 110:1381–1386.

Christin, P.-A., E. Spriggs, C. P. Osborne, C. A. E. Strömberg, N. Salamin, and E. J. Edwards. 2014. Molecular dating, evolutionary rates, and the age of the grasses. Syst.

Biol. 63:153–165.

Clayton, W. D., and S. A. Renvoize. 1986. Genera graminum. Grasses of the World. Kew Bull. Addit. Ser. 13.

Clayton, W. D., M. S. Vorontsova, K. T. Harman, and H. Williamson. 2016. GrassBase-The online world grass flora. Kew: The Board of Trustees, Royal Botanic Gardens, Kew.

Cooper, N., G. H. Thomas, C. Venditti, A. Meade, and R. P. Freckleton. 2016. A cautionary note on the use of Ornstein Uhlenbeck models in macroevolutionary studies. Biol. J. Linn. Soc. 118:64–77. Oxford University Press.

Cotton, J. L., W. P. Wysocki, L. G. Clark, S. A. Kelchner, J. C. Pires, P. P. Edger, D. Mayfield-Jones, and M. R. Duvall. 2015. Resolving deep relationships of PACMAD grasses: a phylogenomic approach. BMC Plant Biol. 15:1–11. BMC Plant Biology.

Crepet, W. L. ., and G. D. . Feldman. 1991. The Earliest Remains of Grasses in the Fossil Record. Am. J. Bot. 78:1010–1014.

Donoghue, M. J., and E. J. Edwards. 2014. Biome Shifts and Niche Evolution in Plants. Annu. Rev. Ecol. Evol. Syst. 45:547–572.

Doyle, J. A., P. K. Endress, and G. R. Upchurch. 2008. Early Cretaceous monocots: A phylogenetic evaluation. Acta Mus. Nat. Pragae, Ser. B, Hist. Nat 64:59–872.

Edwards, E. J., C. P. Osborne, C. A. E. Strömberg, S. A. Smith, and C. G. Consortium. 2010. The Origins of C_4 Grasslands: Integrating Evolutionary and Ecosystem Science. Science (80-.). 328:587–591.

Edwards, E. J., and S. A. Smith. 2010. Phylogenetic analyses reveal the shady history of C_4 grasses. Proc. Natl. Acad. Sci. 107:2532–2537.

Givnish, T. J. 1979. On the adaptive significance of leaf form. Pp. 375–407 *in* O. T. Solbrig, S. Jain, G. B. Johnson, and P. H. Raven, eds. Topics in plant population biology. Columbia University Press, New York, New York.

Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, A. Jarvis, G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. 25:1965–1978.

Iles, W. J. D., S. Y. Smith, M. A. Gandolfo, and S. W. Graham. 2015. Monocot fossils suitable for molecular dating analyses. Bot. J. Linn. Soc. 178:346–374.

Ives, A. R., and T. Garland Jr. 2009. Phylogenetic logistic regression for binary dependent variables. Syst. Biol. 59:9–26. Oxford University Press.

Jacobs, B. F., J. D. Kingston, and L. L. Jacobs. 1999. The origin of grass-dominated ecosystems. Ann. Missouri Bot. Gard. 86:590–643.

Jones, S. S., S. V Burke, and M. R. Duvall. 2014. Phylogenomics, molecular evolution, and estimated ages of lineages from the deep phylogeny of Poaceae. Plant Syst. Evol. 300:1421–1436. Springer.

Kark, S., and B. J. van Rensburg. 2006. Ecotones: marginal or central areas of transition? Isr. J. Ecol. Evol. 52:29–53.

Katoh, K., K. Misawa, K. Kuma, and T. Miyata. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30:3059–3066.

Kellogg, E. A. 2001. Evolutionary history of the grasses. Plant Physiol. 125:1198–1205.

Kellogg, E. A. 2015. The families and genera of vascular plants: Volume 13. Flowering Plants Monocots: Poaceae. Springer International.

Kriticos, D. J., V. Jarošik, and N. Ota. 2015. Extending the suite of bioclim variables: A proposed registry system and case study using principal components analysis.

Liu, H., and C. P. Osborne. 2015. Water relations traits of C_4 grasses depend on phylogenetic lineage, photosynthetic pathway, and habitat water availability. J. Exp. Bot. 66:761–773.

Miller, M. A., W. Pfeiffer, and T. Schwartz. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Pp. 1–8 *in* 2010 Gateway Computing Environments Workshop, GCE 2010. 14 Nov. 2010, New Orleans, LA.

Moritz, C., J. L. Patton, C. J. Schneider, and T. B. Smith. 2000. Diversification of rainforest faunas: an integrated molecular approach. Annu. Rev. Ecol. Syst. 31:533–563.

Niklas, K. J. 1999. A mechanical perspective on foliage leaf form and function. New Phytol. 143:19–31.

Orme, D. 2013. The caper package : comparative analysis of phylogenetics and evolution in R. R Packag. version 0.5, 2 1–36.

Osborne, C. P., and R. P. Freckleton. 2009. Ecological selection pressures for C_4 photosynthesis in the grasses. Proc. Biol. Sci. 276:1753–60.

Oyarzabal, M., J. M. Paruelo, P. Federico, M. Oesterheld, and W. K. Lauenroth. 2008. Trait differences between grass species along a climatic gradient in South and North America. J. Veg. Sci. 19:183–192.

Paradis, E., J. Claude, and K. Strimmer. 2004. APE: Analyses of phylogenetics and evolution

in R language. Bioinformatics 20:289–290.

