Global Change Biology

Global Change Biology (2013) **19**, 184–196, doi: 10.1111/gcb.12037

# Improving our understanding of environmental controls on the distribution of $C_3$ and $C_4$ grasses

STEPHANIE PAU*, ERIKA J. EDWARDS† and CHRISTOPHER J. STILL‡§

*National Center for Ecological Analysis and Synthesis (NCEAS), 735 State Street, Suite 300, Santa Barbara, CA 93101, USA, †Department of Ecology and Evolutionary Biology, Brown University, Providence, RI 02912, USA, ‡Department of Geography, University of California, Santa Barbara, CA 93106-4060, USA, §Forest Ecosystems and Society, Oregon State University, 321 Richardson Hall, Corvallis, OR 97331-5752, USA

## Abstract

A number of studies have demonstrated the ecological sorting of $C_3$ and $C_4$ grasses along temperature and moisture gradients. However, previous studies of $C_3$ and $C_4$ grass biogeography have often inadvertently compared species in different and relatively unrelated lineages, which are associated with different environmental settings and distinct adaptive traits. Such confounded comparisons of $C_3$ and $C_4$ grasses may bias our understanding of ecological sorting imposed strictly by photosynthetic pathway. Here, we used MaxEnt species distribution modeling in combination with satellite data to understand the functional diversity of $C_3$ and $C_4$ grasses by comparing both large clades and closely related sister taxa. Similar to previous work, we found that $C_4$ grasses showed a preference for regions with higher temperatures and lower precipitation compared with grasses using the $C_3$ pathway. However, air temperature differences were smaller (2 °C vs. 4 °C) and precipitation and % tree cover differences were larger (1783 mm vs. 755 mm, 21.3% vs. 7.7%, respectively) when comparing $C_3$ and $C_4$ grasses within the same clade vs. comparing all $C_4$ and all $C_3$ grasses (i.e., ignoring phylogenetic structure). These results were due to important differences in the environmental preferences of $C_3$ BEP and PACMAD clades (the two main grass clades). Winter precipitation was found to be more important for understanding the distribution and environmental niche of $C_3$ PACMADs in comparison with both $C_3$ BEPs and $C_4$ taxa, for which temperature was much more important. Results comparing closely related $C_3$–$C_4$ sister taxa supported the patterns derived from our modeling of the larger clade groupings. Our findings, which are novel in comparing the distribution and niches of clades, demonstrate that the evolutionary history of taxa is important for understanding the functional diversity of $C_3$ and $C_4$ grasses, and should have implications for how grasslands will respond to global change.

*Keywords:* BEP, *Echinochloa*, land surface temperature, niche conservatism, *Oplismenus*, PACMAD, temperature crossover model

*Received 27 April 2012 and accepted 5 September 2012*

## Introduction

$C_3$ and $C_4$ grasses are two fundamental plant functional types (PFTs) that play important and distinct roles in ecosystem functions such as global terrestrial productivity and water cycling. Although $C_4$ plants account for 20–25% of global terrestrial productivity (Still *et al.*, 2003a), large uncertainties remain regarding their response to climate variability and future global change. Part of this uncertainty stems from our ability to characterize the environmental, ecological, and evolutionary controls on $C_3$ and $C_4$ grass distributions.

Numerous studies have demonstrated the ecological sorting of $C_3$ and $C_4$ grasses along spatial gradients, particularly temperature gradients (reviewed in Sage & Monson, 1999). The prevailing explanation for temperature controls on $C_3$-$C_4$ distributions is the biochemically based temperature crossover model (Ehleringer *et al.*, 1997), which predicts that $C_4$ plants will be more competitive and thus more abundant in grassland regions where the mean monthly air temperature exceeds 22 °C (Collatz *et al.*, 1998; Still *et al.*, 2003a). Although the temperature crossover model is a powerful approach for predicting regions and climates where $C_4$ grasses should be dominant over $C_3$ grasses, it has several drawbacks. For example, certain regions where $C_4$ grasses are known to be dominant, such as the cooler high plateaus of South Africa, are not captured by this approach because they do not experience mean monthly temperatures in excess of 22 °C (O'Connor & Bredenkamp, 1997). Another limitation of the crossover temperature model is that $C_3$ and $C_4$ grasses are each assigned separate but uniform temperature thresholds; thus, the large functional and ecological diversity that exists in grasses is not captured.

Correspondence: Stephanie Pau, tel. 00+1 805 892 2500, fax 00+1 805 892 2510, e-mail: pau@nceas.ucsb.edu

There is long-standing observational evidence of $C_3$-$C_4$ partitioning along moisture gradients (Chazdon, 1978; Vogel *et al.*, 1986). Precipitation amount and drought tolerance should be important in $C_3$-$C_4$ distributions because physiological and biochemical differences related to photosynthetic pathway produce higher rates of carbon uptake per unit of transpired water (i.e., photosynthetic water-use efficiency) in $C_4$ grasses (Pearcy & Ehleringer, 1984). More importantly, the interaction between temperature and precipitation should be a key in understanding the distribution of these PFTs. Both seasonal and interannual variation in temperature and precipitation have been shown to be strongly related to $C_3$ and $C_4$ grass distributions, with $C_3$ grasses typically active during the cool-wet season and $C_4$ grasses active during the warm-dry season (Teeri & Stowe, 1976; Paruelo & Lauenroth, 1996; Tieszen *et al.*, 1997; Davidson & Csillag, 2003; Winslow *et al.*, 2003; Still *et al.*, 2003b; von Fischer *et al.*, 2008).

Despite well-established patterns along temperature and moisture gradients, research on $C_3$ and $C_4$ grass ecology can benefit from recent advances in comparative methods and in the development of well-resolved grass phylogenies (Edwards *et al.*, 2007; Edwards & Still, 2008; Edwards & Smith, 2010). Previous studies of $C_3$ and $C_4$ ecology have often inadvertently compared species in different and relatively unrelated lineages. Most $C_3$ grasses are found in two distinct lineages whose last common ancestor likely lived between 50 and 80 Mya (Vicentini *et al.*, 2008). Likewise, there are an estimated 22–24 independent origins of the $C_4$ pathway (Grass Phylogeny Working Group II, 2012). Consequently, there should be considerable diversity within $C_3$ and $C_4$ groups associated with different lineages, distinct environmental settings, and associated adaptive traits that confound our understanding of ecological differences imposed strictly by photosynthetic pathway (Edwards & Still, 2008; Edwards & Smith, 2010; Taylor *et al.*, 2010). In other words, photosynthetic pathway may be associated with other traits conserved through evolutionary history. For example, recent work has suggested that most grasses are warm-climate specialists, regardless of photosynthetic pathway, and a closer look at the environmental context associated with $C_4$ origins indicates that differences in habitat aridity and irradiance may have played a larger role than temperature in $C_4$ evolution (Edwards & Smith, 2010). Similarly, experimental evidence has demonstrated greater water-use efficiency in grasses with the $C_4$ pathway compared with closely related $C_3$ sister taxa (Taylor *et al.*, 2010). Thus, we hypothesize that temperature differences in $C_3$ and $C_4$ grasses may not be as important as differences in water-use efficiency or irradiance. By comparing the environmental preferences of $C_3$ and $C_4$ grasses within and among evolutionary lineages, we can refine our understanding of differences due principally to photosynthetic pathway as well as highlight functional diversity across lineages. Phylogenetically structured comparisons should improve our ability to predict the distribution, abundance, and ecological success of $C_4$ photosynthesis and associated ecosystem functions in response to future climate change.

Here, we investigate the ecological setting and environmental niche of $C_3$ and $C_4$ grasses in a phylogenetic context – comparing clades and sister taxa – to provide novel information on the distribution and functional aspects of these grasses. To explore how a variety of environmental controls – beyond air temperature – define $C_3$ and $C_4$ niches and distributions, we utilized a species distribution model (SDM) approach (this is also known as ecological niche modeling – see Franklin, 2009). SDMs utilize environmental information gathered from geo-referenced collection localities to understand the environmental correlates of species as an indication of their ecological requirements and to model their geographic distributions (Franklin, 2009). We used this approach to quantify and assess the abiotic factors that control $C_3$ and $C_4$ distributions (the fundamental niche) based on species' localities that incorporate biotic limiting factors (the realized niche) (see Franklin, 2009 for a discussion of niche concept in SDMs). Species distribution modeling is an increasingly popular methodology applied to problems in conservation biology, biogeography, ecology, and systematics (reviewed in Guisan & Thuiller, 2005). We extend the SDM approach to better understand the ecological and functional diversity of higher levels of taxonomic organization. Using niche modeling within a phylogenetic context to examine major PFTs and large clades (as opposed to individual species) is a novel application of this modeling approach (Huntley *et al.*, 2004; Guisan & Thuiller, 2005; Chapman & Purse, 2011; see Methods and Discussion for further discussion of modeling the aggregated niches of clades).

