



# C<sub>4</sub> Photosynthesis Promoted Species Diversification during the Miocene Grassland Expansion

Elizabeth L. Spriggs<sup>\*aa</sup>, Pascal-Antoine Christin<sup>ab</sup>, Erika J. Edwards

Department of Ecology and Evolutionary Biology, Brown University, Providence, Rhode Island, United States of America

## Abstract

Identifying how organismal attributes and environmental change affect lineage diversification is essential to our understanding of biodiversity. With the largest phylogeny yet compiled for grasses, we present an example of a key physiological innovation that promoted high diversification rates. C<sub>4</sub> photosynthesis, a complex suite of traits that improves photosynthetic efficiency under conditions of drought, high temperatures, and low atmospheric CO<sub>2</sub>, has evolved repeatedly in one lineage of grasses and was consistently associated with elevated diversification rates. In most cases there was a significant lag time between the origin of the pathway and subsequent radiations, suggesting that the 'C<sub>4</sub> effect' is complex and derives from the interplay of the C<sub>4</sub> syndrome with other factors. We also identified comparable radiations occurring during the same time period in C<sub>3</sub> Pooid grasses, a diverse, cold-adapted grassland lineage that has never evolved C<sub>4</sub> photosynthesis. The mid to late Miocene was an especially important period of both C<sub>3</sub> and C<sub>4</sub> grass diversification, coincident with the global development of extensive, open biomes in both warm and cool climates. As is likely true for most "key innovations", the C<sub>4</sub> effect is context dependent and only relevant within a particular organismal background and when particular ecological opportunities became available.

**Citation:** Spriggs EL, Christin P-A, Edwards EJ (2014) C<sub>4</sub> Photosynthesis Promoted Species Diversification during the Miocene Grassland Expansion. PLoS ONE 9(5): e97722. doi:10.1371/journal.pone.0097722

**Editor:** Robin Allaby, University of Warwick, United Kingdom

**Received:** August 15, 2013; **Accepted:** April 24, 2014; **Published:** May 16, 2014

**Copyright:** © 2014 Spriggs et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported by National Science Foundation grant IOS0843231 to EJE and a Marie Curie IOF 252568 fellowship to PAC. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: elizabeth.spriggs@yale.edu

<sup>aa</sup> Current address: Department of Ecology and Evolutionary Biology, Yale University, New Haven, Connecticut, United States of America

<sup>ab</sup> Current address: Department of Animal and Plant Sciences, University of Sheffield, Sheffield, United Kingdom

## Introduction

Within flowering plants, the grasses (Poaceae) are a remarkable clade, in terms of both species richness and ecological breadth. Comprising over 11,000 species, grasses are exceptionally diverse and a dominant feature of most open habitats throughout the world. Although many share a common morphological form, important physiological differences define various groups of grasses and act to sort these into environmental types.

Grasses living in tropical and subtropical grassland or savanna systems almost exclusively utilize the C<sub>4</sub> photosynthetic pathway [1–3]. This trait is a complex modification over the ancestral C<sub>3</sub> pathway that confers an advantage in open, hot, and dry conditions by concentrating CO<sub>2</sub> inside plant cells and preventing high levels of photorespiration [4]. C<sub>4</sub> photosynthesis characterizes several ecologically dominant, species-rich lineages, suggesting that the C<sub>4</sub> trait may also promote lineage diversification, via either a reduction in extinction rate, an increase in speciation rate, or a combination of both. In the past decade, molecular phylogenies have revealed the existence of three species-poor grass lineages successively sister to the rest of Poaceae and have placed the bulk of grass diversity in either the BEP or PACMAD clade (Figure 1) [5–10]. All of the 22–24 C<sub>4</sub> origins occur within the PACMAD clade, while the similarly sized BEP is entirely C<sub>3</sub> [5–10]. This clustering of all C<sub>4</sub> origins in one of the two major grass lineages may be partly due to increased evolutionary accessibility to the C<sub>4</sub>

trait in this clade, based on a shared set of leaf anatomical attributes [11].

In this study, we use phylogenetic comparative methods on large datasets to test for the effect of C<sub>4</sub> photosynthesis on diversification rates within grasses. While a densely sampled phylogeny of the entire grass lineage is central to accurately identifying shifts in diversification, most previous phylogenetic efforts have concentrated on relatively small subgroups, with the result that few markers are consistently sampled throughout the lineage, and many are difficult to align across distantly related taxa [[e.g. 12–20]]. Previous investigations of grass diversification rates have been hindered by this data structure and have included molecular data for less than 5% of grass diversity [21,22]. To incorporate as many species as possible without introducing large amounts of missing data into the sequence alignments, we constructed 14 separate phylogenies, corresponding to the main lineages inside grasses, and each built with a unique, optimal set of markers. Using a well-resolved backbone phylogeny [10], these were combined into a set of trees that contained 3,595 taxa (Figure 1), encompassing about 30% of the estimated diversity in Poaceae [23]. Using these phylogenies, we found a strong and significant effect of C<sub>4</sub> photosynthesis on diversification. We also explored these trees to identify shifts in diversification independently of any character state information, and interpret these analyses jointly, in the context of C<sub>4</sub> evolution and Miocene grassland expansion.



**Figure 1. Poaceae phylogeny with 3595 taxa.** C<sub>4</sub> lineages are mapped in blue. Red labels indicate the PACMAD clade, yellow labels indicate the BEP clade, and grey labels indicate the early diverging Poaceae lineages. Lineage names are abbreviated as: O.P. Outlying Panicoideae, Ehrh. Ehrhartoideae, Ar.M. Arundinoideae+Micrairoideae, and Arist. Aristidoideae. doi:10.1371/journal.pone.0097722.g001

## Materials and Methods

### Sequence Mining and Matrix Assembly

The majority of recent phylogenetic work in Poaceae has focused on specific subfamilies or genera and has employed a variety of fast-evolving chloroplast and nuclear markers (e.g. [12–20]). The nature of these studies has resulted in a wealth of sequence data for Poaceae, but many markers are both poorly sampled across the entire group and difficult to align across the entire clade. To circumvent the phylogenetic problems that arise from such data, specifically poor alignments, large amounts of

missing sites, and large matrices ill-suited to computationally intensive analyses, we subdivided the tree-building approach. Fourteen sub-trees were constructed separately and subsequently inserted into a fossil-calibrated backbone phylogeny. This approach relies heavily on recent work in the grasses that has resolved deep relationships among the subfamilies and clarified discrepancies in various molecular dating efforts [6–8,10,24].

Sequence data was collected from Genbank with the PHLAWD tool (<http://phlawd.net/> [25]) using the plant GenBank database generated in March 12, 2012. To avoid synonymy problems, all genus names were transformed to those accepted by the Kew

taxonomic database, using the GrassBase [23] synonymy database. Because the taxonomic classification in Genbank is not consistent with the latest developments in grass taxonomy, clades based on GenBank names are not always monophyletic. Species were, therefore, sorted into groups based on previous studies [10] and inspected on preliminary phylogenetic trees as necessary. In general, monophyletic groups were defined to correspond to traditionally recognized clades. The Bambusoideae, Ehrhartoideae, Chloridoideae, Danthoioideae, Andropogoneae, Paspaleae, and Paniceae were all used. The species-poor sister clades Arundinoideae and Micrairoideae were combined, as were the outlying Panicoideae *sensu* GPWGII 2012 [10]. The Pooideae was too large to analyze in one piece, so after marker selection, 3 monophyletic clades were separated from the Pooideae backbone and each was analyzed individually. Two representatives of each separated clade were retained with the remaining backbone Pooideae so that their monophyly and divergence date could be constrained, and the separated lineage could be reinserted later. PHLAWD was then used to create alignments for the most frequently sampled gene regions in each of the 14 clades using a coverage threshold of 0.4 and an identity threshold of 0.1. The three plastid markers *matK*, *ndhF*, and *rbcL* were included in each group and an additional 2 to 10 gene regions were added depending on the group sampled (Table S2). In total, 35 gene regions were incorporated in the analysis (sampling information in Table S1, S2).