Pennell, M. W., J. M. Eastman, G. J. Slater, J. W. Brown, J. C. Uyeda, R. G. FitzJohn, M. E. Alfaro, and L. J. Harmon. 2014. geiger v2. 0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. Bioinformatics 30:2216–2218. Oxford University Press.

Pohl, R. W. 1965. Dissecting equipment and materials for the study of minute plant structures. Rhodora 67:95–96.

Poinar, G. 2011. Silica bodies in the Early Cretaceous Programinis laminatus (Angiospermae: Poales). Palaeodiversity 4:1–6.

Poinar, G., S. Alderman, and J. Wunderlich. 2015. One hundred million year old ergot : psychotropic compounds in the Cretaceous ? Paleodiversity 8:13–19.

Poinar, G. O. 2003. Early Cretaceous grass-like monocots in Burmese amber. Aust. Syst. Bot. 497–504.

Prasad, V., C. A. E. Strömberg, A. D. Leaché, B. Samant, R. Patnaik, L. Tang, D. M. Mohabey, S. Ge, and A. Sahni. 2011. Late Cretaceous origin of the rice tribe provides evidence for early diversification in Poaceae. Nat. Commun. 2:480.

R Core Team. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rambaut, A., and A. J. Drummond. 2009. Tracer 1.5. 0. MCMC trace analysis tool.

Rao, X., and R. A. Dixon. 2016. The differences between NAD-ME and NADP-ME subtypes of C_4 photosynthesis: more than decarboxylating enzymes. Front. Plant Sci. 7:1–9.

Rasband, W. 2012. ImageJ. U. S. Natl. Institutes Heal. Bethesda, Maryland, USA //imagej.nih.gov/ij/.

Revell, L. J. 2012. phytools: An R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. 3:217–223.

Ryser, P., and S. Wahl. 2001. Interspecific variation in RGR and the underlying traits among 24 grass species grown in full daylight. Plant Biol. 3:426–436.

Saarela, J. M., S. V. Burke, W. P. Wysocki, M. D. Barrett, L. G. Clark, J. M. Craine, P. M. Peterson, R. J. Soreng, M. S. Vorontsova, and M. R. Duvall. 2018. A 250 plastome phylogeny of the grass family (Poaceae): topological support under different data partitions. PeerJ 6:e4299.

Schilthuizen, M. 2000. Ecotone: speciation-prone. Trends Ecol. Evol. 15:130–131.

Schneider, J., E. Döring, K. W. Hilu, M. Röser, S. Taxon, N. May, and D. Elke. 2014. Phylogenetic structure of the grass subfamily Pooideae based on comparison of plastid matK gene-3′ trnK exon and nuclear ITS sequences phylogenetic structure of the grass subfamily Pooideae based on sequences. Taxon 58:405–424.

Smith, T. B., R. K. Wayne, D. J. Girman, and M. W. Bruford. 1997. A role for ecotones in generating rainforest biodiversity. Science (80-.). 276:1855–1857.

Stamatakis, A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313.

Staver, A. C., S. Archibald, and S. A. Levin. 2011. The global extent and determinants of savanna and forest as alternative biome states. Science (80-.). 334:230–232.

Stayton, C. T. 2015. The definition, recognition, and interpretation of convergent evolution, and two new measures for quantifying and assessing the significance of convergence. Evolution (N. Y). 69:2140–2153.

Stebbins, G. L. 1972. The evolution of the grass family. Pp. 5–23 *in* The biology and utilization of grasses.

Strömberg, C. A. E. 2005. Decoupled taxonomic radiation and ecological expansion of open-habitat grasses in the Cenozoic of North America. PNAS 102:11980–11984.

Strömberg, C. A. E. 2011. Evolution of grasses and grassland ecosystems. Annu. Rev. Earth Planet. Sci. 39:517–544.

Teisher, J. K., M. R. McKain, B. A. Schaal, and E. A. Kellogg. 2017. Polyphyly of Arundinoideae (Poaceae) and evolution of the twisted geniculate lemma awn. Ann. Bot. 120:725–738.

Thomasson, J. R. 1985. Miocene Fossil Grasses : Possible Adaptation in Reproductive Bracts (Lemma and Palea). Ann. Missouri Bot. Gard. 72:843–851.

Vicentini, A., J. C. Barber, S. S. Aliscioni, L. M. Giussani, and E. a. Kellogg. 2008. The age of the grasses and clusters of origins of C_4 photosynthesis. Glob. Chang. Biol. 14:2963–2977.

Walther, H. 1974. Ergänzungen zur Flora von Seifhennersdorf, Sachsen. II. Teil – Abh. Abh. Staatl. Mus. Miner. Geol. Dresden 21:143–185.

Wright, I. J., N. Dong, V. Maire, I. C. Prentice, M. Westoby, R. V Gallagher, B. F. Jacobs, R. Kooyman, E. A. Law, M. R. Leishman, Ü. Niinemets, P. B. Reich, L. Sack, R. Villar, and P. Wilf. 2017. Global climatic drivers of leaf size. Science (80-.). 357:917–921.

Wu, Y., H.-L. You, and X.-Q. Li. 2017. Dinosaur-associated Poaceae epidermis and phytoliths from the Early Cretaceous of China. Natl. Sci. Rev.

Wysocki, W. P., L. G. Clark, L. Attigala, E. Ruiz-Sanchez, and M. R. Duvall. 2015. Evolution of the bamboos (Bambusoideae; Poaceae): a full plastome phylogenomic analysis. BMC Evol. Biol. 15:50.