We also exploited several satellite-derived datasets to characterize the environmental niches of these grasses; such datasets should provide several advantages over station-based air temperature and moisture climatologies. This may be especially true over large regions of the tropics where meteorological stations are sparse and can lead to large interpolation errors (New *et al.*, 2002; Daly, 2006). Remotely sensed datasets are continuously observed and not modeled or interpolated, and they are often freely available. As demonstrated by Buermann *et al.* (2008), including remotely sensed data can improve model predictions for some species (see also Phillips *et al.*, 2006; Bradley & Fleishman, 2008;

Gillespie *et al.*, 2008; Bisrat *et al.*, 2011). One satellite dataset we use is a measure of the surface skin temperature, a quantity that has rarely been used for SDMs despite its obvious linkage to surface microclimates, and we demonstrate its potential importance here.

The Hawaiian Islands provide an ideal setting for this work because of their broad climatic gradients encompassed in a small geographic region. This allows us to capture the response of these grasses to a wide range of environmental variability within a discrete background setting. The majority of the Hawaiian grass flora (~200 species) is nonnative and has been introduced in the last 100–150 years (Wagner *et al.*, 1999). Thus, the present-day grass communities in Hawaii have assembled relatively recently and grass distributions should be the result primarily of ecological sorting, and not a result of insular biogeographical processes that would limit the generality of our results to other regions (Edwards & Still, 2008).

## Materials and methods

### *Species and environmental data*

We used 3595 digitized herbarium collections for Poaceae across the main seven Hawaiian Islands (Kauai, Oahu, Molokai, Lanai, Maui, Kahoolawe, and Hawaii), representing 152 species and over a century of collections from the Smithsonian Flora of the Hawaiian Islands Website and The Bishop Museum Herbarium (details on this dataset are provided in Edwards & Still, 2008). These specimens were assigned to either the 'BEP' or 'PACMAD' clade (most species of Poaceae belong to one or the other clade). 'PACMAD' is an acronym for Panicoideae, Aristidoideae, Chloridoideae, Micrairioideae, Arundinoidae, and Danthonioideae lineages. Our BEP occurrences refer largely to the Pooideae lineage because Bambusoideae (bamboos) and Ehrhartoideae, which make up the majority of BEP aside from Pooideae, were not well represented in the herbaria collections. We used the MaxEnt SDM (Phillips *et al.*, 2006), described below, to better understand key environmental and ecological differences between: (1) all $C_3$ taxa (combining both BEP and PACMAD clades) and $C_4$ taxa, (2) $C_3$ PACMADs and $C_3$ BEPs, and (3) closely related sister taxa in the PACMAD clade that differ principally in photosynthetic pathway. For sister taxa, we focused our efforts on the $C_3$ grass genus *Oplismenus* (grouping *O. compositus* and *O. hirtellus*) and the $C_4$ grass genus *Echinochloa* (grouping *E. colona* and *E. crusgalli*). These three comparisons illustrate typical comparisons between $C_3$ and $C_4$ grasses irrespective of evolutionary lineage (case 1 above), the diversity among clades that are commonly grouped (case 2), and finally niche differences associated primarily with photosynthetic pathway in closely related sister taxa (case 3).

When describing the aggregate niche of a clade, either all species in a clade should be equally represented in the occurrence data or the dominance of certain species (i.e., their ecological abundance) may be considered because they contribute to a larger proportion of the niche space occupied by the clade. When using herbaria collections, widely collected species may indicate ecological abundance (in our case, the exceptions are bamboos and *Ehrhata*, which could be considered common, but are not widely collected – C. Imada, personal communication) so that equal sampling may not be necessary. Nonetheless, we further examined the niche of each group ($C_3$ BEP, $C_3$ PACMAD, and $C_4$ PACMAD) by randomly subsampling an even number of occurrences from each species in the group (see Supporting Information for details).

Remote sensing data were obtained from a variety of satellite-based sensors. We used the NASA Shuttle Radar Topography Mission (SRTM) 90 m resolution (C-band) digital elevation model (DEM) obtained from the Consortium for Spatial Information (http://srtm.csi.cgiar.org) version 4.1 for our elevation layer. Elevation should in theory capture some interactions among temperature, moisture availability (associated with both the temperature lapse rate and orographic precipitation that primarily affects mid to high elevations), solar radiation (which is high at elevations above the cloud inversion layer), and also variations in $CO_2$ partial pressure (which decreases with increasing altitude). However, because elevation is considered an indirect gradient that has no direct physiological basis for predicting plant distributions (Austin & Smith, 1989), but is nonetheless an effective proxy variable that captures many processes, we present results of variable importance both with and without elevation as a predictor. As a proxy for both tree cover and habitat openness (i.e., high and low light grass environments), we used the Vegetation Continuous Fields (VCF) product (MOD44B, Collection 4, version 3) from the year 2005 (the most recent year available), which provides a continuous measure of percent canopy cover (hereafter referred to as '% tree cover'). The % tree cover product is derived from monthly composites of all seven reflectance bands in the NASA Moderate Resolution Imaging Spectroradiometer (MODIS) Surface Reflectance data at 500 m spatial resolution using global training data and phenological metrics in a regression tree (Hansen *et al.*, 2003). We also included the Terra MODIS Land Surface Temperature (LST) product (MOD11A2) Collection 5, which provides 8-day composite data at 1 km resolution. Daytime LST data were screened using the 'QC_Day' scientific dataset for only 'good quality' pixels (i.e., not contaminated by clouds or aerosols – Wan 2002). For each land pixel across the Hawaiian Islands, a 10-year average of daytime LST data, referred to hereafter as 'LST', was created using screened data from February 2000-February 2010. We also calculated the coefficient of variation (called 'LST CV') across the entire 8-day time series. Mean annual precipitation (MAP), mean monthly precipitation, mean annual air temperature (MAT), and mean monthly temperature climate grids were provided at 250 m resolution by T. Giambelluca (Giambelluca *et al.*, 1986). The SRTM, % tree cover, and LST data were resampled to a 250 m grid cell spatial resolution in ArcGIS using a nearest neighbor approach to match the majority of the environmental layers and to better capture the spatial scale of the plant collection localities. All data preprocessing was performed using ArcGIS v.9.3 (Environmental Systems Research Institute, Redlands, CA, USA) or IDL v.7.1 (Exelis Visual Information Solutions, Boulder, CO, USA).

### *Species distribution/ecological niche modeling*

We used the freely available software package known as MaxEnt, which is a statistical learning method based on the principles of maximum entropy modeling to predict habitat suitability for the organism(s) of interest (Phillips *et al.*, 2006; Phillips & Dudík, 2008). The approach is conservative by assuming a probability distribution of maximum entropy (closest to uniform) with the constraint that the expected value of each environmental feature (and their interactions) fits the empirical average for that feature estimated from the occurrence points (Elith *et al.*, 2011). MaxEnt was developed for use with presence-only species data, thus making it appropriate to use with herbarium records. Its performance has been demonstrated to be consistently high compared with other species distribution modeling methods (Elith *et al.*, 2006; Hernandez *et al.*, 2008). This technique also allows flexibility and varying degrees of detail compared with the more mechanistic, but categorical crossover temperature approach described above. For instance, with MaxEnt we can predict the present-day geographic distributions of plant functional types ($C_3$ vs. $C_4$ grasses), as well as for individual grass species and grass clades. Another advantage of MaxEnt is its ability to utilize a wide range of environmental layers for its predictions. In particular, MaxEnt's regularization method, which balances model fit and model complexity, can deal with a large number of correlated variables, although including variables that are known to be ecologically irrelevant is not recommended (see Elith *et al.*, 2011).

We evaluated model performance based on AUC (area under the curve of the receiver-operating characteristic plot) values. In presence-only models like MaxEnt, AUC values describe the probability that the model scores a random presence site higher than a random background site (Phillips *et al.*, 2006). In other words, AUC provides a measure of how well the model accurately predicts the probability of occurrence across a landscape. Twenty-five percent of randomly chosen sample records were set aside for testing the fit of the model, whereas the remaining sample records were used to train the model. We also examined jackknife tests of variable importance, which assess the relative importance of each environmental variable in estimating the species distribution. Each environmental variable is successively omitted and a model is created with the remaining variables to examine how much independent information that predictor contained. A model is also created with each variable in isolation to examine how useful that predictor alone is for estimating the species distribution. In general, patterns of variable importance based on jackknife tests were similar in their ranked contributions to training gain, test gain, and AUC gain, and thus we only show results from the contribution to training gain because it represents how well the model fits 75% of the data. MaxEnt also provides response curves showing how its prediction depends on each environmental variable used in isolation to account for potential strong correlations among and between environmental variables. These response curves indicate how each environmental variable may limit a taxon's niche. The MaxEnt prediction is a logistic probability of presence that is based on presence-only data; however, from here on we refer to the MaxEnt output as simply 'probability of presence'. This logistic output assumes that a value of '0.5' indicates 'typical' presence, i.e., a 50% chance of the taxa being present in the suitable areas (see Phillips & Dudík, 2008 and Elith *et al.*, 2011 for detailed discussion of the MaxEnt output).