Once the alignments were complete, the software trimAl [26] was used to remove sites with more than 70% missing data for each gene region and the MEGA software [27] was used to manually edit the alignment where necessary. In each group, the alignments were concatenated with Phyutility [28] and species names were checked against the GrassBase [23] synonymy database. A small number of names were referenced in Tropicos [29] but not in GrassBase [23], and were consequently considered to be recently described species. Synonyms, misspellings, subspecies, and varieties were manually removed whenever possible to leave a single representative sequence per accepted species. At this point, RAxML software [30] was used to build a tree with 20 maximum likelihood searches, retaining the tree with the highest likelihood score across them. The phylogeny inferred for each group was manually inspected to identify taxa that had very long branches, representing potential errors. The sequences of these taxa were inspected by BLAST searches against GenBank, and putatively erroneous sequences, corresponding to either sequencing or identification errors, were removed.

### Tree Building and Molecular Dating

To estimate the age of the main grass lineages, dating analyses were first performed with a dataset of three previously sampled chloroplast genes and 543 taxa covering the entire grass family [10]. The software BEAST 1.7.2 [31] was run under a GTR+G+I substitution model, a Yule process for the prior distribution of node ages and a log-normal distribution for the prior on evolutionary rates among branches. Time-calibrated trees were obtained with two contrasting hypotheses for the placement of fossils [24]. Under calibration #1, which is based only on macrofossil calibrations and does not take into account fossil phytoliths whose placement is somewhat controversial [32], the crown age of the BEP-PACMAD clade followed a normal calibration density with a mean of 51.2 Ma and a standard deviation of 6.0 Ma [24]. Under calibration #2, which incorporates fossil phytoliths [32], the age of this same node followed a normal calibration density with a mean of 82.4 Ma and a standard deviation of 7.5 Ma [24]. In this second analysis, we

also constrained the stem of Oryzae to obtain dates compatible with phytolith fossil evidence [32], using an exponential distribution with a mean of 10 Ma offset by 67 Ma. For these two analyses, the topology was not fixed, except for the monophyly of the ingroup (all taxa except *Pharus*). Trees were sampled every 5,000 generations for 15,000,000 generations after a burn-in period of 5,000,000 generations. Convergence, effective sample size, and the adequacy of the burn-in period were assessed using Tracer 1.5 [31].

A phylogeny was then inferred separately for each previously defined group of grasses using the software BEAST as described above [31]. Crown node ages were fixed (uniform prior with range of 0.01 around the fixed value) to the dates obtained from the Bayesian consensus phylogeny estimated from the 543-taxon dataset (above), under calibration #1. All trees were then scaled to match the dates under calibration #2. All subsequent analyses were performed on both sets of time-calibrated phylogenetic trees. The monophyly of the ingroup was enforced to ensure proper rooting. For each dataset, two independent Markov Chain Monte Carlo analyses were run for 10–50 million generations, sampling every 1000–5000 generations, depending on the size of the dataset. Convergence, effective sample size, and the adequacy of the burn-in period were assessed using Tracer [31]. A burn-in period of 2,500,000–6,000,000 generations was chosen, again depending on the size of the dataset. For clades of over 150 taxa, convergence from random starting trees was extremely slow, and so the best of our previous 20 maximum likelihood RAxML trees was dated using non-parametric rate smoothing in r8s [33] and used as a starting point for each run.

For each group, the maximum clade credibility tree was selected with TreeAnnotator [31] and the node heights of this tree were scaled in R to match each of the dating hypotheses by multiplying all branch lengths by the fraction (hypothesis root age/current root age). The calibrated phylogenetic trees were then manually inserted into the associated backbone phylogeny of 543 grasses [10], preserving the deep relationships among the groups and forming a set of all-inclusive, ultrametric phylogenies with 3595 species each. With 544 genera represented, this tree contains more than 29% of the species and 71.2% of the recognized genera in Poaceae. Of the missing genera, only 6 have more than 10 species [23].

To take into account both phylogenetic uncertainty and variation in dating hypotheses, we repeated diversification analyses on 100 topologies drawn randomly from the population of trees sampled post burn-in by BEAST for each of our 14 groups. A unique, calibrated phylogeny for each group was scaled and added to each of our two backbone phylogenies to produce 100 alternative phylogenies of the grasses under each set of dating conditions.

### Diversification Analyses

Three approaches were used to analyze the patterns of diversification in Poaceae. First, the BiSSE (Binary State Speciation and Extinction) method [34,35] specifically evaluated the relationship between photosynthetic type and diversification rate. Second, log-scale species richness was compared among sister groups with different photosynthetic types using a Wilcoxon sign ranked test [36]. Third, turboMEDUSA, a likelihood method implemented in R [37], was used to locate and quantify shifts in diversification rates across Poaceae independently of any character state information. Since all of the C<sub>4</sub> origins occur within the PACMAD portion of Poaceae, our focus is on this clade, although we also ran analyses across the entire phylogeny.

To effectively evaluate diversification patterns, it was necessary to determine the richness and distribution of Poaceae species on our phylogeny. Although taxonomic issues remain unsettled in certain areas of Poaceae phylogeny, we were able to approximate the size of most genera using the accepted names in GrassBase [23]. Unless otherwise demonstrated, genera were assumed to be monophyletic and occasionally small genera nested in larger ones were merged. For each genus, the species with the most sequence data was selected as the representative of that group and was assigned the richness of the entire genus. In genera with both C<sub>3</sub> and C<sub>4</sub> taxa, we divided the genus into the minimum number of clades such that the C<sub>3</sub> taxa and the C<sub>4</sub> taxa were monophyletic and each represented by a single tip in our phylogeny. According to estimates from GrassBase [23], the phylogeny inferred in this study contains ~25% of the known *Panicum* species. This genus is, however, highly polyphyletic [38], and diverse sections have been segregated into new genera in the past few years [38–44]. To cope with this uncertainty, the number of *Panicum* species was equally spread among all *Panicum* tips in our phylogeny and the well-supported monophyletic groups of *Panicum* were subsequently collapsed. Using this approach, we were able to assign 11,554 species (95.5% of Poaceae) to a specific tip on our tree [23].

An additional difficulty lay in the potential tendency for large clades on short branches to throw off diversification estimates. Therefore, large clades of over 190 species were split among several representatives. The genus *Poa*, for instance, which contains over 550 species, was divided evenly among the tips corresponding to *Poa pratensis*, *Poa annua*, and *Poa colensoi*. Even with similar subdivisions, excessively small state probabilities occasionally caused the BiSSE likelihood calculations to fail. In these cases, the groups were further subdivided among additional representatives or combined with a sister group to increase the subtending branch length.

Using TurboMEDUSA, the number of shifts in diversification rates was first estimated with the default AICc threshold on our genus level tree (8.4547). Each representative tip was assigned the same richness value used for the BiSSE analyses. This approach suggested 24 shifts, some of which were located on extremely short branches leading to a single tip, with a relatively small number of species. These shifts were no longer identified with a more conservative threshold of 10.5, which suggested 18 shifts in diversification. These shifts were considered more reliable and are reported here.

Pruning our large phylogeny down to single representatives of each genus allowed us to include information about unsampled diversity in our analyses, but it also reduced a substantial amount of branching structure and information. For example, using this approach with TurboMEDUSA precludes the identification of shifts that might occur closer to the tips, within genera for instance. We therefore performed a complimentary TurboMEDUSA analysis on the complete 3595 species tree. We also ran BiSSE analyses on the unpruned, 3595 tip tree, accommodating unsampled diversity by reporting our overall sampling frequency (0.2973 for Poaceae, and 0.2966 for PACMAD [23]) which BiSSE then used in calculations.