Yamanaka, S., H. Takeda, S. Komatsubara, F. Ito, H. Usami, E. Togawa, and K. Yoshino. 2009. Structures and physiological functions of silica bodies in the epidermis of rice plants. Appl. Phys. Lett. 95:123703. AIP.

Yates, M. J., G. Anthony Verboom, A. G. Rebelo, and M. D. Cramer. 2010. Ecophysiological significance of leaf size variation in Proteaceae from the Cape Floristic Region. Funct. Ecol. 24:485–492.

TABLES

Table 1. Ancestral state estimates (ASE) of habitat (probability for each of three habitat states for key nodes) and the mean and 95% CI for the ASE of mean annual precipitation. Reconstructions for habitat and leaf shape are based on a summary of 1000 stochastic map simulations performed by the *make.simmap* function in Phytools v.0.6. The reconstruction of precipitation was performed using the ace function in ape v.5.2.

Crown nodes of clade	Open habitat	Forest Margins	Understory	Linear leaves	Lanceolate leaves	Ovate leaves	Annual precipitation (mm) (and 95% CI)
Flagellariaceae + Joinvilleaceae + Ecdeiocoleaceae + Poaceae	0%	15%	85%	18%	51%	31%	1648 (1301-1994)
Joinvilleaceae + Ecdeiocoleaceae + Poaceae	0%	18%	82%	18%	51%	31%	1622 (1332-1912)
Poaceae	0%	22%	78%	16%	51%	34%	1647 (1413-1881)

Anomochlooideae	1%	20%	79%	9%	32%	60%	1693 (1385-2001)
Spikelet Clade ((Pharoideae + (Puelioideae+Core Poaceae))	1%	34%	65%	18%	53%	29%	1647 (1415-1880)
Pharoideae	1%	24%	76%	1%	38%	61%	2070 (1815-2326)
Bistigmatic clade (Puelioideae+PACMAD+ BOP)	2%	61%	38%	27%	62%	12%	1608 (1392-1824)
Puelioideae	1%	24%	75%	2%	22%	76%	2065 (1781-2348)
Core Grasses (PACMAD+BOP)	3%	74%	23%	37%	60%	3%	1495 (1304-1687)
BOP (Bambusoideae, Oryzoideae, Pooideae)	3%	73%	24%	39%	59%	2%	1381 (1217-1546)
Oryzoideae	3%	64%	33%	52%	46%	1%	1414 (1218-1610)
Oryzeae	21%	65%	14%	61%	36%	3%	1361 (1174-1549)
Bambusoideae	2%	59%	39%	11%	72%	17%	1480 (1287-1673)
Arundinarieae	0%	93%	7%	0%	99%	0%	1449 (1339-1560)
Bambuseae	0%	95%	5%	0%	99%	0%	1969 (1848-2089)
Olyreae	0%	23%	77%	0%	26%	74%	1858 (1684-2032)
Pooideae	52%	45%	3%	78%	22%	0%	1332 (1168-1495)
Stipeae	96%	4%	0%	100 %	0%	0%	783 (591-976)
Core Pooids	91%	9%	0%	100 %	0%	0%	830 (703-958)

Poeae	100 %	0%	0%	100 %	0%	0%	818 (701-935)
Triticodae	100 %	0%	0%	100 %	0%	0%	721 (583-859)
PACMAD (Panicoideae, Aristidoideae, Chloridoideae, Micrairoideae, Arundinoideae, Danthonioideae)	85%	15%	0%	88%	12%	0%	1402 (1237-1566)
ACMAD (Aristidoideae, Chloridoideae, Micrairoideae, Arundinoideae, Danthonioideae)	100%	0%	0%	100 %	1%	0%	1219 (1040-1398)
Aristidoideae	99%	1%	0%	99%	1%	0%	864 (635-1093)
Micrairoideae	100 %	0%	0%	95%	5%	0%	1366 (1185-1546)
Arundinoideae	100 %	0%	0%	99%	1%	0%	1067 (891-1244)
Danthonioideae	100 %	0%	0%	100 %	0%	0%	941 (765-1117)
Chloridoideae	100 %	0%	0%	100 %	0%	0%	808 (649-966)
Earliest Chloridoid C_4 node	100 %	0%	0%	100 %	0%	0%	780 (643-917)
Panicoideae	88%	11%	0%	90%	10%	0%	1447 (1285-1609)
Andropogoneae + Arundinelleae	100 %	0%	0%	99%	1%	0%	1587 (1447-1727)
Paniceae	99%	1%	0%	98%	2%	0%	1386 (1248-1525)
Paspaleae	100 %	0%	0%	100 %	0%	0%	1488 (1346-1630)

Table 2. Transition rates with 95% confidence intervals between three habitat types estimated using maximum likelihood in the R package corHMM 1.22 (upper line in each cell) and the average (minimum and maximum) number of transitions between states estimated from 1000 simulations in the R package phytools 0.6 (bottom line).

	Open Habitat	Forest Margins	Forest Understory
Open habitat	--	0.0033 (0.0013-0.0053) 31.09 (20-45)	0.0007 (-0.0007-0.002) 6.30 (0-15)
Forest margins	0.0124 (-0.0013-0.0262) 24.65 (12-41)	--	0.0212 (0.0067-0.0357) 41.78 (22-61)
Understory	0 (-0.0153-0.0153) 0 (0-0)	0.0129 (-0.0055-0.0313) 24.24 (12-47)	--

Table 3. Transition rates with 95% confidence intervals between three leaf shapes estimated using maximum likelihood in the R package corHMM 1.22 (upper line in each cell) and the average (minimum and maximum) number of transitions between states estimated from 1000 simulations in the R package phytools 0.6 (bottom line).