We then quantified niche overlap and niche breadth of $C_3$ PACMAD, $C_3$ BEP, and $C_4$ grasses using ENMTools v.1.3 (Warren *et al.*, 2008). The degree of niche overlap was evaluated using common indices, *I*, *D*, and relative rank statistics, which all range from 0 (completely different niches) to 1 (identical niches) (Krebs, 1989). Whereas the *I* and D statistics examine the difference between habitat suitability scores for models of different taxa at each grid cell, the relative rank statistic considers the ranking of habitat suitability in each grid cell relative to other grid cells regardless of the absolute difference between models (see Warren *et al.*, 2008 for full discussion of these metrics). We quantified niche breadth using a common measure of 'inverse concentration' (Warren *et al.*, 2008; Levins, 1968; Krebs, 1989). To illustrate niche overlap, we created a divergence map of the MaxEnt models to highlight differences between $C_3$ members of the two major grass clades (PACMADs vs. BEPs). The MaxEnt typical default prevalence of 0.5 is not comparable across rare and common taxa because of the sampling effort required to obtain the presence data (in presence-absence data, sampling bias affects both presence and absence, therefore the effect should be canceled out – Elith *et al.*, 2011). Thus, we modified the default prevalence of each group using the proportion of the number of occurrences for each group divided by the total number of grass specimens from herbaria collections (the time period of collection has been similar for all taxa). $C_4$ grasses have been the most widely collected grass specimens (66.6% of collections), followed by $C_3$ BEPs (23.9%), and $C_3$ PACMADs (9.5%). This estimate of abundance assumes that collection intensity has been the same for each clade and that the number of herbaria specimens approximate ecological abundance (which is likely true for abundant species, whereas rare species are possibly over represented in herbaria collections, C. Imada, personal communication – see above).

Finally, we offer a comparison of MaxEnt model predictions against those made by the crossover temperature model (Ehleringer *et al.*, 1997; Collatz *et al.*, 1998; Still *et al.*, 2003a). This model is the dominant approach for predicting climates where $C_4$ grasses should have higher photosynthetic rates and thus should outcompete $C_3$ grasses, and it has been used to map the global distribution of $C_3$ and $C_4$ grasslands in the present (Collatz *et al.*, 1998; Still *et al.*, 2003a) and at smaller spatial scales in the geologic past (Fox & Koch, 2004; Osborne & Beerling, 2006). The crossover temperature approach also allows one to include the impacts of past and future atmospheric $CO_2$ variations on photosynthetic performance and thus potential $C_3$ and $C_4$ grass distributions (Collatz *et al.*, 1998). This is difficult with a species distribution model like MaxEnt, which at best implicitly includes a $CO_2$ effect on plant distributions. We created a temperature crossover prediction by summing the number of months that meet the $C_4$ climate criteria in our mean monthly climate layers (mean monthly temperature greater than or equal to 22 °C and mean monthly precipitation greater than or equal to 25 mm in that same month) for each pixel. We then assessed the frequency of $C_4$ grass occurrences in those pixels where the $C_4$ climate criteria were met by the crossover temperature model, and we also screened out pixels with > 50% tree cover (i.e., we restricted our analysis to open habitats).

## Results

### Distribution modeling

*All $C_3$ (BEP and PACMAD combined) vs. all $C_4$ taxa.* Both the $C_3$ and $C_4$ models performed moderately well (AUC for training data = 0.83 and 0.76, respectively). Jackknife tests of variable importance showed that winter precipitation and summer air temperature contributed the most to the $C_3$ model training gain when used in isolation, whereas % tree cover and January precipitation decreased gain the most when excluded from the model (Table 1). This combined impact on training gain implies that winter precipitation and summer temperatures were the most important variables for the prediction of $C_3$ distribution, but % tree cover and January precipitation contained the most unique information. When elevation was included in the SDM, elevation contributed the most to training gain and decreased training gain the most when excluded. In comparison, summer temperatures contributed the most to the training gain of the $C_4$ model and training gain decreased the most when December and January precipitation, and LST CV were excluded (Table 1). Thus, winter precipitation and the interannual variability in LST CV contained the most unique information. When elevation was included, it was the second highest contributor to training gain and decreased model gain the most when excluded.

Response curves showed that all $C_3$ taxa appeared to prefer MATs between roughly 10–20 °C, above which probability of presence declined rapidly (sample mean = 16.2 °C). By contrast, the $C_4$ response curve showed a steadily increasing probability of presence with increasing MAT until a sharp drop at about 23 °C (mean of all sample points = 20.2 °C). This sharp drop is due to the fact that almost no pixels in Hawaii experience MATs higher than 23 °C. For all $C_3$ taxa (PACMAD and BEP clades), probability of presence generally increased with increases in MAP (sample mean = 2229 mm), whereas it decreased for $C_4$ taxa (sample mean = 1474 mm). $C_3$ grasses occurred in regions with higher % tree cover (sample mean = 45.8%) and lower LST (sample mean = 24.2 °C) compared with $C_4$ grasses (38.1% and 25.8 °C, respectively). These diverging trends, with $C_4$ taxa preferring warmer, drier, and more open environments, and $C_3$ taxa preferring colder, wetter, and shadier environments, generally agree with previous findings that used different approaches to characterize climatic preferences of $C_3$ and $C_4$ grasses.

*C3 PACMAD vs. $C_3$ BEP.* The $C_3$ PACMAD model performed well (AUC for training data = 0.90), while model performance for the $C_3$ BEP model was slightly lower (AUC for training data = 0.86). These AUC values were slightly higher than for the all $C_3$ model, which grouped many more species and thus represented broader niches than smaller clade groupings. Jackknife tests of variable importance showed that winter precipitation and LST contributed the most to the $C_3$ PACMAD model training gain when these variables were used in isolation meaning that these were the most effective variables for predicting the distribution of PACMADs (Table 1). However, % tree cover decreased training gain the most when excluded from the model, demonstrating that % tree cover is a distinct variable containing information that is not included in the other variables. Interestingly, winter precipitation, which represents the wet season when the majority of precipitation falls, was found to be more important than summer precipitation for the $C_3$ PACMAD model; also, air temperature in any month contributed very little to the model training gain for this clade. In comparison to the $C_3$ PACMAD model, air temperature variables were the primary contributors to $C_3$ BEP model and model gain decreased the most when % tree cover and LST were excluded, again suggesting that % tree cover and LST contain the most unique information of the environmental variables (Table 1). When elevation was included in the model, it was the primary contributor to the $C_3$ BEP model gain, and model gain decreased the most when elevation was not included.

Response curves showed that the predicted probability of presence for $C_3$ PACMADs increased with air temperature up to about 16 °C, but declined rapidly above about 21 °C (sample mean = 18.2 °C) (Fig. 1a). In comparison, $C_3$ BEPs showed a preference for cooler regions (sample mean = 15.5 °C; Fig. 1d) compared with $C_3$ PACMADs with relatively high probability of presence at low MAT, which declined rapidly above 20 °C. Probability of presence for $C_3$ PACMADs increased with MAP (sample mean = 3257 mm; Fig. 1b). By contrast, $C_3$ BEPs exhibited a general preference for relatively drier regions (MAP sample mean = 1895 mm), although the MAP response curve for this group did show an increase above 4000 mm (Fig. 1e). For both groups, the high probability of presence in regions with very high precipitation (above 6000 mm $yr^{-1}$) is a result of the few background pixels (i.e., small land area) representing those environmental conditions – in other words there are few samples

**Table 1** Jackknife tests of variable importance for model training gain (i.e., how well the model fits 75% of the data), which assess the relative importance of each variable in estimating the species distribution model. Each variable is successively omitted ('training gain without variable') and also used in isolation ('training gain with single variable'). MAT = mean annual temperature; MAP = mean annual precipitation; LST = land surface temperature; LST CV = coefficient of variation of land surface temperature; months are referred to by three letter abbreviations. See Methods for details about environmental variables