Tree inference, dating analyses, and diversification analyses were conducted on the OSCAR HPC cluster at Brown University and the Louise HPC cluster at Yale University. Sequence matrices, trees, and character matrices have all been deposited on dataDryad doi:10.5061/dryad.74b5d.

## Results

### BiSSE

Our BiSSE analyses provided extremely strong support for the evolution of C<sub>4</sub> photosynthesis increasing diversification rates in grasses. All of the BiSSE tests on the BEAST maximum credibility tree strongly rejected the model of equal diversification rates for C<sub>3</sub> and C<sub>4</sub> taxa. This was irrespective of how we accommodated unsampled diversity, whether we analyzed PACMAD separately or together with all of Poaceae, and whether we calibrated our phylogeny with phytoliths or with less-controversial macrofossils (Table 1, Table S3). In most cases, the best-fitting model was a 6-parameter model in which both speciation and extinction rates were different for C<sub>3</sub> and C<sub>4</sub> taxa, but this was often only marginally better than models where C<sub>3</sub> and C<sub>4</sub> lineages differed only in either speciation or extinction rates. Regardless, equal diversification rates were soundly rejected, and in all cases, C<sub>4</sub> diversification rates were inferred to be higher. This C<sub>4</sub> effect can also clearly be seen in a lineage-through-time plot (Figure S1).

The replicate BiSSE analyses on 100 trees from the posterior distribution indicate that these results are also robust to phylogenetic uncertainty. Across 100 replicate PACMAD trees, equal diversification rates were strongly rejected ( $p < 0.01$ ) in all but one tree regardless of whether the missing diversity was distributed proportionally or by genus (Figure 2; Figure S2). Poaceae-wide trees similarly provided additional support, but only when analyses were performed using the full 3,595 taxon tree and missing species were distributed evenly across the tips (Figure S2). Lack of support in these trees when diversity was distributed by genus is probably due to the extremely high numbers of species in C<sub>3</sub> BEP genera like *Festuca*, *Poa*, and *Stipa* which were each clustered at a few tips in the genus-level analyses.

In the PACMAD clade, when the missing species diversity was distributed by genus, BiSSE estimated a net C<sub>4</sub> diversification rate of 0.1458 spp/my and a net C<sub>3</sub> rate of 0.0951 spp/my in the maximum credibility tree under the macrofossil-dating hypothesis (Table 1). When the missing species diversity was instead distributed proportionally, both diversification rates were estimated to be much higher (0.2407 spp/my for C<sub>4</sub>, 0.1677 spp/my for C<sub>3</sub>) (Table S3). When the entire Poaceae tree was used, the estimated C<sub>3</sub> and C<sub>4</sub> rates were very similar to those identified when using the PACMAD tree alone, with significantly higher C<sub>4</sub> rates of diversification ( $p < .01$ ; Table S3).

Under the phytolith-based dating calibration, the results from all analyses were consistent with those based on the macrofossil dates, with the obvious exception that actual net rates of diversification were estimated to be much lower, because grasses were inferred to be older. Similar contrasts between the C<sub>3</sub> and C<sub>4</sub> net diversification rates were evident, and models of equal diversification were rejected under the same conditions at similar levels of confidence (Table S3, Figure S2).

### Sister Group Comparisons

Nearly half of the C<sub>4</sub> lineages in grasses are sister to groups that contain both C<sub>3</sub> and C<sub>4</sub> taxa; however, 12 have exclusively C<sub>3</sub> sister clades and could be compared directly (Table S4). Of these, the C<sub>4</sub> group is equally or more diverse in ten cases and the log-scaled species richness is significantly greater in C<sub>4</sub> groups (Wilcoxon sign ranked test  $p$ -value = 0.0067). While the independent C<sub>4</sub> lineages differ greatly in both age and species richness, they are consistently more diverse than their C<sub>3</sub> sisters (Table S4).

**Table 1.** Parameters inferred for the PACMAD sampling frequency (proportional) tree.

Model	Parameters	Speciation Rates		Extinction Rates		Net Diversification Rates		Transition Rates		Ln Likelihood
		C <sub>3</sub>	C <sub>4</sub>	C <sub>3</sub>	C <sub>4</sub>	C <sub>3</sub>	C <sub>4</sub>	To C <sub>3</sub>	To C <sub>4</sub>	
<b>Dating Hypothesis 1</b>	equal diversification	0.6122	0.6122	0.4076	0.4076	0.2047	0.2047	0.0104	0.0001	-4803.4520
	<b>equal speciation</b>	<b>0.5991</b>	<b>0.5991</b>	<b>0.4488</b>	<b>0.3652</b>	<b>0.1503</b>	<b>0.2339</b>	<b>0.0083</b>	<b>0.0000</b>	<b>-4792.1430</b>
<b>Dating Hypothesis 2</b>	equal extinction	0.5600	0.6219	0.3963	0.3963	0.1637	0.2256	0.0095	0.0000	-4794.8060
	<b>full model</b>	<b>0.6924</b>	<b>0.5667</b>	<b>0.5539</b>	<b>0.3267</b>	<b>0.1386</b>	<b>0.2400</b>	<b>0.0068</b>	<b>0.0001</b>	<b>-4790.4060</b>
equal diversification	4	0.9383	0.9383	0.8718	0.8718	0.0665	0.0665	0.0046	0.0000	-5840.7290
	<b>5</b>	<b>0.9300</b>	<b>0.9300</b>	<b>0.8962</b>	<b>0.8399</b>	<b>0.0338</b>	<b>0.0901</b>	<b>0.0038</b>	<b>0.0000</b>	<b>-5828.5740</b>
equal extinction	5	0.8404	0.8964	0.8051	0.8051	0.0353	0.0913	0.0039	0.0000	-5831.2130
	<b>6</b>	<b>0.3627</b>	<b>0.3011</b>	<b>0.2857</b>	<b>0.1667</b>	<b>0.0770</b>	<b>0.1344</b>	<b>0.0037</b>	<b>0.0001</b>	<b>-5827.0790</b>

Values as inferred under a 4-parameter model (speciation and extinction rates both constrained to be equal), two 5-parameter models (extinction and speciation rates allowed to differ in each respectively), and a 6-parameter model (separate speciation and extinction rates for C<sub>3</sub> and C<sub>4</sub> taxa). doi:10.1371/journal.pone.0097722.t001

## turboMEDUSA

The distinctiveness of several C<sub>4</sub> lineages was also highlighted by turboMEDUSA (Figure 3, Table S5). Under the macrofossil dating hypothesis, when the missing species diversity was distributed by genus, the inferred diversification rate was low in early diverging grass lineages (0.036 spp/my), increased in the common ancestor of the BEP and PACMAD clades (0.143 spp/my and further in a derived clade (core Panicoideae) containing 14–16 C<sub>4</sub> origins (0.220 spp/my; Figure 3). Within the PACMAD lineage, there were an additional five accelerations in diversification rate, four of which occurred within C<sub>4</sub> clades, and the fifth occurring slightly before two subsequent C<sub>4</sub> origins.