	Linear	Lanceolate	Ovate
Linear	--	0.016 (0.025-0.007) 40.9 (26-58)	0 (0.017--0.017) 0 (0-0)
Lanceolate	0.004 (0.005-0.002) 37.8 (21-60)	--	0.013 (0.023-0.002) 27 (20-37)
Ovate	0 (0.001--0.001) 0 (0-0)	0.011 (0.018-0.005) 12.6 (3-29)	--

Table 4. Tests of phylogenetic signal (Pagel's λ, and Blomberg's kappa) for leaf shape and size based on 10,000 simulations; mean and standard error of values for leaf outline, w:l ratio, length, width and area, and the results of phylogenetic ANOVA between C_3 and C_4 species and between species in open habitats vs. forest margins and forest understories.

	Pagel's λ P(λ)=0	Blomberg's K P(K)=0	C_3 Mean (SE)	C_4 Mean (SE)	Significant difference C_3 vs. C_4	Open Habitat Mean (SD)	Forest Margins Mean (SE)	Forest Understory Mean (SE)	Significant difference Habitat
Leaf Outline	--	K=0.766†, P=0.001	--	--	P=0.008	--	--	--	P<0.001*
W:L Ratio	λ=0.963 P<0.001	K=1.148 P<0.001*	0.09 (0.007)	0.052 (0.006)	P=0.495	0.04 (0.0027)	0.112 (0.009)	0.217 (0.022)	P=0.001*
Length (cm)	λ<0 P=1	K=0.331 P=0.018*	15.82 (1.12)	13.55 (1.23)	P=0.69	15.31 (1.21)	14.75 (1.07)	13.93 (1.84)	P=0.93
Width (cm)	λ= 0.861 P<0.001*	K=0.993 P<0.001*	1.22 (0.11)	0.54 (0.05)	P=0.35	0.517 (0.052)	1.466 (1.145)	2.669 (0.363)	P=0.001*
Area (cm^2)	λ=0.674 P<0.001	K=0.519 P<0.001*	18.48 (4.19)	6.81 (1.19)	P=0.382	10.469 (3.973)	17.695 (3.063)	30.55 (6.113)	P=0.002*

†Calculated using the multivariate version of the K-statistic (Adams 2014a)

Table 5. Phylogenetic generalized least squares results for explanatory models of leaf shape and size variation. An asterisk indicates that the model was significant after Bonferroni correction for multiple tests.

Regression Model	**AICc**	**F [DF]**	**R^2**	**P**
Leaf shape (outline) ~ Habitat (3-categories)		26.76 [2,203]	0.21	<0.001*
Leaf shape (outline) ~ Photosynthesis type		3.54 [1,204]	0.017	=0.014
Leaf shape (outline) ~ W:L Ratio		1130.7 [1,204]	0.847	<0.001*
Leaf shape (outline) ~ log(Area)		1.297 [1,204]	0.006	=0.424
Leaf shape (outline) ~ log(Length)		20.541 [1,204]	0.091	<0.001*
Leaf shape (outline) ~ log(Width)		32.879 [1,204]	0.139	<0.001*
W:L Ratio ~ Habitat (3)		29.27 [2,203]	0.216	<<0.0001*
Log(width) ~ Habitat (3)	442.86	24.28 [2,203]	0.185	<<0.0001*
Log(width) ~ Habitat (3) + Annual mean moisture index	432.73	22.72 [3,202]	0.241	<<0.0001*
Log(width) ~ Habitat (3) + Highest weekly solar radiation	429.9	23.3 [3,202]	0.246	<<0.0001*
Log(width) ~ Habitat (3) + Annual mean moisture index + Highest weekly radiation	430.33	18.36 [4,201]	0.253	<<0.0001*
Log(area) ~ Habitat (2b)	666.89	16.67 [1,204]	0.071	<<0.0001*
Log(area) ~ Habitat (2b) + Annual mean moisture index	656.38	17.54 [2,203]	0.139	<<0.0001*
Log(area) ~ Habitat (2b) + Highest weekly solar radiation	647.41	50.05 [2,202]	0.324	<<0.0001*
Log(area) ~ Habitat (2b) + Annual mean moisture index + Highest weekly solar radiation	647.88	33.99 [3,202]	0.326	<<0.0001*

Table 6. Leaf shape and size parameters (optimum value estimated from the AICc weighted average from six models of evolution (BM, BMS, OU1, OUM, OUMV): mean and 95% CI for leaf shape (width:length ratio) and size (length, width and area) for each parameter. Parameters were estimated for leaf traits in each habitat using AICc weights from all models with positive eigenvalues with > 1% AICc weight.