| $C_3$ training gain without variable | | $C_3$ training gain with single variable | | $C_4$ training gain without variable | | $C_4$ training gain with single variable | |
|---|---|---|---|---|---|---|---|
| %Tree cover | 0.5610 | Feb precipitation | 0.1599 | Dec precipitation | 0.3454 | Jun temperature | 0.1534 |
| Jan precipitation | 0.5622 | May temperature | 0.1477 | Jan precipitation | 0.3466 | May temperature | 0.1522 |
| LST CV | 0.5808 | Dec precipitation | 0.1469 | LST CV | 0.3479 | Jul temperature | 0.1522 |
| Sep precipitation | 0.5832 | Nov temperature | 0.1368 | %Tree cover | 0.3507 | Aug temperature | 0.1517 |
| LST | 0.5834 | Jan precipitation | 0.1360 | Aug precipitation | 0.3509 | Sep temperature | 0.1498 |
| Jun precipitation | 0.5841 | Jun temperature | 0.1347 | LST | 0.3513 | MAT | 0.1446 |
| Feb precipitation | 0.5857 | Apr temperature | 0.1342 | Apr temperature | 0.3532 | Oct temperature | 0.1430 |
| Jul precipitation | 0.5874 | Jul temperature | 0.1342 | Sep precipitation | 0.3533 | Dec temperature | 0.1421 |
| Aug precipitation | 0.5877 | MAT | 0.1335 | Mar temperature | 0.3544 | Apr temperature | 0.1419 |
| Mar temperature | 0.5892 | Oct temperature | 0.1330 | Jun precipitation | 0.3546 | Mar temperature | 0.1409 |
| Feb temperature | 0.5899 | Jan temperature | 0.1304 | Nov precipitation | 0.3555 | Nov temperature | 0.1401 |
| Dec precipitation | 0.5900 | Mar temperature | 0.1294 | May precipitation | 0.3558 | Feb temperature | 0.1353 |
| Mar precipitation | 0.5909 | Dec temperature | 0.1288 | Dec temperature | 0.3558 | Jan temperature | 0.1338 |
| Oct temperature | 0.5914 | Aug temperature | 0.1281 | Oct precipitation | 0.3564 | Dec precipitation | 0.1006 |
| MAP | 0.5920 | Mar precipitation | 0.1275 | Jul precipitation | 0.3566 | Jan precipitation | 0.0702 |
| Aug temperature | 0.5922 | Apr precipitation | 0.1274 | May temperature | 0.3566 | Nov precipitation | 0.0601 |
| Oct precipitation | 0.5922 | Sep temperature | 0.1266 | Aug temperature | 0.3569 | Sep precipitation | 0.0600 |
| Jun temperature | 0.5926 | Feb temperature | 0.1224 | Nov temperature | 0.3571 | Aug precipitation | 0.0575 |
| May precipitation | 0.5926 | Nov precipitation | 0.1177 | Mar precipitation | 0.3573 | Jun precipitation | 0.0550 |
| Dec temperature | 0.5935 | LST | 0.1050 | Jan temperature | 0.3573 | Apr precipitation | 0.0540 |
| Apr precipitation | 0.5937 | MAP | 0.0994 | Sep temperature | 0.3577 | May precipitation | 0.0529 |
| Jul temperature | 0.5940 | Sep precipitation | 0.0976 | Feb temperature | 0.3579 | Feb precipitation | 0.0521 |
| May temperature | 0.5940 | Aug precipitation | 0.0967 | Oct temperature | 0.3581 | Mar precipitation | 0.0509 |
| Apr temperature | 0.5945 | Jul precipitation | 0.0955 | Apr precipitation | 0.3582 | Jul precipitation | 0.0494 |
| Nov temperature | 0.5952 | Oct precipitation | 0.0718 | Jun temperature | 0.3582 | Oct precipitation | 0.0431 |
| Jan temperature | 0.5955 | %Tree cover | 0.0717 | Feb precipitation | 0.3583 | LST | 0.0385 |
| Nov precipitation | 0.5956 | May precipitation | 0.0674 | MAT | 0.3584 | MAP | 0.0379 |
| Sep temperature | 0.5968 | Jun precipitation | 0.0639 | MAP | 0.3585 | LST CV | 0.0336 |
| MAT | 0.5985 | LST CV | 0.0069 | Jul temperature | 0.3588 | %Tree cover | 0.0134 |

| $C_3$ PACMAD training gain without variable | | $C_3$ PACMAD training gain with single variable | | $C_3$ BEP training gain without variable | | $C_3$ BEP training gain with single variable | |
|---|---|---|---|---|---|---|---|
| %Tree cover | 0.9338 | Jan precipitation | 0.5277 | %Tree cover | 0.7290 | Jan temperature | 0.2038 |
| LST | 0.9778 | Feb precipitation | 0.4956 | LST | 0.7331 | Nov temperature | 0.1993 |
| LST CV | 0.9882 | Dec precipitation | 0.4841 | Jan precipitation | 0.7379 | Feb temperature | 0.1970 |
| Jun precipitation | 1.0004 | MAP | 0.4734 | Sep precipitation | 0.7424 | Apr temperature | 0.1949 |
| Sep precipitation | 1.0018 | Nov precipitation | 0.4651 | Dec precipitation | 0.7452 | May temperature | 0.1948 |
| Jan precipitation | 1.0025 | LST | 0.4559 | Aug temperature | 0.7485 | Dec temperature | 0.1943 |
| Aug precipitation | 1.0046 | Oct precipitation | 0.4541 | Jun precipitation | 0.7493 | Mar temperature | 0.1926 |
| Dec precipitation | 1.0072 | Mar precipitation | 0.4387 | LST CV | 0.7494 | MAT | 0.1925 |
| Jul precipitation | 1.0090 | Apr precipitation | 0.4318 | Mar precipitation | 0.7496 | Oct temperature | 0.1834 |
| Nov precipitation | 1.0120 | Sep precipitation | 0.4110 | Apr precipitation | 0.7504 | Aug temperature | 0.1809 |
| May precipitation | 1.0129 | Jun precipitation | 0.4008 | Jun temperature | 0.7506 | Jun temperature | 0.1782 |
| Sep temperature | 1.0138 | May precipitation | 0.3961 | MAP | 0.7508 | Sep temperature | 0.1774 |
| Jan temperature | 1.0146 | %Tree cover | 0.3810 | Jul precipitation | 0.7511 | Jul temperature | 0.1733 |
| Mar precipitation | 1.0153 | Jul precipitation | 0.3665 | Feb precipitation | 0.7519 | Sep precipitation | 0.1282 |
| Oct temperature | 1.0160 | Aug precipitation | 0.3606 | Aug precipitation | 0.7520 | Jul precipitation | 0.1258 |
| May temperature | 1.0164 | May temperature | 0.1863 | Nov precipitation | 0.7523 | Aug precipitation | 0.1169 |
| Dec temperature | 1.0178 | Oct temperature | 0.1701 | Jan temperature | 0.7526 | Nov precipitation | 0.1074 |

**Table 1** (continued)

| $C_3$ PACMAD training gain without variable | | $C_3$ PACMAD training gain with single variable | | $C_3$ BEP training gain without variable | | $C_3$ BEP training gain with single variable | |
|---|---|---|---|---|---|---|---|
| Aug temperature | 1.0180 | LST CV | 0.1653 | Sep temperature | 0.7526 | Feb precipitation | 0.1073 |
| Oct precipitation | 1.0187 | Aug temperature | 0.1621 | MAT | 0.7530 | Dec precipitation | 0.1072 |
| Apr precipitation | 1.0188 | Jul temperature | 0.1596 | Dec temperature | 0.7532 | Apr precipitation | 0.0967 |
| Feb temperature | 1.0188 | Jun temperature | 0.1565 | May temperature | 0.7536 | Mar precipitation | 0.0916 |
| MAP | 1.0194 | Mar temperature | 0.1525 | Apr temperature | 0.7541 | Jun precipitation | 0.0808 |
| Feb precipitation | 1.0200 | Sep temperature | 0.1513 | May precipitation | 0.7547 | MAP | 0.0799 |
| Apr temperature | 1.0204 | Dec temperature | 0.1513 | Oct precipitation | 0.7551 | Jan precipitation | 0.0792 |
| Nov temperature | 1.0214 | MAT | 0.1496 | Oct temperature | 0.7554 | Oct precipitation | 0.0742 |
| Jun temperature | 1.0232 | Jan temperature | 0.1473 | Nov temperature | 0.7558 | May precipitation | 0.0716 |
| MAT | 1.0236 | Feb temperature | 0.1452 | Feb temperature | 0.7565 | LST | 0.0697 |
| Mar temperature | 1.0256 | Nov temperature | 0.1451 | Mar temperature | 0.7592 | %Tree cover | 0.0560 |
| Jul temperature | 1.0276 | Apr temperature | 0.1380 | Jul temperature | 0.7605 | LST CV | 0.0547 |