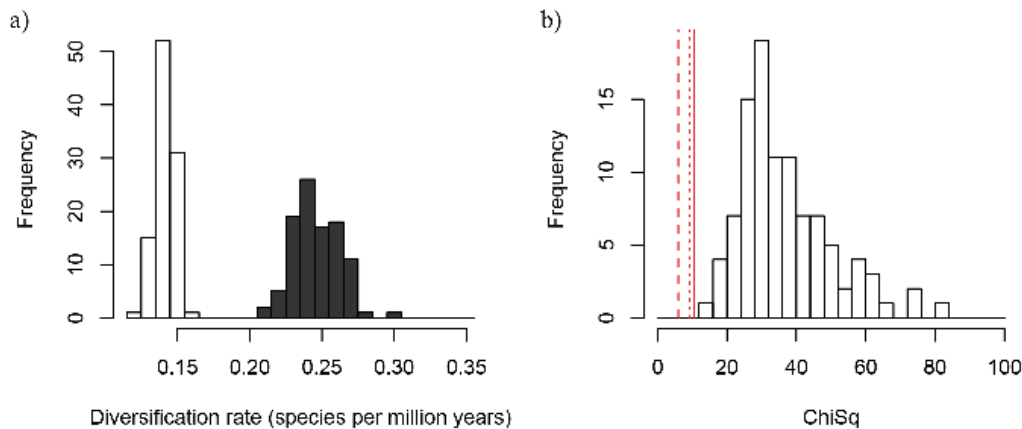
In addition to these PACMAD radiations, turboMEDUSA also detected increases within the cold-adapted Pooideae grasses (Figure 3; Table S5). Although the fastest rate inferred for grasses was in a C<sub>4</sub> genus, *Tripogon*, several other exceptionally high rates were found in the C<sub>3</sub> Pooideae resulting in young and highly diverse taxa such as *Agrostis*, *Poa*, *Elymus*, and *Festuca*. These BEP clade radiations appeared to be concurrent with many of the warm-climate C<sub>4</sub> radiations (Figure 4). The alternative dating hypothesis based on phytoliths identified precisely the same shifts, but all of the rates were slower and the timing of the shifts was earlier (Table S5, Figure 4).

The turboMEDUSA results from our 3595-tip tree (not including any missing species) were generally similar although the exact location of many of the shifts differed significantly (Table S6). In all, 12 shifts in diversification rate were identified in the full tree. There was a clear acceleration from a slow background rate (0.009 spp/my) into the BEP+PACMAD clade (0.164 spp/my), followed by a series of more recent increases. In the PACMAD, there were 5 accelerations, 1 of which occurred within the entirely C<sub>4</sub> lineage Andropogoneae, and 3 of which were nested within C<sub>4</sub> genera (*Muhlenbergia* and *Paspalum*). The final PACMAD acceleration was at the base of the C<sub>3</sub> Danthionioideae, a lineage known, like the Pooideae, for its prominence in cool-climate grasslands [9]. In the BEP clade, accelerations within the Bambusoideae and Pooideae were followed by even more rapid diversification in *Festuca*, the Stipeae, and within *Arundinaria*.

## Discussion

### C<sub>4</sub> photosynthesis promotes elevated diversification rates

In this study, we have approached the effect of C<sub>4</sub> photosynthesis on diversification rates by employing a variety of statistical tests that evaluate diversification patterns in very different manners. The BiSSE analysis provides strong statistical evidence that photosynthetic type influences diversification and that across the entire Poaceae tree, C<sub>4</sub> lineages have radiated faster than C<sub>3</sub> lineages. This evidence is compelling, but it does not take into account variation in clade-specific diversification that may be unrelated to C<sub>4</sub> photosynthesis and entirely dependent on other factors (*e.g.* generation time, dispersal strategies). However, our sister group comparisons confirmed that C<sub>4</sub> lineages are statistically more speciose than their respective C<sub>3</sub> sister clades. This indicates that within any given lineage background, the evolution of C<sub>4</sub> photosynthesis tends to increase the number of descendant species. The “turboMEDUSA” approach estimates the position on the phylogeny where shifts in diversification rates occurred, independent of any trait information [37]. These analyses located multiple accelerations in diversification rate across Poaceae, with many in C<sub>4</sub> PACMAD clades. Taken at face value, the location of most of these identified shifts suggests that there is a significant delay between the origin of the C<sub>4</sub> pathway and subsequent C<sub>4</sub>



**Figure 2. Histograms of BiSSE model inferences based on 100 replicate PACMAD trees.** Each tree had 1774 taxa, and the missing diversity was represented as a proportion (sampling frequency). Black bars indicate C<sub>4</sub> rates, white bars indicate C<sub>3</sub> rates. The panels show: a. Net diversification rates derived from a 6-parameter model, b. Chi Squared values derived from ANOVA comparison of a 6-parameter model and a 4-parameter (equal diversification) model for each tree. The red lines indicate significance values of .05, .01, and .005. doi:10.1371/journal.pone.0097722.g002

radiations. A similar pattern of delayed, Miocene shifts in C<sub>4</sub> grasses was found in a previous diversification study using other methods [22]. Our results are also consistent with other “delayed rise” scenarios that seem to be a pattern identified in many groups of organisms [45–49].

C<sub>4</sub> photosynthesis is a physiological trait that has no obvious links with speciation, so how can it contribute to high diversification rates? The grass lineage is exceptionally diverse for its age, and some of the diversification rates we present here are among the highest reported in the literature under either calibration scheme (Table S5) [50,51]. The C<sub>4</sub> trait might have affected diversification by increasing speciation rates, decreasing extinction rates, or both. While our data do not statistically support one scenario over another, extinction rates are more consistently lower in C<sub>4</sub> lineages across a broad sampling of alternative phylogenetic trees, and lower extinction in C<sub>4</sub> lineages is generally favored, although not always strongly, by BiSSE (Table S3). We suggest that the elevated competitive ability of C<sub>4</sub> plants in hot, open environments [1] allowed newly formed species to survive in a range of environments, thus lowering extinction rates. Other grass-specific traits, such as their propensity for asexual reproduction, polyploidy, and long distance dispersal of seeds by wind might have acted as species-producing mechanisms in various clades throughout Poaceae [21,52]. Under this view, net C<sub>4</sub> diversification rates are higher, because fewer C<sub>4</sub> species generated by these life history mechanisms went extinct than did similarly generated C<sub>3</sub> species. The protracted spread of open systems might also have promoted diversification via a gradually growing patchwork of suitable habitats that allowed for repeated allopatric speciation events. In the tropics, the C<sub>4</sub> trait would have increased the likelihood of successful establishment in these open areas characterized by warm temperatures, moderate drought, and high radiation loads [9,53].