	Parameter	Open	Forest margins	Understory
W:L Ratio	Optimum (95% CI)	0.0376 (0.0308-0.0435)	0.0823 (0.0558-0.1087)	0.2242 (0.1756-0.2732)
	Alpha α (95% CI)	0.0495 (0.0312-0.0976)	0.0495 (0.0312-0.0976)	0.0495 (0.0312-0.0976)
	-log α (95% CI) root scaled to 1	−2.27 (−1.60 - −3.43)	−2.27 (−1.60 - −3.43)	−2.27 (−1.60 - −3.43)
	Sigma σ2 (95% CI)	0 (0-0.0001)	0.0003 (0.0002-0.0006)	0.0009(0.0004-0.0017)
Length	Optimum cm (95% CI)	11.71 (10.05-13.72)	13.53 (10.84-16.87)	11.15 (7.67-16.1)
	Alpha α (95% CI)	0.27 (0.04-1.03)	0.27 (0.04-1.03)	0.27 (0.04-1.03)
	-log α (95% CI) root scaled to 1	−1.52 (−0.43 - −2.2)	−1.52 (−0.43 - −2.2)	−1.52 (−0.43 - −2.2)
	Sigma σ2 (95% CI)	0.33 (0.07-1.27)	0.24 (0.05-0.97)	0.26 (0.06-1.07)
Width	Optimum cm (95%)	0.6 (0.34-0.87)	1.29 (0.65-1.88)	3.11 (2.3-3.97)
	Alpha α (95% CI)	0.03 (0.02-0.05)	0.03 (0.02-0.05)	0.03 (0.02-0.05)
	-log α (95% CI) root scaled to 1	−3.23 (−1.65 - −5.26)	−3.23 (−1.65 - −5.26))	−3.23 (−1.65 - −5.26)
	Sigma σ2 (95% CI)	0.02 (0.01-0.03)	0.08 (0.05-0.13)	0.14 (0.06-0.25)
Area	Optimum cm2 (95% CI)	3.29 (2.58-4.25)	13.53 (9.36-20.16)	21.04 (13.49-35.14)
	Alpha α (95% CI)	1.23 (1.05-5.97)	1.23 (1.05-5.97)	1.23 (1.05-5.97)
	-log α (95% CI) root scaled to 1	−5.27 (−2.68 - −5.87)	−5.27 (−2.68 - −5.87)	−5.27 (−2.68 - −5.87)
	Sigma σ2 (95% CI)	2.03 (1.18-237.96)	1.46 (1.08-15.87)	1.48 (1.08-11.47)

FIGURE CAPTIONS

Figure 1. Ancestral state estimation of habitat type and leaf shape in the grass family. a) Maximum likelihood ancestral state estimation of habitat type: open (full sun), forest margins, and forest understory estimated from 578 extant grass genera and three outgroup genera. b) Ancestral state reconstruction of leaf shape with leaf shape expressed as a discrete variable: linear, lanceolate or ovate. Discrete leaf-states based on the genus averages of width:length ratio from 578 extant grass genera and two outgroup genera. Reconstructions are based on a summary of 1000 stochastic map simulations performed by the *make.simmap* function in Phytools v.0.6.

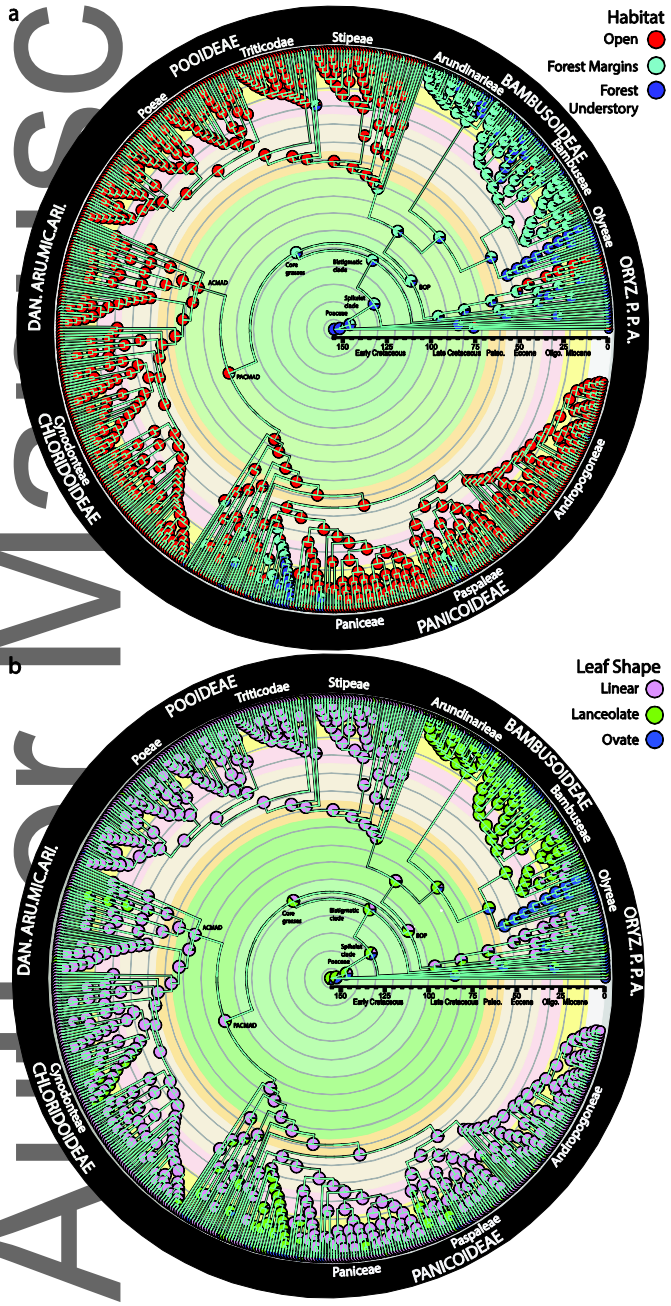

Figure 2. Maximum likelihood ancestral state reconstruction of photosynthesis type (C_3 or C_4) estimated from 578 extant grass genera and three outgroup genera. b) Ancestral state reconstruction of genus-averages of mean annual precipitation estimated from 564 extant grass genera and three outgroup genera. Reconstructions for photosynthesis type are based on a summary of 1000 stochastic map simulations performed by the *make.simmap* function in Phytools v.0.6. The reconstruction of precipitation was performed using the ace function in ape v.5.2.