| *Oplismenus* training gain without variable | | *Oplismenus* training gain with single variable | | *Echinochloa* training gain without variable | | *Echinochloa* training gain with single variable | |
|---|---|---|---|---|---|---|---|
| %Tree cover | 0.9205 | %Tree cover | 0.3911 | %Tree cover | 1.2222 | May temperature | 0.8690 |
| LST | 0.9312 | May temperature | 0.3893 | Jan precipitation | 1.2608 | Jul temperature | 0.8612 |
| LST CV | 0.9419 | Jun temperature | 0.3712 | Aug precipitation | 1.2809 | Dec temperature | 0.8537 |
| Jun temperature | 0.9561 | Dec temperature | 0.3695 | Jun precipitation | 1.2870 | Oct temperature | 0.8303 |
| Jun precipitation | 0.9591 | Oct temperature | 0.3682 | Sep precipitation | 1.2876 | Sep temperature | 0.8292 |
| Sep precipitation | 0.9607 | Nov temperature | 0.3656 | Jul temperature | 1.2881 | Nov temperature | 0.8275 |
| Aug precipitation | 0.9649 | Aug temperature | 0.3652 | LST | 1.2902 | MAT | 0.8252 |
| Sep temperature | 0.9662 | Jul temperature | 0.3614 | Oct temperature | 1.2918 | Aug temperature | 0.8230 |
| Jan precipitation | 0.9694 | Sep temperature | 0.3541 | Mar precipitation | 1.2944 | Mar temperature | 0.8192 |
| Mar precipitation | 0.9697 | MAT | 0.3482 | LST CV | 1.2966 | Jun temperature | 0.8066 |
| Oct temperature | 0.9699 | Apr temperature | 0.3361 | May precipitation | 1.2969 | Apr temperature | 0.8022 |
| May precipitation | 0.9715 | Feb temperature | 0.3329 | Nov temperature | 1.2977 | Feb temperature | 0.7882 |
| Jan temperature | 0.9753 | Mar temperature | 0.3302 | May temperature | 1.2982 | Jan temperature | 0.7578 |
| Dec precipitation | 0.9762 | LST | 0.3090 | Dec temperature | 1.2984 | Jan precipitation | 0.2338 |
| Jul temperature | 0.9762 | Jan temperature | 0.2963 | Feb temperature | 1.2988 | Dec precipitation | 0.2311 |
| Apr precipitation | 0.9765 | LST CV | 0.2599 | Feb precipitation | 1.2990 | Sep precipitation | 0.2089 |
| Feb temperature | 0.9766 | Dec precipitation | 0.2429 | Jan temperature | 1.2990 | Aug precipitation | 0.1899 |
| Jul precipitation | 0.9767 | Jan precipitation | 0.2422 | Sep temperature | 1.2990 | %Tree cover | 0.1457 |
| Feb precipitation | 0.9770 | Nov precipitation | 0.2275 | Oct precipitation | 1.2992 | Jul precipitation | 0.1425 |
| MAP | 0.9772 | Oct precipitation | 0.2133 | Nov precipitation | 1.2992 | Feb precipitation | 0.1224 |
| Oct precipitation | 0.9772 | MAP | 0.1943 | Dec precipitation | 1.2992 | LST CV | 0.1150 |
| Nov precipitation | 0.9772 | Feb precipitation | 0.1914 | Jun temperature | 1.2992 | May precipitation | 0.1111 |
| Dec temperature | 0.9773 | Mar precipitation | 0.1866 | Aug temperature | 1.2992 | Jun precipitation | 0.1025 |
| May temperature | 0.9774 | Apr precipitation | 0.1646 | MAP | 1.2995 | Nov precipitation | 0.0916 |
| Mar temperature | 0.9776 | May precipitation | 0.1322 | Jul precipitation | 1.2995 | Mar precipitation | 0.0871 |
| Aug temperature | 0.9790 | Sep precipitation | 0.1307 | Apr precipitation | 1.2996 | MAP | 0.0712 |
| MAT | 0.9796 | Jul precipitation | 0.1189 | Mar temperature | 1.2996 | LST | 0.0689 |
| Apr temperature | 0.9796 | Jun precipitation | 0.1166 | Apr temperature | 1.2997 | Apr precipitation | 0.0615 |
| Nov temperature | 0.9798 | Aug precipitation | 0.0975 | MAT | 1.2999 | Oct precipitation | 0.0518 |

occurrences in regions of very high precipitation, but there are also few background pixels (pseudo absences used in presence-only SDM) from which the Maxent probability of presence is calculated. $C_3$ PACMADs occurred in regions with higher % tree cover (sample mean = 59.4%) and lower LST (sample mean = 22.1 °C) compared with $C_3$ BEPs, which occurred in more open regions with less tree cover (sample mean = 41.4%) and higher LST (sample mean = 24.8 °C).

*$C_3$ vs. $C_4$ sister taxa (Oplismenus vs. Echinochloa).* Model performance for both *Oplismenus* and *Echinochloa* was high (training AUC = 0.91 and 0.92, respectively), reflecting the smaller niche and geographic distribution

of these genera compared to the larger clades. The most important variable contributing to the training gain for *Oplismenus* was % tree cover, followed by air temperature variables in no particular month (Table 1). Model training gain decreased the most when % tree cover and LST were excluded. In comparison, air temperature variables by far contributed the most to *Echinochloa* training gain (with summer temperatures contributing slightly more) and decreased the most when % tree cover was excluded (Table 1).

The modeled probability of presence in response to MAT increased exponentially with higher temperatures for both *Oplismenus* and *Echinochloa* (Fig. 1j and m); however, *Oplismenus* began increasing at lower temperatures and reached a peak at a lower temperature (about 21 °C) compared with *Echinochloa* (which peaked at about 23 °C; sample means for *Oplismenus* and *Echinochloa* occurrences were 20.7 °C and 22.3 °C, respectively). Both genera declined sharply at air temperatures above 23 °C. *Oplismenus* showed a lower probability of presence in the driest regions, but response increased and peaked at about 2700 mm, at which point there was a gradual decline in probability of presence (Fig. 1k). In comparison, *Echinochloa* exhibited a high probability of presence at low MAP values and quickly declined at MAP values greater than 600 mm (Fig. 1n). There was a large difference in the % tree cover values of *Oplismenus*, which occurred in much more closed habitats (sample mean = 62.0%) compared with *Echinochloa*, which occurred in much more open habitats (sample mean = 31.3%). Relatedly, *Oplismenus* occurred in regions of lower LST (sample mean = 23.2 °C) compared with *Echinochloa* (sample mean = 25.9 °C), as would be expected due to the higher tree cover and transpiration of *Oplismenus* environments.

### *Niche overlap and niche breadth*

Niche overlap tests revealed that $C_4$ distributions overlap slightly more with $C_3$ BEP distributions than they do with $C_3$ PACMAD distributions ($I = 0.86$ vs. 0.82 and $D = 0.60$ vs. 0.56, respectively). This is interesting given that the $C_3$ and $C_4$ PACMAD taxa share a common evolutionary history compared with $C_4$ and $C_3$ BEP taxa, and instead indicate differences likely due to photosynthetic pathway. These differ from the relative rank statistic, which shows that $C_4$ distributions overlap more with $C_3$ PACMAD distributions than with $C_3$ BEPs (0.54 vs. 0.49, respectively). This is not surprising because the relative rank metric is sensitive to regions estimated to be of intermediate suitability for both $C_3$ PACMADs and $C_4$ taxa (Warren *et al.*, 2008), and $C_3$ PACMAD taxa occur in lower elevations geographically closer to $C_4$ taxa in comparison with $C_3$ BEP taxa (i.e., geographic space). The *I* and *D* metrics appear more sensitive to the similar preference of $C_4$ taxa and $C_3$ BEPs for low % tree cover, warmer LST, and drier regions (i.e., ecological space), which is also illustrated in the response curve differences discussed above. As illustrated by the divergence map between the $C_3$ PACMAD and $C_3$ BEP distributions, $C_3$ BEPs appear to occur in both warmer and drier leeward regions of islands as well as very high elevations, where $C_3$ PACMADs have a low probability of occurrence (Fig. 2). Niche breadth analysis revealed that $C_3$ PACMADs occupy the narrowest niche based on the inverse concentration metric (0.47), whereas $C_4$ taxa have the broadest niche (0.75), and $C_3$ BEPs were intermediate in niche breadth (0.60). In contrast, *Echinochloa* species ($C_4$) had a narrower niche (0.30) compared with *Oplismenus* species (0.42).

### *MaxEnt vs. crossover temperature model*

The simpler but physiologically based crossover temperature model arguably performs better than the MaxEnt model of $C_4$ distribution. The greatest number of $C_4$ occurrences corresponded to those grid cells that met the $C_4$ climate criteria for 12 months of the year (and were screened for tree cover <50%), and at least some occurrences were found in each $C_4$ climate-months bin (Fig. 3a). In comparison, the greatest number of $C_4$ occurrences corresponded to intermediate logistic probabilities of presence (between 0.4 and 0.7) from MaxEnt output (Fig. 3b). The crossover temperature model primarily captured abiotic limitations (although the crossover temperature of 22 °C is based on relative $C_3$-$C_4$ photosynthetic production), whereas the MaxEnt model used species localities that resulted from both abiotic and biotic interactions. Therefore, the crossover temperature model captured a larger proportion of the land area as suitable and more occurrences were likely to fall into these regions. Interestingly, the number of $C_4$ occurrences declined sharply at higher MaxEnt probabilities of presence (0.8–0.9). One explanation may be that the MaxEnt probability of presence is dependent upon the total area that is represented by different environmental conditions (i.e., background samples). For example, there were a large number of $C_4$ occurrences in low elevation regions, but these regions were also a large proportion of the total area being modeled, therefore, the MaxEnt predicted probability in these regions will be lower than in regions where the ratio of occurrences in a particular environment to the total area of that environmental condition is higher.