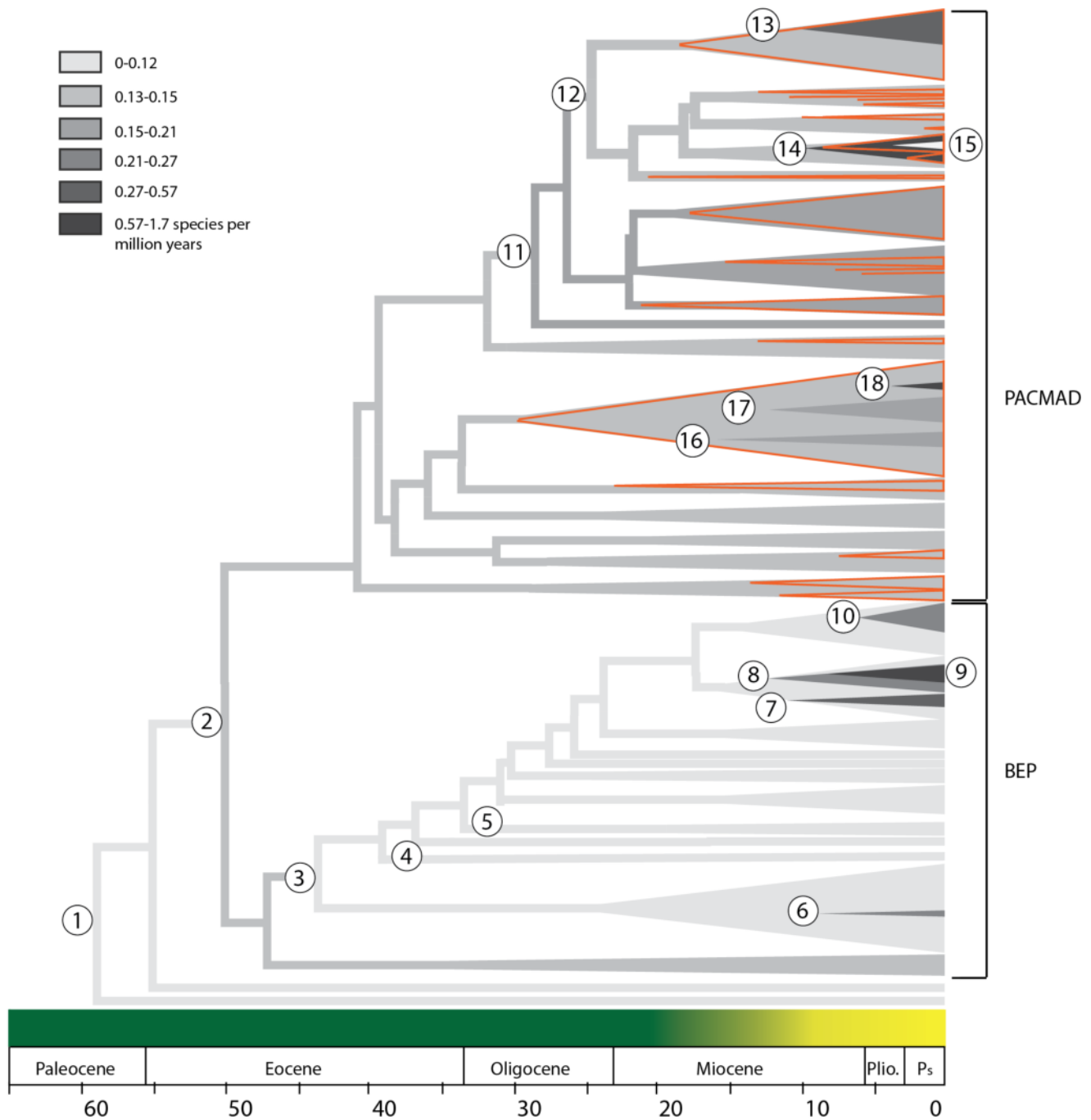
#### Environmental changes and diversification of C<sub>4</sub> grasses

Interestingly, interpreting the context and environmental conditions in which the MEDUSA-identified shifts occurred depends heavily upon the tree-calibrations, and whether or not the recent phylogenetic placement of phytolith fossils is indeed accurate [32]. Under the macrofossil dating hypothesis, without phytolith evidence, C<sub>4</sub> origins appear almost entirely after the Oligocene atmospheric CO<sub>2</sub> drop (with the possible exception of

the Chloridoideae) [54], and almost all are within the time period in which C<sub>3</sub> grasslands are believed to have existed on various continents [55–62,2,63]. Under this younger dating scenario, C<sub>4</sub> photosynthesis may have evolved coincidentally with movement into already established C<sub>3</sub>-dominated grasslands. Previous authors have found evidence that C<sub>4</sub> origins are correlated with shifts to drier, more open habitats, which is congruent with this scenario [9,64]. Under this timescale, all of our pinpointed C<sub>4</sub> radiations occurred between about 5 and 16 Ma, during the time period in which the fossil record indicates a rapid ecological spread of C<sub>4</sub> grasses [65–70,2]. This dating scenario suggests concurrent diversification and rise to dominance of C<sub>4</sub> species, with rapid radiations occurring well after C<sub>4</sub> origins, in open, grassy biomes.

The dating hypothesis based on phytolith evidence suggests slightly different drivers of both C<sub>4</sub> evolution and of diversification. In this older scenario, at least five C<sub>4</sub> origins predate the fossil record of widespread open systems. In this case, C<sub>4</sub> photosynthesis might still have evolved in high-radiation, open habitats, but these areas would have been rare and fragmented across the landscape. Even less intuitive is that some of these origins would have occurred when atmospheric CO<sub>2</sub> levels were very high, which would result in generally lower levels of photorespiration and a weaker selection pressure for evolution of the C<sub>4</sub> pathway. Regardless, in this case, both the C<sub>4</sub> and Pooid radiations would have begun around 15–25 Ma, during a time period that roughly coincides with the global appearance of open C<sub>3</sub>-dominated grasslands on various continents, but when C<sub>4</sub> species were rare [65–70,2]. It would suggest that C<sub>4</sub> species diversified rapidly before they became dominant features of the landscape. Only the shift in *Tripogon* could be coincident with the ecological spread of C<sub>4</sub> grasses. Interestingly, this second version of events shares some similarities to phytolith-based paleo-ecological reconstructions in North America, which suggest that substantial grass taxonomic diversity predated the late Miocene C<sub>4</sub> grassland expansion by 23–27 Ma [60].

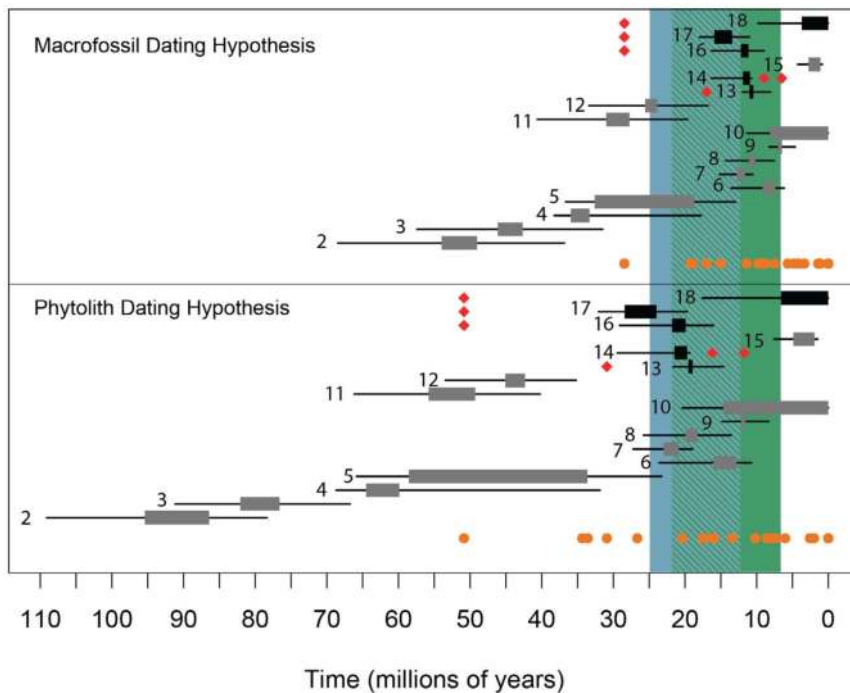
Both scenarios are reasonable, and although there is not enough confidence in the phylogenetic placement of phytolith fossils [32] to prefer the older timescale, we view the abundant grass phytolith record as a remarkable resource that promises to reveal much more about the timing of past events in grass history [60,61,71]. Regardless of the timeframe, we want to express the genuine uncertainty inherent in any of these point estimates of diversification



**Figure 3. Simplified representation of shifts in diversification rates across Poaceae based on calibration #1.** Darker shades of grey indicate higher rates of diversification. Red triangles indicate the approximate phylogenetic placement of C<sub>4</sub> lineages. The left point of each triangle corresponds to the stem age of the inferred shift. The transition from dark green to yellow across the bottom indicates the average timing of the rise of open, grassland habitats on different continents [2]. Rate shifts correspond to Table S5 and are labeled as follows: 1) background diversification rate, 2) BEP+PACMAD 3) Bambusoideae+Pooideae, 4) early diverging Pooideae, 5) Phaneospermateae, 6) *Perrierbambus*+*Bonia* clade 7) Poeae 2 clade, 8) *Poa*+*Alopecurus* clade, 9) *Agrostis*+*Calamagrostis* clade, 10) *Festuca*, 11) Core Paniceae, 12) Andropogoneae+Paspaleae, 13) *Sorghum*+*Andropogon* clade, 14) *Axonopus*+*Paspalum* clade, 15) *Poecilostachys*, 16) *Eragrostis* clade, 17) *Spartina* clade, 18) *Tripogon* (Table S5). doi:10.1371/journal.pone.0097722.g003

rate shifts. The turboMEDUSA analyses were extremely sensitive to many variables, and a slightly different taxon sampling, AICc threshold, branch-length distribution, or distribution of missing taxa would all lead to different nodes being identified—sometimes extremely different nodes (Table S6). This makes this sort of approach to diversification analyses fairly difficult to interpret.