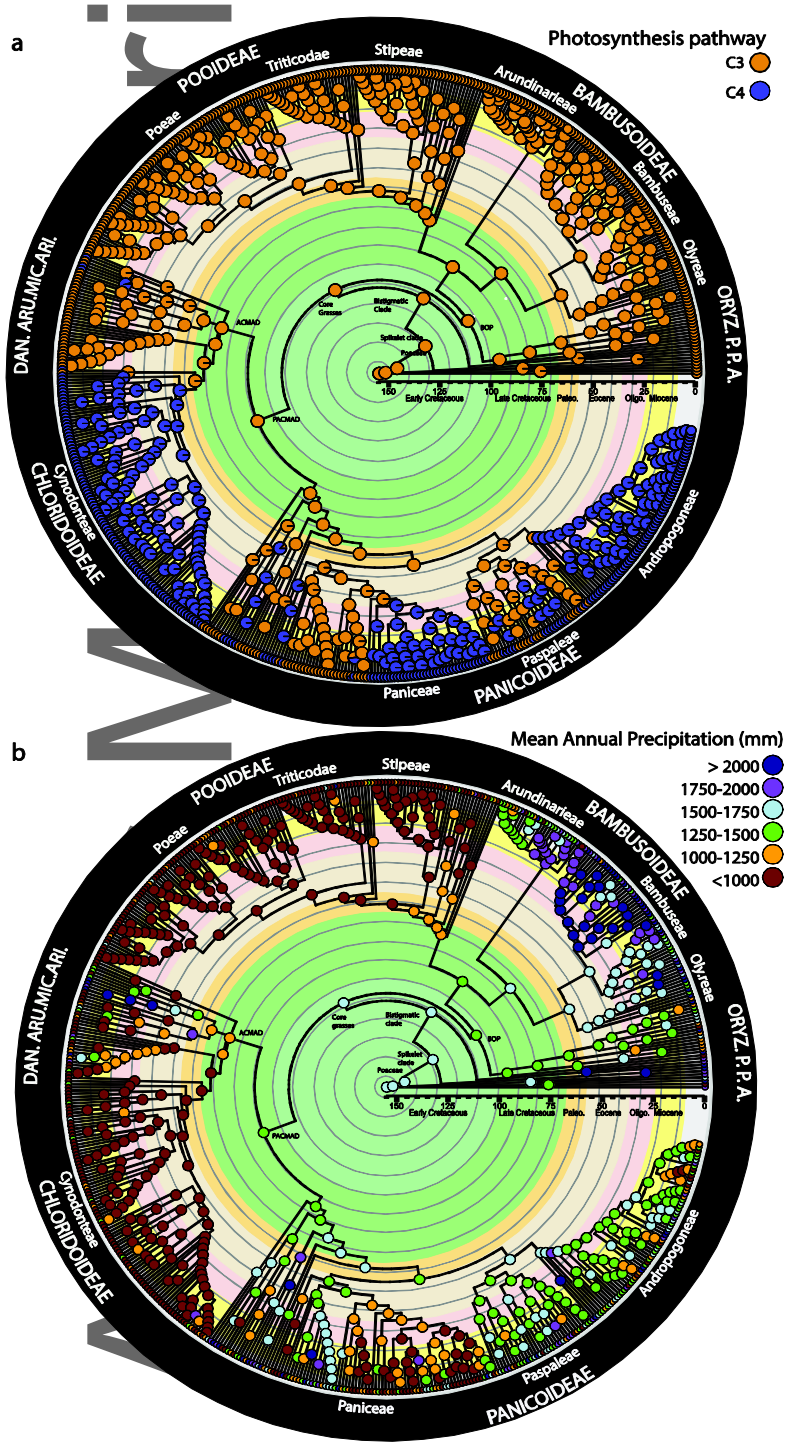

Figure 3. a) Plot of the first and second principal components (PC 1 & 2) for leaf shape (outlines) represented as a phylomorphospace. Leaf outlines represent variation along the x axis. Point colors indicate habitat type (red = open, light blue = forest margin, dark blue = forest understory). b.) Procrustes aligned outlines of 206 composite leaves. The green outline indicates the mean shape of all aligned specimens.

