Sensitivity of leaf size and shape to climate: global patterns and paleoclimatic applications

Daniel J. Peppe1,2, Dana L. Royer1, Bárbara Cariglino3,4, Sofia Y. Oliver1, Sharon Newman1, Elias Leight1, Grisha Enikolopov1, Margo Fernandez-Burgos1, Fabianny Herrera5, Jonathan M. Adams6, Edwin Correa7, Ellen D. Currano8, J. Mark Erickson9, Luis Felipe Hinojosa10, John W. Hoganson11, Ari Iglesias12, Carlos A. Jaramillo7, Kirk R. Johnson13, Gregory J. Jordan14, Nathan J. B. Kraft15, Elizabeth C. Lovelock16,17, Christopher H. Lusk18, Ülo Niinemets19, Josep Pen˜uelas20, Gillian Rapson21, Scott L. Wing22 and Ian J. Wright18

1Department of Earth and Environmental Sciences, Wesleyan University, Middletown, CT 06459, USA; 2Department of Geology, Baylor University, Waco, TX 76798, USA; 3Department of Geosciences, Pennsylvania State University, University Park, PA 16802, USA; 4Museo Argentino de Ciencias Naturales ‘B. Rivadavia’, Buenos Aires, Argentina; 5Florida Museum of Natural History and Department of Biology, University of Florida, Gainesville, FL 32611, USA; 6Department of Biological Sciences, Seoul National University, Seoul 151, Korea; 7Smithsonian Tropical Research Institute, Apartado Postal 0843-03092, Balboa, Ancon, Panama; 8Department of Geology, Miami University of Ohio, Oxford, OH 45056, USA; 9Department of Geology, St Lawrence University, Canton, NY 13617, USA; 10Faculty of Science, University of Chile, Institute of Ecology and Biodiversity, Santiago, Chile; 11North Dakota Geological Survey, Bismarck, ND 58505, USA; 12Facultad de Ciencias Naturales y Museo, Universidad Nacional de La Plata, 1900 La Plata, Argentina; 13Denver Museum of Nature and Science, Denver, CO 80205, USA; 14School of Plant Science, University of Tasmania, Private Bag 55, Hobart 7001, Tas., Australia; 15Biodiversity Research Centre, University of British Columbia, Vancouver, BC V6T 1Z4, Canada; 16John Day Fossil Beds National Monument, National Parks Service, Kimberly, OR 97848, USA; 17Department of Earth Science, University of California Santa Barbara, Santa Barbara, CA 93106, USA; 18Department of Biological Sciences, Macquarie University, Sydney, NSW 2109, Australia; 19Institute of Agricultural and Environmental Sciences, Estonian University of Life Sciences, Tartu 51014, Estonia; 20Global Ecology Unit CCREAF-CEAB-CSIC, Center for Ecological Research and Forestry Applications, Universitat Autónoma de Barcelona, Barcelona, Spain; 21Ecology Group, Institute of Natural Resources, Massey University, Private Bag 11222, Palmerston North, New Zealand; 22Department of Paleobiology, Smithsonian Institution, National Museum of Natural History, Washington, DC 20013, USA

Author for correspondence:
Daniel J. Peppe
Tel: +1 254 7102629
Email: daniel_peppe@baylor.edu

Received: 4 August 2010
Accepted: 23 November 2010

New Phytologist (2011) 190: 724–739
doi: 10.1111/j.1469-8137.2010.03615.x

Key words: climate proxies, leaf lifespan, leaf physiognomy, paleobotany, paleoclimate, phylogeny, precipitation, temperature.

Summary

- Paleobotanists have long used models based on leaf size and shape to reconstruct paleoclimate. However, most models incorporate a single variable or use traits that are not physiologically or functionally linked to climate, limiting their predictive power. Further, they often underestimate paleotemperature relative to other proxies.
- Here we quantify leaf–climate correlations from 92 globally distributed, climatically diverse sites, and explore potential confounding factors. Multiple linear regression models for mean annual temperature (MAT) and mean annual precipitation (MAP) are developed and applied to nine well-studied fossil floras.
- We find that leaves in cold climates typically have larger, more numerous teeth, and are more highly dissected. Leaf habit (deciduous vs evergreen), local water availability, and phylogenetic history all affect these relationships. Leaves in wet climates are larger and have fewer, smaller teeth. Our multivariate MAT and MAP models offer moderate improvements in precision over univariate approaches (± 4.0 vs 4.8°C for MAT) and strong improvements in accuracy. For example, our provisional MAT estimates for most North American fossil floras are considerably warmer and in better agreement with independent paleoclimate evidence.
- Our study demonstrates that the inclusion of additional leaf traits that are functionally linked to climate improves paleoclimate reconstructions. This work also illustrates the need for better understanding of the impact of phylogeny and leaf habit on leaf–climate relationships.
Introduction