Historical patterns of diversification across such a large number of species over such a long period of time are necessarily highly complex and varied, and it seems unrealistic to assume that there are discrete, abrupt and identifiable shifts in lineage diversification through time. That said, it seems just as unrealistic to assign one diversification rate to C<sub>4</sub> grasses and a second to C<sub>3</sub> grasses. What



**Figure 4. Timing of shifts in diversification rate across both dating hypotheses.** Both grey and black rectangles indicate shifts in diversification rate bounded by the estimated stem and crown node ages for the branch where the shift occurred. Error bars are determined by the 95% confidence interval for each age estimate. Each shift is numbered and corresponds to the shifts in both Figure 3 and Table S5. The black rectangles indicate accelerations that occurred within C<sub>4</sub> clades or immediately before C<sub>4</sub> origins. Orange dots indicate the crown node ages for each of the estimated 24 origins of C<sub>4</sub> photosynthesis in Poaceae. The red diamonds are the origins that are associated with subsequent rate shifts. The blue area indicates the time when grasslands are estimated to have arisen on various continents [55–62,2,63], and the green area is the time when C<sub>3</sub> grasslands were replaced by C<sub>4</sub> grasses [65–70,2]. Overlap between the two is indicated by diagonal hatches. doi:10.1371/journal.pone.0097722.g004

feels preferable in this latter case is that this is a more diffuse analysis, averaged across all C<sub>4</sub> and C<sub>3</sub> branches throughout the tree, and it is a direct hypothesis that can be accepted or rejected. The specific branches and dates of the shifts identified with turboMEDUSA will surely vary as more complete phylogenies are developed, but we suspect that the BiSSE results and sister group comparisons will prove to be quite robust.

### The effect of C<sub>4</sub> photosynthesis is context dependent

Although our results indicate that C<sub>4</sub> photosynthesis increased diversification, its effect varies among lineages. Not every C<sub>4</sub> clade experienced higher rates of diversification, and when rate accelerations occurred, they were presumably long after the initial C<sub>4</sub> origin. The delay in C<sub>4</sub> grass diversification might be the result of dependence on the development of a series of other adaptations to dry, open landscapes before C<sub>4</sub> grasses became highly competitive. This corresponds to previous arguments made that C<sub>4</sub> photosynthesis is itself only one component in suites of characters that confer ecological success or dominance [2]. Perhaps in C<sub>4</sub> lineages, the right combination of traits for rapid diversification did not emerge until after considerable time had passed since the origin. Alternatively, it is well known that the C<sub>4</sub> advantage is highly context dependent [1,72], and C<sub>4</sub> grasses might not have had the opportunity to diversify before open systems expanded in the Miocene.

In addition to the PACMAD C<sub>4</sub> radiations, we identified a series of concurrent non-C<sub>4</sub> radiations in the BEP clade. These occurred mainly within the Pooideae, particularly in lineages with well-established cold climate tolerance and temperate zone diversity. Interestingly, none of the C<sub>3</sub> accelerations occurred in

lineages that occupy the same climate space as C<sub>4</sub> grasses. At higher latitudes, the drought and cold tolerant Pooideae could have conceivably exploited similarly expanding open biomes, but under a cooler climatic regime where C<sub>4</sub> photosynthesis is not adaptive (but see Still et al. 2014 for a new perspective on temperature tolerances of Poooid grasses [73]). This indicates that while not all diverse grass lineages are C<sub>4</sub>, it is primarily C<sub>4</sub> clades that are able to take advantage of the warm tropical, open biomes.

### Concluding thoughts

In studies of lineage diversification and character traits, the case of C<sub>4</sub> photosynthesis is exceptional. There is an unusually high number (>66) of origins, providing a rare opportunity to test replicates and to identify the C<sub>4</sub> effect independent of other clade-specific adaptations [74,4]. Our sister group comparisons provide the best possible control for aspects of evolutionary history in 12 cases, and the complementary BiSSE analysis integrates over all of the origins and all of the heterogeneity in the phylogeny. The evolutionary history of grasses is inextricably linked with climatic and ecosystem changes throughout the Miocene that resulted in the global rise of grasslands [75,2]. The evolution of C<sub>4</sub> photosynthesis has long been recognized as an essential element of the ecological success of grasses in warm, open regions [1–3], and here we present compelling evidence that the C<sub>4</sub> pathway has also behaved as a “key innovation”, promoting elevated rates of lineage diversification during the assembly of the world’s grassy biomes.



## Supporting Information

**Figure S1 Histograms of BiSSE model inferences A–F.** histograms from BiSSE analyses run on trees under dating hypothesis 1 (macrofossil), G.–N. are histograms from BiSSE analyses run on trees under dating hypothesis 2 (phylolith). All are based on the results from 100 replicated phylogenies with the missing species richness distributed either proportionally (sampling frequency) or as unresolved clades. A.–B. PACMAD, unresolved clades; C.–D. Poaceae sampling frequency; E.–F. Poaceae unresolved clades; G.–H. PACMAD sampling frequency; I.–J. PACMAD unresolved clades; K.–L. Poaceae sampling frequency; M.–N. Poaceae unresolved clades.  
(PDF)

**Figure S2 LTT time plot, using the 3,595 taxon tree.** Showing the accumulation of a) all Poaceae species (black), b) C<sub>4</sub> species (blue), and c) C<sub>3</sub> species through time (green).  
(PDF)

**Table S1 The proportion of species represented by molecular data in our phylogeny.**  
(DOC)

**Table S2 The number of taxa and genes used in phylogenetic analyses for each subclade.**  
(DOC)

**Table S3 Maximum credibility tree results for each set of BiSSE model comparisons.** Bold indicates the preferred model(s).  
(DOC)

**Table S4 C<sub>4</sub> taxa with entirely C<sub>3</sub> sister clades.** Reconstructions based on character reconstructions that assume

the irreversibility of C<sub>4</sub> and are consistent with previous studies [10]. Bold indicates that a C<sub>4</sub> clade is at least as diverse as its C<sub>3</sub> sister.

(DOC)

**Table S5 Rate shifts inferred by turboMEDUSA with a pruned-to-genus tree (553 tips).** Shifts inferred with an AIC threshold of 10.5. The shift number does not correspond to the order of the shifts, but instead match Figure 3 and Figure 4. Bold indicates an acceleration in diversification in a C<sub>4</sub> lineage. Asterisks mark clades that were picked up in the full tree analysis as well (Table S6).

(DOC)

**Table S6 Rate shifts inferred by turboMEDUSA with the 3595-tip tree. Shifts inferred with an AIC threshold of 17.** The shift number does not correspond to the order of the shifts. Bold indicates an acceleration in diversification in a C<sub>4</sub> lineage.