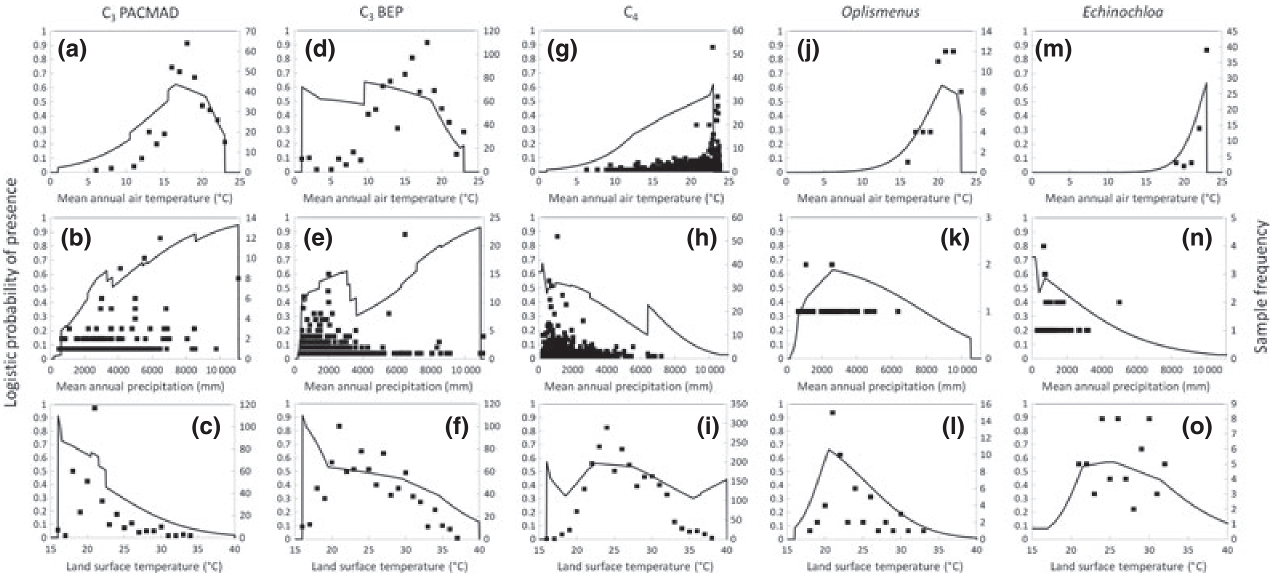

**Fig. 1** MaxEnt environmental response curves of probability of presence (solid line; left axis) with all other variables set to their average value for the occurrence points. Sample frequency (boxes; right axis) showing the number of occurrences along environmental gradients. The high probability of presence of $C_3$ PACMADs and BEPs in regions with very high precipitation (above 6000 mm yr$^{-1}$) is likely the result of few background pixels (i.e., small land area) representing those environmental conditions – there are few samples occurrences in regions of very high precipitation, but there are also few background pixels from which the MaxEnt probability of presence is calculated.

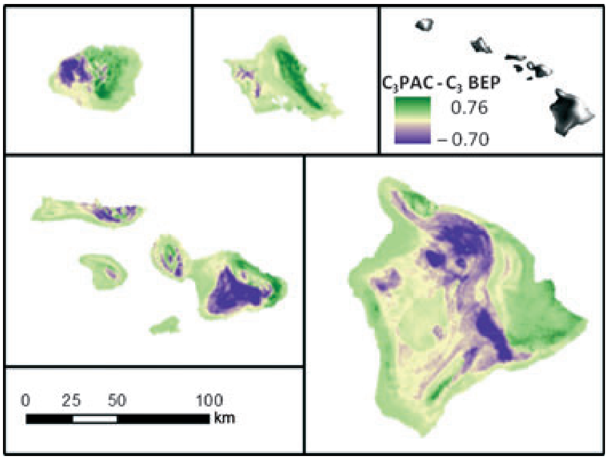

**Fig. 2** Divergence map showing regions of similarity and difference between $C_3$ PACMAD and $C_3$ BEP grass distributions ($C_3$ PACMAD model – $C_3$ BEP model) on the Hawaiian Islands based on MaxEnt models including elevation. Although we report results with and without elevation when evaluating variable importance (see Methods), the inclusion of elevation resulted in slightly better models based on AUC, and thus we include it for mapping purposes only. Positive values indicate greater PACMAD probability of presence, negative values indicate greater BEP probability of presence, and the light tan color indicates regions of similar probabilities. Inset (upper right) showing mean annual precipitation across islands with wettest regions in white and driest regions in black.

## Discussion

We expected that the predicted environmental niches and geographic distributions of $C_3$ and $C_4$ grasses would differ due to the differential temperature sensitivities and resource-use efficiencies shown for each plant type by many previous studies. As expected, on average $C_4$ taxa show a preference for regions with higher temperatures and lower precipitation compared with $C_3$ taxa in both the BEP and PACMAD clades (Fig. 1). However, consistent with our hypothesis, we found considerable differences in the environmental preferences and niches of $C_3$ BEPs and $C_3$ PACMADs (Table 1; Figs 1 and 2) and these differences were obscured when modeling $C_3$ grasses as a whole, confounding the ecological sorting due to $C_3$ and $C_4$ photosynthetic pathway. Indeed, our comparison between $C_3$ BEPs and $C_3$ PACMADs, which suggests many ecological differences unrelated to photosynthetic pathway, is not unexpected given that the last common ancestor of these clades lived over ~50 Mya (Vicentini *et al.*, 2008).

Our results show that $C_4$ (PACMAD) taxa are more ecologically distinct from $C_3$ PACMADs than from $C_3$ BEPs. The distinction between $C_4$ and $C_3$ PACMADs highlights differences in resource-use (e.g., light, water, and nitrogen) associated with photosynthetic pathway and these differences are supported in the comparison

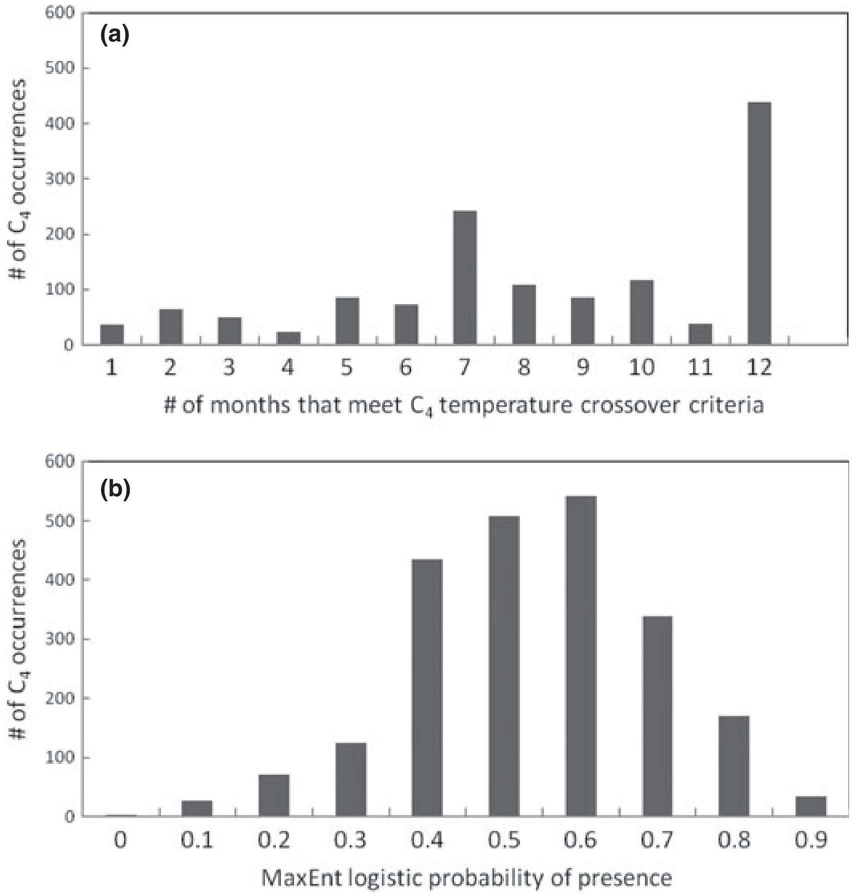

**Fig. 3** Frequency of $C_4$ grass occurrences in pixels that met the temperature crossover criteria (a) and across the MaxEnt probabilities of occurrence for a $C_4$ model (b) in the Hawaiian Islands. Number of months that met the $C_4$ temperature crossover criteria in mean monthly climate layers were summed (mean monthly temperature greater than or equal to 22 °C and mean monthly precipitation greater than or equal to 25 mm in that same month) for each pixel (pixels with> 50% tree cover were excluded).

among closely related $C_3$-$C_4$ sister taxa. Precipitation, in particular winter precipitation, was found to be more important for understanding the distribution of $C_3$ PACMADs in comparison to $C_3$ BEPs and $C_4$ taxa, for which temperature variables were much more important (Table 1). Because of the differences in the BEP and PACMAD clades, the importance of winter precipitation over temperature variables was muted in the jackknife results when modeling all $C_3$ grasses as a group, although winter precipitation still ranked as slightly more important than summer precipitation, and this is supported by previous work examining $C_3$-$C_4$ phenology (Paruelo & Lauenroth, 1996; Tieszen *et al.*, 1997; Davidson & Csillag, 2003; von Fischer *et al.*, 2008).