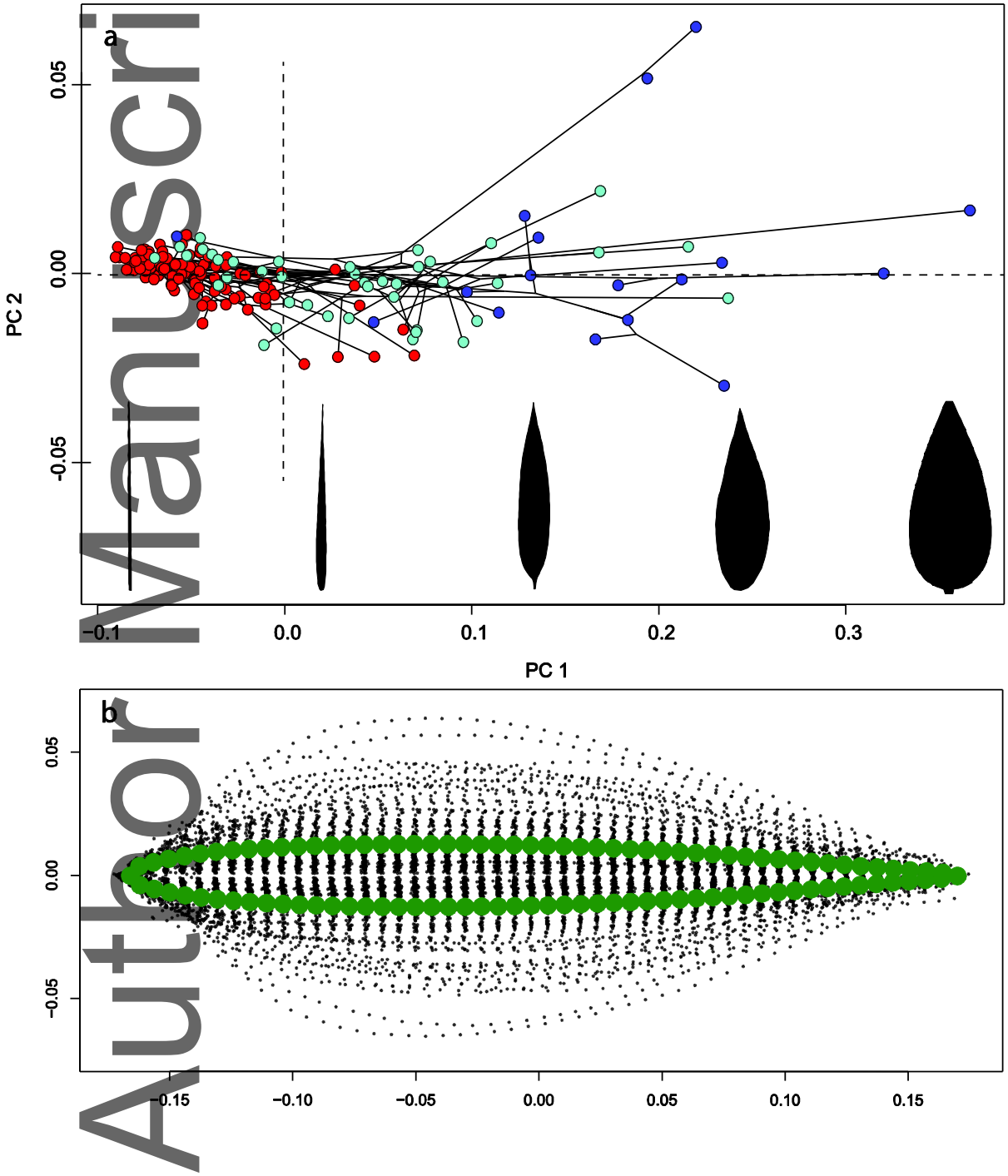

Figure 4. Ancestral state estimation of grass leaf shape showing a sample of the leaf outlines used in the analysis. Branch colors indicate reconstructed values of the first principal component of leaf shape across the phylogeny.