(DOC)

## Acknowledgments

We thank NESCent, the Grass Phylogeny Working Group II, Jurrian deVos, Michael Donoghue, David Chatelet, Matt Ogburn, Monica Arakaki, Radika Bhaskar, and Laura Garrison for input and discussion.

## Author Contributions

Conceived and designed the experiments: ELS PAC EJE. Performed the experiments: ELS PAC. Analyzed the data: ELS PAC. Wrote the paper: ELS PAC EJE.

## References

- Ehleringer JR, Cerling TE, Helliker BR (1997) C<sub>4</sub> photosynthesis, atmospheric CO<sub>2</sub> and climate. *Oecologia* 112: 285–299.
- Edwards EJ, Osborne CP, Strömberg CAE, Smith SA (2010) C<sub>4</sub> Grasses Consortium. The origins of C<sub>4</sub> grasslands: integrating evolutionary and ecosystem science. *Science* 328: 587–591.
- Still CJ, Berry JA, Collatz GJ, DeFries RS (2003) Global distribution of C<sub>3</sub> and C<sub>4</sub> vegetation: carbon cycle implication. *Global Biogeochemical Cycles* 17: 1006–1021.
- Sage RF, Sage TL, Kocacinar F (2012) Photorespiration and the evolution of C<sub>4</sub> photosynthesis. *Annu. Rev. Plant Biol.* 63: 19–47.
- Clark LG, Zhang WP, Wendel JF (1995) A phylogeny of the grass family (Poaceae) based on *ndhF* sequence data. *Syst. Biol.* 20(4): 436–460.
- Grass Phylogeny Working Group (2001) Phylogeny and subfamilial classification of the grasses (Poaceae). *Ann. Mo. Bot. Gard.* 88: 373–457.
- Vicentini A, Barber JC, Alisconi SS, Giussani LM, Kellogg EA (2008) The age of the grasses and clusters of origins of C<sub>4</sub> photosynthesis. *Glob. Change Biol.* 14: 2963–2977.
- Christin PA, Besnard G, Samaritani E, Duvall MR, Hodkinson TR, et al. (2008) Oligocene CO<sub>2</sub> decline promoted C<sub>4</sub> photosynthesis in grasses. *Current Biology* 18: 37–43.
- Edwards EJ, Smith SA (2010) Phylogenetic analyses reveal the shady history of C<sub>4</sub> grasses. *PNAS* 107(6): 2532–2537.
- Grass Phylogeny Working Group II (GPWGII) (2012) New grass phylogeny resolves deep evolutionary relationships and discovers C<sub>4</sub> origins. *New Phytol.* 193(2): 304–12.
- Christin PA, Osborne CP, Chatelet DS, Columbus TJ, Besnard G, et al. (2013) Anatomic enablers and the evolution of C<sub>4</sub> photosynthesis in grasses. *PNAS* 101(4): 1381–1386.
- Catalan P, Kellogg EA, Olmstead RG (1997) Phylogeny of Poaceae subfamily Pooideae based on chloroplast *ndhF* gene sequences. *Mol. Phylogenet. Evol.* 8(2): 150–166.
- Catalan P, Torrecilla P, Rodriguez JAL, Olmstead RG (2004) Phylogeny of the festucoid grasses of subtribe Loliinae and allies (Poeae, Pooideae) inferred from *ITS* and *trnL-F* sequences. *Mol. Phylogenet. Evol.* 31(2): 517–541.
- Galley C, Linder HP (2007) The phylogeny of the pentaschistis clade (Danthonioideae, Poaceae) based on chloroplast DNA, and the evolution and loss of complex characters. *Evolution* 61(4): 864–884.
- Jacobs S, Bayer R, Everett J, Arriaga M, Barkworth M, et al. (2007) Systematics of the tribe Stipeae (Gramineae) using molecular data. *Aliso* 23: 349–361.
- Quintanar A, Castroviejo S, Catalan P (2007) Phylogeny of the tribe Aveneae (Pooideae, Poaceae) inferred from plastid *trnT-F* and nuclear *ITS* sequences. *Am. J. Bot.* 94(9): 1554–1569.
- Gillespie IJ, Soreng RJ, Bull RD, Jacobs SWL, Refulio-Rodriguez NF (2008) Phylogenetic relationships in subtribe Poinae (Poaceae, Poaeae) based on nuclear *ITS* and plastid *trnT-trnL-trnF* sequences. *Botany Botanique* 86(8): 938–967.
- Sungkaew S, Stapleton C, Salamin N, Hodkinson T (2009) Non-monophyly of the woody bamboos (Bambuseae; Poaceae): a multi-gene region phylogenetic analysis of Bambusoideae s.s. *J. Plant Res.* 122(1): 95–108.
- Cerros T, Columbus JT, Barker NP (2011) Phylogenetic relationships of Aristida and relatives (Poaceae, Aristidoideae) based on non-coding chloroplast (*trnL-F*, *rpl16*) and nuclear (*ITS*) DNA sequences. *Am. J. Bot.* 98: 1868–1886.
- Morrone O, Aagesen L, Scataglieni MA, Salariato DL, Denham SS (2012) Phylogeny of the Paniceae (Poaceae: Panicoideae): integrating plastid DNA sequences and morphology into a new classification. *Cladistics* 28(4): 333–356.
- Salamin N, Davies TJ (2004) Using supertrees to investigate species richness in grasses and flowering plants. In Bininda-Emonds, O.R.P (ed), *Phylogenetic supertrees*.
- Bouchenak-Khelladi Y, Verboom GA, Hodkinson TR, Salamin N, Francois O (2009) The origins and diversification of C<sub>4</sub> grasses and savanna-adapted ungulates. *Glob. Change Biol.* 15(10): 2397–2417.
- Clayton WD, Vorontsova MS, Harman KT, Williamson H. Grassbase (2006) The Online World Grass Flora. <http://www.kew.org/data/grasses-db.html>. Accessed 01 Jun 2011.
- Christin PA, Spriggs E, Osborne CP, Strömberg CAE, Salamin N, et al. (2014) Molecular dating, evolutionary rates, and the ages of the grasses. *Syst. Biol.* 63: 153–165.
- Smith SA, Beaulieu J, Donoghue MJ (2009) Mega-phylogeny approach for comparative biology: an alternative to supertree and supermatrix approaches. *BMC Evol. Biol.* 9: 37.
- Capella-Gutierrez S, Silla-Martinez JM, Gabalon T (2009) TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25: 1972–1973.
- Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24: 1596–1599.
- Smith SA, Dunn C (2008) Phytulity: a phyloinformatics tool for trees, alignments, and molecular data. *Bioinformatics* 24(5): 715–716.
- Missouri Botanical Garden. Tropicos. <http://www.tropicos.org>. [accessed June 2012].