Consistent with results by Edwards & Still (2008) and Edwards & Smith (2010), air temperature differences were smaller (2 °C vs. 4 °C) and precipitation and % tree cover differences were larger (1783 mm vs. 755 mm, 21.3% vs. 7.7%, respectively) when comparing $C_3$ and $C_4$ grasses within a phylogenetic context (i.e., within the same PACMAD clade vs. all $C_3$). These results are relevant for studies of grassland response to future global change because they suggest that the distribution and abundance of $C_4$ grasses should be more sensitive to differences in precipitation and associated changes in tree cover, whereas the sensitivity to air temperature attributed to photosynthetic pathway may not be as large as previously thought.

The niche partitioning among $C_3$ PACMADs, $C_3$ BEPs, and $C_4$ taxa highlights the interactions among temperature, water availability, and tree cover in controlling their distributions. Our results clearly show that $C_3$ PACMADs prefer mid-elevation, closed canopy regions where there is greater precipitation but air temperatures are also warmer; in comparison, high-elevation regions where $C_3$ BEPs are more likely to occur are cold, dry, and open habitats. Thus, even in dry environments where $C_4$ species are thought have a competitive advantage due to their higher water-use efficiency (Ehleringer & Monson, 1993), temperature limitations inherent to the tropical origins of $C_4$ taxa may exclude these grasses at high elevation sites where $C_3$ BEPs exist.

Results comparing closely related $C_3$-$C_4$ sister taxa (in the PACMAD clade), *Oplismenus* and *Echinochloa*, support the patterns derived from our modeling of the larger clade groupings. Both taxa had lower elevational distributions resulting in higher probabilities of presence in drier regions compared with $C_3$ PACMADs as a whole; however, the temperature response profiles were similar to the larger clade responses (Fig. 1). It is noteworthy that the sister taxa comparison emphasized large differences in the light preferences of $C_3$ compared with $C_4$ grasses. *Oplismenus* preferred regions with greater tree cover, lower light levels, and also cooler air temperatures in contrast with *Echinochloa*, which preferred open habitats with lower tree cover and higher air temperature. These findings are consistent with the advantages typically conferred by the $C_4$ pathway in high light and high temperature environments and also agree well with Edwards & Smith (2010) and Osborne and Freckleton (2009) who showed that $C_3$ PACMAD taxa are preferentially growing in shady habitats associated with high tree cover compared with closely related $C_4$ sister taxa, which occur in drier and more open habitats.

Response curves, sample histograms, a higher AUC value, and the niche breadth analysis together show that $C_3$ PACMADs have a narrower environmental niche compared with $C_3$ BEPs and $C_4$ taxa. If organisms with a narrow niche breadth and smaller geographic range are most threatened by increases in temperature (Urban *et al*., 2012), then $C_3$ PACMADs are most likely to be impacted by climate change. On the other hand, $C_3$ BEPs may be at risk because temperatures on Hawaii at higher elevations are increasing more rapidly compared with lower elevations due to an increase in minimum temperatures and a compression of the diurnal temperature range (Giambelluca *et al*., 2008). Furthermore $C_3$ BEPs were most sensitive to temperature as shown by jackknife results. Although $C_3$ BEPs generally occurred in colder air temperature environments compared with $C_3$ PACMADs, both showed clear declines in predicted probability of presence at air temperatures above 21 °C. In this regard, both groups are distinct from $C_4$ taxa, which showed a modeled increase in presence with increasing air temperature.

Recent experimental work has demonstrated that the invasion success of a $C_4$ grass can be enhanced by increasing temperature and aridity (Chuine *et al*., 2012), suggesting that the invasion of $C_4$ grasses on Hawaii may be accelerated with increasing temperature and aridity throughout the islands. Our results suggest that the success of invasive $C_4$ grasses on Hawaii may benefit from distinct physiological advantages over $C_3$ PACMAD compared with $C_3$ BEP grasses, even in the face of increasing $CO_2$, which should favor $C_3$ grasses through $CO_2$ fertilization. It has, however, been suggested that $C_4$ plants may be less phenotypically plastic compared with $C_3$ plants and thus may be less capable of acclimating to environmental change (Sage & McKown, 2006). Our comparisons of *Oplismenus* and *Echinochloa* indicate that $C_4$ grasses may have a narrower niche, in contrast with the larger clade groupings, and thus may be less capable of responding to climate change by shifting in space or time. Future work investigating the niche diversity both within and across clades could begin to answer key questions regarding the rate of change in niche evolution and how much genetic variation is present within species and clades for adaptation to changing climate (Holt, 2009).

Although using SDMs to understand the environmental niche of clades can provide novel information on the evolutionary bounds of functional traits and related ecological processes, careful interpretation of results is warranted when there is not a balanced representation of each species in a clade. When using herbaria collections, widely collected species potentially overrepresent the environmental preferences of the entire clade. However, widely collected herbaria specimens can indicate ecological abundance (C. Imada, personal communication – see Methods). Thus, when considering the aggregate niche of a clade, dominant species represent a larger proportion of the niche space occupied by the clade. Furthermore, evenly sampling across all species limits the number of occurrences that can be used in the models. However, we further examined the niche of each group ($C_3$ BEP, $C_3$ PACMAD, and $C_4$ PACMAD) by randomly subsampling an even number of occurrences from each species in the group, and in our case model results were similar (see Supporting Information for details).

Finally, to our knowledge, this is the first niche/SDM approach that has compared air temperature and LST. Such datasets have only recently become available for an extended period (the MODIS LST record is now over a decade long). Our results highlight the ecological relevance of LST in explaining biotic distributions. LST was an especially important predictor for $C_3$ PACMADs, ranking much higher than any of the air temperature measures (Table 1). LST is likely a more relevant biological quantity than air temperature for many scientific questions, as it is the actual temperature of the object or surface of interest (often referred to as the 'skin temperature'). Because any organism's thermodynamic temperature can frequently deviate from air temperature, it is of equal or greater importance to measure skin temperature. As many biological processes (such as respiration rates) and ecologically important environmental attributes (such as vapor pressure deficit, which is critical for estimating evapotranspiration and its influence on species distributions), are nonlinearly dependent on temperature, substantial errors can be introduced when using air temperature in place of the skin temperature. Satellite-derived measures of skin temperature are also strongly related to the hydroclimatology of the land surface (Anderson *et al*., 2007; Karneili *et al*., 2010). LST should be tightly correlated with % tree cover and MAP as these variables are strongly related through transpirational cooling (Mildrexler *et al*., 2006). For all taxa, LST and % tree cover consistently ranked high in decreasing model performance when these variables were excluded, demonstrating the unique information contained in these variables. Importantly, the difference in LST was larger when comparing $C_3$ PACMAD with $C_4$ grasses than when comparing all $C_3$ grasses with $C_4$ grasses (3.7 °C vs. 1.6 °C) unlike air temperature differences, which were smaller. This is possibly explained by the drier, sunnier and more open conditions of $C_3$ BEPs, which would result in higher LST values than $C_3$ PACMADs.

## Acknowledgment

We thank T. Giambelluca for use of his climate data for Hawaii, Clyde Imada (Bishop Museum), and three anonymous reviewers for their comments on a previous draft. This work was conducted while S. P. was a Postdoctoral Associate at the National Center for Ecological Analysis and Synthesis, a Center funded by NSF (Grant #EF-0553768), the University of California, Santa Barbara, and the State of California. C. J. S. acknowledges the support of a NASA New Investigator Program award that supported this research.

## References

Anderson MC, Norman JM, Mecikalski JR, Otkin JA, Kustas WP (2007) A climatological study of evapotranspiration and moisture stress across the continental United States based on thermal remote sensing: 1. Model formulation. *Journal of Geophysical Research*, **112**, D10117.

Austin MP, Smith TM (1989) A new model for the continuum concept. *Vegetatio*, **83**, 35–47.

Bisrat SA, White MA, Beard KH, Cutler DR (2011) Predicting the distribution potential of an invasive frog using remotely sensed data in Hawaii. *Diversity and Distributions*, **18**, 648–660.

Bradley BA, Fleishman E (2008) Can remote sensing of land cover improve species distribution modelling? *Journal of Biogeography*, **35**, 1158–1159.

Buermann W, Saatchi S, Smith TB, Zutta BR, Chaves JA, Milá B, Graham CH (2008) Predicting species distributions across the Amazonian and Andean regions using remote sensing data. *Journal of Biogeography*, **35**, 1160–1176.

Chapman DS, Purse BV (2011) Community versus single-species distribution models for British plants. *Journal of Biogeography*, **38**, 1524–1535.

Chazdon RL (1978) Ecological aspects of the distribution of $C_4$ grasses in selected habitats of Costa Rica. *Biotropica*, **10**, 265–269.