The sizes and shapes (physiognomy) of leaves correlate strongly with temperature and moisture from global to local scales, and there are biological bases for these relationships (Bailey & Sinnott, 1915, 1916; Webb, 1968; Lewis, 1972; Givnish, 1979, 1984; Wolfe, 1979, 1993; Hall & Swaine, 1981; Richards, 1996; Wilf, 1997; Wilf et al., 1998; Jacobs, 1999, 2002; Feild et al., 2005; Traiser et al., 2005; Royer & Wilf, 2006). Paleobotanists have long used these leaf–climate correlations to develop proxies for reconstructing paleoclimate (Bailey & Sinnott, 1915, 1916; Dilcher, 1973; Wing & Greenwood, 1993; Wolfe, 1993, 1995; Wilf, 1997; Wilf et al., 1998; Jacobs, 1999, 2002; Kowalski & Dilcher, 2003; Traiser et al., 2005; Adams et al., 2008).

One key leaf–climate association is between leaf teeth and both temperature and local water availability (Baker-Brosh & Peet, 1997; Feild et al., 2005; Royer & Wilf, 2006). The percentage of woody, non-monocotyledonous angiosperms (woody dicots) at a site with toothed leaves (Bailey & Sinnott, 1916; Wolfe, 1979; Wilf, 1997), as well as variables related to tooth count and tooth size (Huff et al., 2003; Royer et al., 2005), all negatively correlate with mean annual temperature (MAT). The prevalence of leaf teeth in cool climates is potentially an adaptation for increased carbon uptake through enhanced sap flow early in the growing season (Billings, 1905; Bailey & Sinnott, 1916; Wolfe, 1993; Baker-Brosh & Peet, 1997; Wilf, 1997; Royer & Wilf, 2006). In cold environments, this early-season pulse in sap flow may allow plants with toothed leaves to maximize the duration of their growing seasons; in warmer climates, the potential benefit is outweighed by the attendant water costs (Wing et al., 2000; Royer & Wilf, 2006). The relationship between leaf teeth and enhanced sap flow may also help explain why, at a given MAT, toothed species are sometimes more abundant in locally wet environments where the water cost associated with teeth may be less important (the ‘freshwater-margin effect’ in and near swamps, and near lakes and streams; Wolfe, 1993; Burnham et al., 2001; Kowalski & Dilcher, 2003; Greenwood, 2005; Royer et al., 2009a). Teeth may also release excess root pressure through guttation, preventing the flooding of intercellular spaces in the leaf lamina and, in cooler climates, freeze–thaw embolisms (Feild et al., 2005).

Leaf size is also sensitive to climate: site-mean leaf size typically scales with water availability and, to a lesser degree, temperature (Webb, 1968; Dilcher, 1973; Dolph & Dilcher, 1980a,b; Givnish, 1984; Greenwood, 1992; Wilf et al., 1998). Energy balance models predict that for a given level of radiation and wind speed, leaf temperatures are higher in large canopy leaves because of their thicker boundary layers (Vogel, 1968, 1970, 2009; Parkhurst & Loucks, 1972; Givnish, 1979, 1984, 1987; Gates, 1980). Warmer leaf temperatures promote both photosynthesis and transpiration; thus, plants in drier climates tend to have smaller leaves to reduce evaporative cooling, while in more humid climates larger leaves are common because the attendant water cost is less critical (Givnish, 1984).