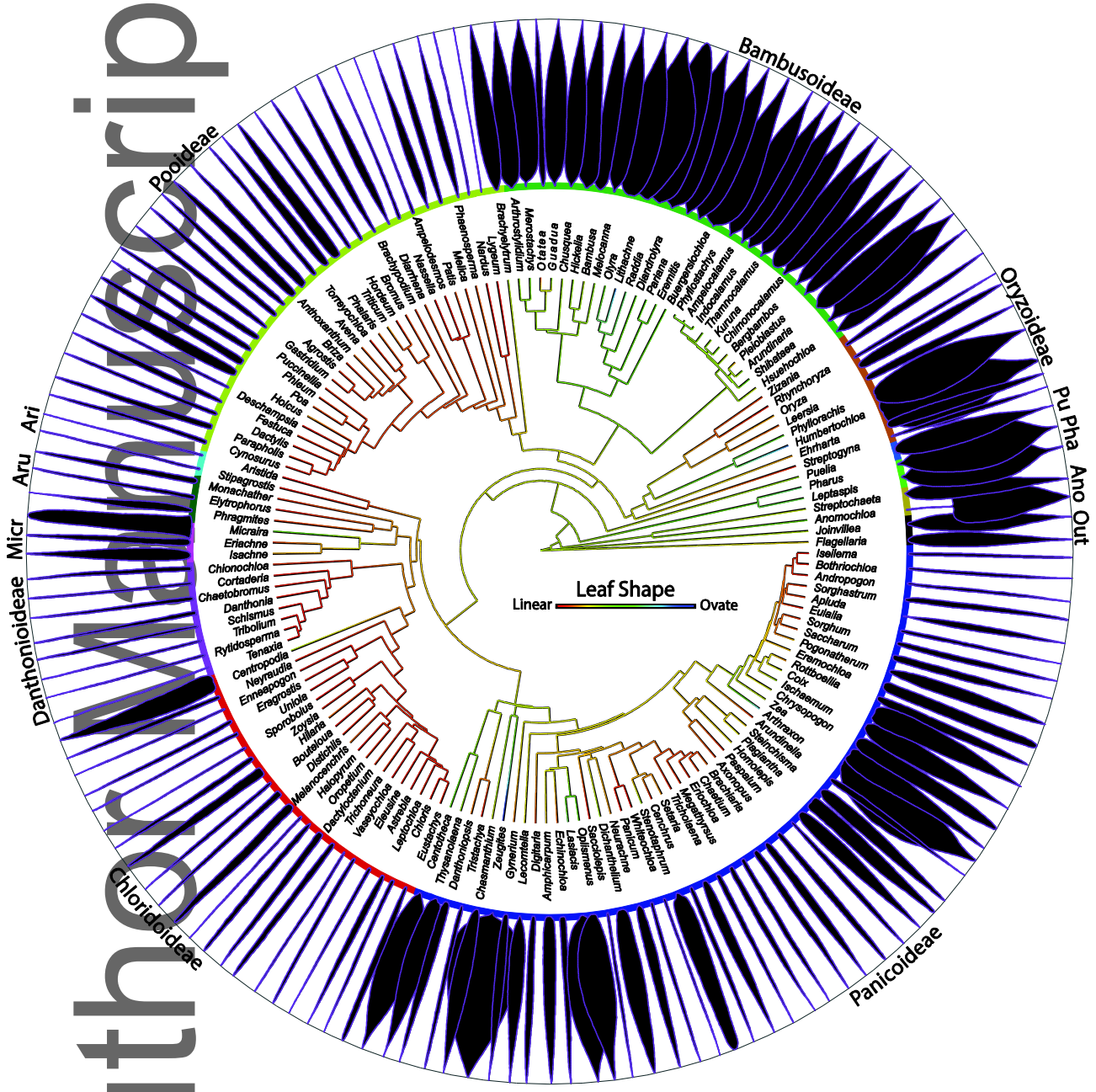

Figure 5. Violin plots of the leaf width:length ratio and log transformed leaf size variables by habitat and photosynthesis type. Different letters above the plots indicate groups with significantly different distributions under phylogenetic ANOVA.

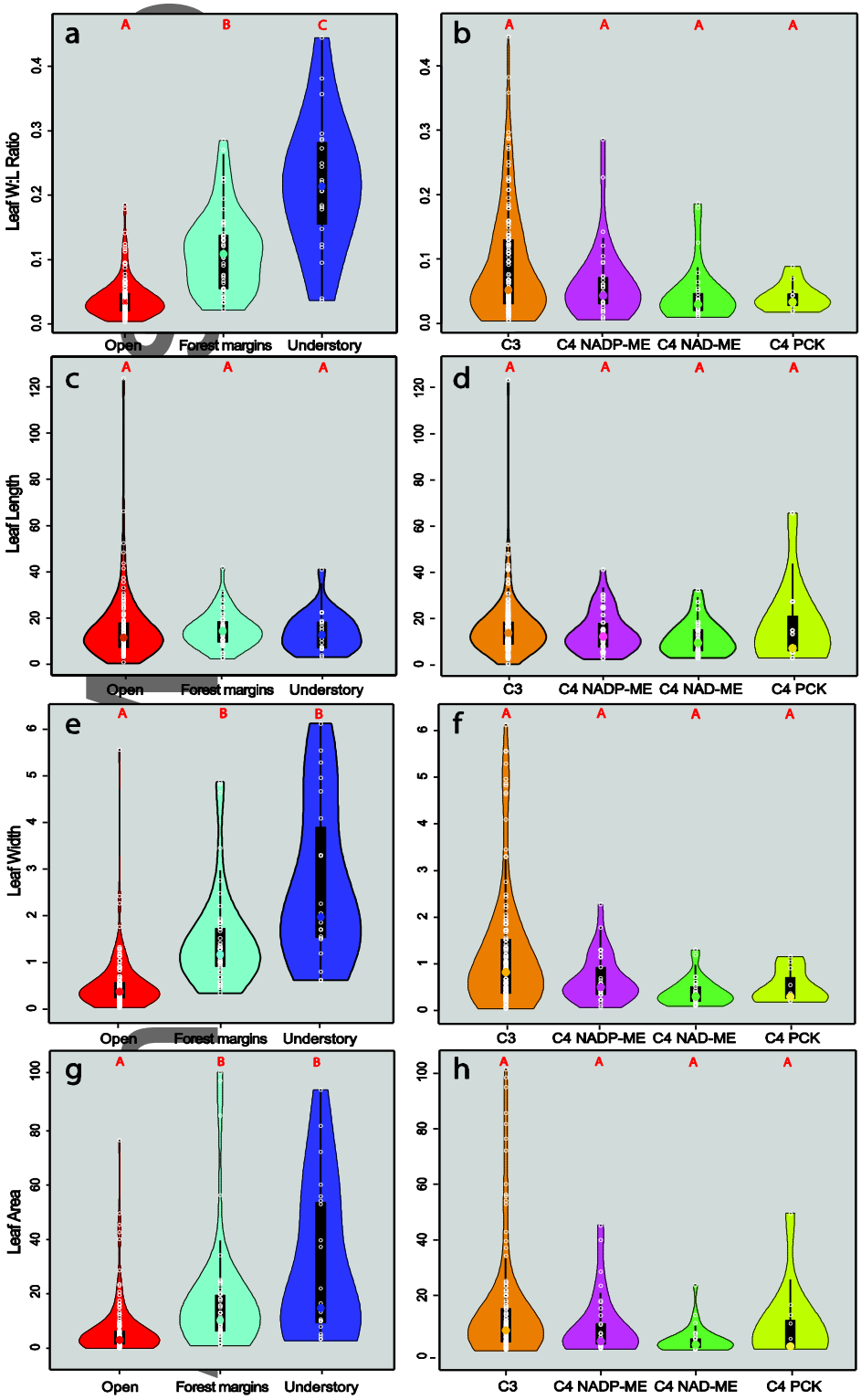

Figure 6. Scatter plot of leaf width and area against highest weekly solar irradiance and mean annual moisture index. Phylogenetic Generalized Least Squares (PGLS) analyses were conducted separately for each of the three habitat categories. '*' indicates a significant (P< 0.01) regression with PGLS. Open habitat species=red, forest margin species = light blue, understory species = dark blue. PGLS was conducted in the R package caper v.0.5.2.