Chuine I, Morin X, Sonié L *et al*. (2012) Climate change might increase the invasion potential of the alien $C_4$ grass Setaria parviflora (Poaceae) in the Mediterranean Basin. *Diversity and Distributions*, **18**, 661–672.

Collatz GJ, Berry JA, Clark JS (1998) Effects of climate and atmospheric $CO_2$ partial pressure on the global distribution of $C_4$ grasses : present, past, and future. *Ecology*, **114**, 441–454.

Daly C (2006) Guidelines for assessing the suitability of spatial climate data sets. *International Journal of Climatology*, **26**, 707–721.

Davidson A, Csillag F (2003) A comparison of three approaches for predicting $C_4$ species cover of northern mixed grass prairie. *Remote Sensing of Environment*, **86**, 70–82.

Edwards EJ, Smith SA (2010) Phylogenetic analyses reveal the shady history of $C_4$ grasses. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 2532–2537.

Edwards EJ, Still CJ (2008) Climate, phylogeny and the ecological distribution of $C_4$ grasses. *Ecology Letters*, **11**, 266–276.

Edwards EJ, Still CJ, Donoghue MJ (2007) The relevance of phylogeny to studies of global change. *Trends in Ecology and Evolution*, **22**, 243–249.

Ehleringer JR, Monson RK (1993) Evolutionary and ecological aspects of photosynthetic pathway variation. *Annual Review of Ecology and Systematics*, **24**, 411–439.

Ehleringer JR, Cerling TE, Helliker BR (1997) $C_4$ photosynthesis, atmospheric $CO_2$, and climate. *Oecologia*, **112**, 285–299.

Elith JH, Graham CH, Anderson RP *et al*. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.

Elith J, Phillips SJ, Hastie T, Dudík M, Chee YE, Yates CJ (2011) A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, **17**, 43–57.

von Fischer JC, Tieszen LL, Schimel DS (2008) Climate controls on $C_3$ vs. $C_4$ productivity in North American grasslands from carbon isotope composition of soil organic matter. *Global Change Biology*, **14**, 1141–1155.

Fox D, Koch P (2004) Carbon and oxygen isotopic variability in Neogene paleosol carbonates: constraints on the evolution of the C4-grasslands of the Great Plains, USA. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **207**, 305–329.

Franklin J (2009) *Mapping Species Distributions: Spatial Inference and Prediction*. Cambridge University Press, Cambridge.

Giambelluca TW, Nullet MA, Schroeder TA (1986) *Rainfall Atlas of Hawai'i*. Hawai'i Division of Water and Land Development, Department of Land and Natural Resources, Honolulu.

Giambelluca TW, Diaz HF, Luke MSA (2008) Secular temperature changes in Hawai'i. *Geophysical Research Letters*, **35**, 1–5.

Gillespie TW, Foody GM, Rocchini D, Giorgi AP, Saatchi S (2008) Measuring and modelling biodiversity from space. *Progress in Physical Geography*, **32**, 203–221.

Grass Phylogeny Working Group II (2012) New grass phylogeny resolves deep evolutionary relationships and discovers $C_4$ origins. *New Phytologist*, **193**, 304–312.

Guisan A, Thuiller W (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, **8**, 993–1009.

Hansen M, DeFries R, Townshend JR, Carroll M, Dimiceli C, Sohlberg R (2003) Global percent tree cover at spatial resolution 500 Meters: first results of the MODIS Vegetation Continuous Fields algorithm. *Earth Interactions*, **7**, 1–15.

Hernandez PA, Franke I, Herzog SK *et al*. (2008) Predicting species distributions in poorly-studied landscapes. *Biodiversity and Conservation*, **17**, 1353–1366.

Holt RD (2009) Bringing the Hutchinsonian niche into the 21st century: ecological and evolutionary perspectives. *Proceedings of the National Academy of Sciences*, **106**, 19659–19665.

Huntley B, Green RE, Collingham YC *et al*. (2004) The performance of models relating species geographical distributions to climate is independent of trophic level. *Ecology Letters*, **7**, 417–426.

Karneili A, Agam N, Pinker RT *et al*. (2010) Use of NDVI and land surface temperature for drought assessment: merits and limitations. *Journal of Climate*, **23**, 618–633.

Krebs CJ (1989) *Ecological Methodology*. Harper & Row, New York.

Mildrexler DJ, Zhao M, Running SW (2006) Where are the hottest spots on earth? *EOS Transactions*, **87**, 461–467.

Levins R (1968) *Evolution in changing environments*. *Monographs in populationbiology*, Vol 2. Princeton University Press, Princeton, NJ.

New M, Lister D, Hulme M, Makin I (2002) A high-resolution data set of surface climate over global land areas. *Climate Research*, **21**, 1–25.

O'Connor TG, Bredenkamp GJ (1997) Grassland. In: *Vegetation of Southern Africa* (eds Cowling TM, Richardson D), pp. 215–257. Cambridge University Press, Cambridge.

Osborne CP, Beerling DJ (2006) Nature's green revolution: the remarkable evolutionary rise of $C_4$ plants. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **361**, 173–194.

Osborne CP, Freckleton RP (2009) Ecological selection pressures for C4 photosynthesis in the grasses. Proceedings of the Royal Society. *Biological Sciences*, **276**, 1753–1760.

Paruelo JM, Lauenroth WK (1996) Relative abundance of plant functional types in grasslands and shrublands of North America. *Ecological Applications*, **6**, 1212–1224.

Pearcy RW, Ehleringer JR (1984) Comparative ecophysiology of $C_3$ and $C_4$ plants. *Plant, Cell and Environment*, **7**, 1–13.

Phillips SJ, Dudík M (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, **31**, 161–175.

Phillips SJ, Anderson R, Schapire R (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.

Sage RF, McKown AD (2006) Is $C_4$ photosynthesis less phenotypically plastic than $C_3$ photosynthesis? *Journal of Experimental Botany*, **57**, 303–317.

Sage RF, Monson RK (1999) *$C_4$ Plant Biology*. Academic Press, San Diego, California.

Still CJ, Berry JA, Collatz GJ, DeFries RS (2003a) Global distribution of $C_3$ and $C_4$ vegetation: carbon cycle implications. *Global Biogeochemical Cycles*, **17**, 1006.

Still CJ, Berry JA, Ribas-Carbo M, Helliker BR (2003b) The contribution of $C_3$ and $C_4$ plants to the carbon cycle of a tallgrass prairie: an isotopic approach. *Oecologia*, **136**, 347–359.

Taylor SH, Hulme SP, Rees M, Ripley BS, Woodward FI, Osborne CP (2010) Ecophysiological traits in $C_3$ and $C_4$ grasses: a phylogenetically controlled screening experiment. *The New Phytologist*, **185**, 780–791.

Teeri JA, Stowe L (1976) Climatic patterns and distribution of $C_4$ grasses in North America. *Oecologia*, **23**, 1–12.

Tieszen LL, Reed B, Bliss NB, Wylie B, DeJong D (1997) NDVI, $C_3$ and $C_4$ production, and distributions in Great Plains grassland land cover classes. *Ecological Applications*, **7**, 59–78.

Urban MC, Tewksbury JJ, Sheldon KS (2012) On a collision course: competition and dispersal differences create no-analogue communities and cause extinctions during climate change. *Proceedings of the Royal Society Biological Sciences*, **279**, 2072–2080.

Vicentini A, Barber JC, Aliscioni SS, Giussani LM, Kellogg EA (2008) The age of the grasses and clusters of origins of $C_4$ photosynthesis. *Global Change Biology*, **14**, 2963–2977.

Vogel JC, Fuls A, Danin A (1986) Geographical and environmental distribution of $C_3$ and $C_4$ grasses in the Sinai, Negev, and Judean Deserts. *Oecologia*, **70**, 258–265.

Wagner WL, Herbst DR, Sohmer S (1999) *Manual of the Flowering Plants of Hawai'i* (Revised edn.) University of Hawai'i Press, Honolulu.

Wan Z, Zhang Y, Zhang Q, and Li Z-L (2002) Validation of the land-surface temperature products retrieved from Terra Moderate Resolution Imaging Spectroradiometer data. *Remote Sensing of Environment*, **83**, 163–180.

Warren DL, Glor RE, Turelli M (2008) Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution*, **62**, 2868–2883.

Winslow JC, Hunt ER, Piper SC (2003) The influence of seasonal water availability on global $C_3$ versus $C_4$ grassland biomass and its implications for climate change research. *Ecological Modelling*, **163**, 153–173.

## Supporting Information

Additional Supporting Information may be found in the online version of this article:

**Appendix A.** Evenly representing each species for aggregated clade modeling.
**Figure S1.** Response curves generated when using the same number of occurrence points for each species in each group ($C_3$ PACMAD, $C_3$ BEPs and $C_4$ PACMADs).
**Table S1.** Model performance and training gain results using the same number of occurrence points for each species in each group ($C_3$ BEP, $C_3$ PACMAD and $C_4$ PACMAD).