30. Stamatakis A (2006) RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22(21): 2688–2690.
31. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7: 214.
32. Prasad V, Strömberg CAE, Leaché AD, Samant B, Patnaik R, et al. (2011) Late Cretaceous origin of the rice tribe provides evidence for early diversification in Poaceae. *Nature Commun.* 2: 480.
33. Sanderson MJ (2002) R8s; inferring absolute rates of evolution and divergence times in the absence of a molecular clock. *Bioinformatics* 19: 301–302.
34. Maddison WP, Midford PE, Otto SP (2007) Estimating a binary character's effect on speciation and extinction. *Syst. Biol.* 56: 701–710.
35. Fitzjohn RG, Maddison WP, Otto SP (2009) Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. *Syst. Biol.* 58: 595–611.
36. Vamوسي SM, Vamوسي JC (2005) Endless tests: guidelines for analyzing non-nested sister-group comparisons. *Evol. Ecol. Res.* 7: 567–579.
37. Alfaro ME, Santini F, Brock C, Alamillo H, Dornburg A (2009) Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *PNAS* 106: 13410–13414.
38. Aliscioni SS, Giussani LM, Zuloaga FO, Kellogg EA (2003) A molecular phylogeny of *Panicum* (Poaceae: Paniceae): tests of monophyly and phylogenetic placement within the Panicoideae. *Am. J. Bot.* 90: 796–821.
39. Morrone O, Zuloaga FO, Davidse G, Filgueiras TS (2001) *Canastra*, a new genus of Paniceae (Poaceae, Panicoideae) segregated from *Arthropogon*. *Novon.* 11: 429–436.
40. Morrone O, Scatagliini AM, Zuloaga FO (2007) *Cyphonanthus*, a new genus segregated from *Panicum* (Poaceae: Panicoideae: Paniceae) based on morphological, anatomical, and molecular data. *Taxon* 56: 521–532.
41. Morrone O, Denham SS, Aliscioni SS, Zuloaga FO (2008) *Parodiophylloclaoa*, a new genus segregated from *Panicum* (Paniceae, Poaceae) based on morphological and molecular data. *Syst. Bot.* 33: 66–76.
42. Bess EC, Doust AN, Davidse G, Kellogg EA (2006) *Zuloagaea*, a new genus of neotropical grass within the “bristle clade” (Poaceae: Poaceae). *Syst. Bot.* 31: 656–670.
43. Sede SM, Zuloaga FO, Morrone O (2009) Phylogenetic studies in the Paniceae (Poaceae-Panicoideae): *Ocellochloa*, a new genus from the New World. *Syst. Bot.* 34: 684–692.
44. Sede SM, Morrone O, Aliscioni SS, Giussani LM, Zuloaga FO (2009) *Oncorachis* and *Sclerochlamys*, two new segregated genera from the *Streptostachys* (Poaceae, Panicoideae, Paniceae): a revision based on molecular, morphological and anatomical characters. *Taxon* 58: 365–374.
45. Donoghue MJ (2005) Key innovations, convergence, and success: macroevolutionary lessons from plant phylogeny. *Paleobiology* 31: 77–93.
46. Nel A, Roques P, Nel P, Prokop J, Steyer JS (2007) The earliest holometabolous insect from the Carboniferous: a “crucial” innovation with delayed success (Insect Protomeropina Protomeropidae). *Ann. Soc. Entomol. Fr.* 43(3): 349–355.
47. Bininda-Emonds ORP, Cardillo M, Jones KE, MacPhee RDE, Beck RMD (2008) The delayed rise of present-day mammals. *Nature* 446: 507–512.
48. Marazzi B, Sanderson MJ (2010) Large-scale patterns of diversification in the widespread legume genus *Senna* and the evolutionary role of extrafloral nectarines. *Evolution* 64(12): 3570–3592.
49. Labandeira CC (2011) Evidence for an earliest late Carboniferous divergence time and the early larval ecology and diversification of major Holometabola lineages. *Entomol. Am.* 117(1/2):9–21.
50. Magallón S, Sanderson MJ (2001) Absolute diversification in Angiosperm clades. *Evolution* 55(9): 1762–1780.
51. Valente LM, Savolainen V, Vargas P (2010) Unparalleled rates of species diversification in Europe. *P.R.Soc.B.* 277: 1489–1496.
52. Mason-Gamer RJ (2004) Reticulate evolution, introgression, and intertribal gene capture in an allohexaploid grass. *Syst. Biol.* 63(1): 25–37.
53. Pau S, Edwards EJ, Still CJ (2013) Improving our understanding of environmental controls on the distribution of C<sub>3</sub> and C<sub>4</sub> grasses. *Glob. Change Biol.* 19: 184–196.
54. Zhang YG, Pagani M, Liu Z, Bohaty SM, DeConto R (2013) A 40-million-year history of atmospheric CO<sub>2</sub>. *Phil. Trans. R. Soc. A* 371: 20130096.
55. Retallack GJ (1997) Neogene expansion of the North American prairie. *Palaios* 12: 380–390.
56. Jacobs BF, Kingston JD, Jacobs LL (1999) The origin of grass-dominated ecosystems. *Ann. Missouri Bot. Gard.* 86: 590–643.
57. Kay RF (1999) Revised geochronology of the Casamayoran South American Land Mammal Age: Climatic and biotic implications. *Proc. Natl. Acad. Sci. U.S.A.* 96: 13235–13240.
58. Retallack GJ (2001) Cenozoic expansion of grasslands and climatic cooling. *Geol.* 109: 407–426.
59. Strömberg CAE (2004) Using phytolith assemblages to reconstruct the origin and spread of grass-dominated habitats in the great plains of North America during the late Eocene to early Miocene. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 207: 239–275.
60. Strömberg CAE (2005) Decoupled taxonomic radiation and ecological expansion of open-habitat grasses in the Cenozoic of North America. *Proc. Natl. Acad. Sci. U.S.A.* 102: 11980–11984.
61. Strömberg CAE, Werdelin L, Friis EM, Saraç G (2007) The spread of grass-dominated habitats in Turkey and surrounding areas during the Cenozoic: phytolith evidence. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 250: 18–49.
62. Jiang H, Ding Z (2009) Spatial and temporal characteristics of Neogene palynoflora in China and its implication for the spread of steppe vegetation. *Journal of Arid Environments* 73: 765–772.
63. Palazzesi L, Barreda V (2012) Fossil pollen records reveal a late rise of open-habitat ecosystems in Patagonia. *Nature Communications.* 3: 1294.
64. Osborne CP, Freckleton RP (2009) Ecological selection pressures for C<sub>4</sub> photosynthesis in grasses. *Proc. Roy. Soc. B.* 276(1663): 1753–1760.
65. Quade J, Soloumias N, Cerling TE (1994) Stable isotopic evidence from paleosol carbonates and fossil teeth in Greece for forest or woodlands over the past 11 Ma. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 108: 41–53.
66. Quade J, Cerling TE, Andrews P, Alpagut B (1995) Paleodietary reconstruction of Miocene faunas from Pasalar, Turkey using stable carbon and oxygen isotopes of fossil tooth enamel. *J. Hum. Evol.* 28: 373–384.
67. Cerling TE (1997) Global vegetation change through the Miocene/Pliocene boundary. *Nature* 398: 153–158.
68. Passey BH (2002) Environmental change in the Great Plains: An isotopic record from fossil horses. *J. of Geol.* 110(2): 123–140.
69. Fox DL, Koch PL (2003) Tertiary history of C<sub>4</sub> biomass in the Great Plains, USA. *Geology* 31(9): 809–812.
70. Passey BH (2009) Strengthened East Asian summer monsoons during a period of high-latitude warmth? Isotopic evidence from Mio-Pliocene fossil mammals and soil carbonates from Northern China. *Earth Planet. Sci. Lett.* 277: 443–452.
71. Strömberg CAE, Dunn RE, Madden RH, Kohn MJ, Carlini AA (2013) Decoupling the spread of grasslands from the evolution of grazer-type herbivores in South America. *Nature Communications* 4:1478.
72. De Queiroz A (2002) Contingent predictability in evolution: key traits and diversification. *Syst. Biol.* 51(6): 917–929.
73. Still CJ, Pau S, Edwards EJ (2014) Land surface skin temperature captures thermal environments of C<sub>3</sub> and C<sub>4</sub> grasses. *Global Ecol. Biogeogr.* 23: 286–296.
74. Sage RF, Christin PA, Edwards EJ (2011) The C<sub>4</sub> plant lineages of the planet Earth. *J. Exp. Bot.* 62(9): 3155–3169.
75. Osborne CP (2008) Atmosphere, ecology and evolution: what drove the Miocene expansion of C<sub>4</sub> grasslands? *J. Ecol.* 96: 35–45.