Other factors can affect these leaf–climate relationships. It has been commonly claimed, but never rigorously tested, that deciduous species are more likely to be toothed than evergreen species (Bailey & Sinnott, 1916; Givnish, 1979; Wolfe, 1993; Jacobs, 2002). Shared phylogenetic and/or regional histories of floras may also be important. Multiple studies have noted different leaf–climate relationships in the northern and southern hemispheres, with extant southern hemisphere temperate floras typically having a higher percentage of untoothed species than temperature-equivalent northern hemisphere floras (Greenwood, 1992; Jordan, 1997; Jacobs, 1999, 2002; Kennedy et al., 2002; Kowalski, 2002; Greenwood et al., 2004; Aizen & Ezcurra, 2008; Hinojosa et al., 2010; Steart et al., 2010). These differences may be the result of regional differences in environment, such as soil fertility and thermal seasonality, and/or phylogenetic differences (Wolfe and Upchurch, 1987; Jordan, 1997; Greenwood et al., 2004). Other regional differences in leaf–climate relationships exist, although often the differences are not statistically significant (e.g. Gregory-Wodzicki, 2000; Traiser et al., 2005; Miller et al., 2006; Su et al., 2010).

To address these potential problems, regional calibrations have been developed (for example, see Hinojosa et al., 2010; Su et al., 2010) and make the assumption that leaf–climate relationships within a region were the same as they are now. This is a valid assumption in some cases (e.g. late Neogene and Quaternary floras), but not in others (e.g. Cretaceous and early Cenozoic floras), particularly given the uncertainty in the cause for the difference and the major environmental and evolutionary changes since the Cretaceous. If phylogeny is important, then regional calibrations assume that past lineage composition of the fossil flora was similar to the current composition in the region, and that evolution and extinction subsequent to the deposition of the fossils has not changed leaf–climate relationships in those lineages (Jordan, 1997; Hinojosa et al., 2010; Little et al., 2010). If current environmental drives regional differences, then regional calibrations must assume that critical environmental features, such as soil fertility and thermal seasonality, were the same in the relevant region at the time of deposition of the fossils, another questionable assumption. Overall, the effects of phylogeny and regional environmental differences on leaf–climate correlations are poorly constrained and have rarely been tested in a proper statistical framework (Hinojosa et al., 2010; Little et al., 2010). As more detailed large-scale assessments of the relationship between phylogeny and leaf traits become available (Little et al., 2010), comparing the leaf–climate correlations...
in this and other studies to related methods that incorporate phylogenetic relationships (Felsenstein, 1985; Garland et al., 1992; Westoby et al., 1998) will likely provide additional insights into the ecological and evolutionary forces shaping trait–climate correlations.

The most common leaf physiognomic methods for estimating MAT and mean annual precipitation (MAP), leaf-margin analysis and leaf-area analysis, are each based on a single variable, the percentage of untoothed species at a site and site-mean leaf size, respectively (Wolfe, 1979; Wilf, 1997, 1998; Jacobs, 2002; Miller et al., 2006). Although climate estimates from these methods commonly agree with independent evidence (e.g. Greenwood & Wing, 1995; Wing et al., 2000; Uhl et al., 2003; Wilf et al., 2003a,b; Mosbrugger et al., 2005; Yang et al., 2007; Greenwood et al., 2010), there are many instances where these proxies provide cooler and drier estimates of MAT and MAP than alternative proxy evidence (Utescher et al., 2000; Liang et al., 2003; Fricke & Wing, 2004; Kvacek, 2007; Wing et al., 2009b). Because these are univariate approaches, additional characters may lead to improvements.

To this end, Wolfe (1993, 1995) developed a method called Climate-Leaf Analysis Multivariate Program (CLAMP), which uses 31 categorical leaf states, including leaf-margin and leaf-size categories. The method correlates the characters to climate using canonical correspondence analysis (CCA; Wolfe, 1995). Because CLAMP more thoroughly describes leaf physiognomy, it might be expected to result in more accurate climate estimates than the univariate approaches, but in practice it does not (Jacobs & Deino, 1996; Wilf, 1997; Wiemann et al., 2003; Gregory-Wodzicki, 2000; Kowalski & Dilcher, 2003; Royer et al., 2005; Dilcher et al., 2009; Smith et al., 2009b). This may be caused by errors and biases related to the ambiguity of character definitions, the categorical nature of the character states, weak or non-existent correlations between climate and some character states, and problems related to using CCA in a predictive framework (Jordan, 1997; Wilf, 1997; Wilf et al., 1998, 1999; Green, 2006; Peppe et al., 2010). Thus, although CLAMP is multivariate, it is fraught with systemic problems and does not produce more accurate climate estimates. Other multivariate approaches have been proposed (Wing & Greenwood, 1993; Stranks & England, 1997; Gregory-Wodzicki, 2000), but because they use the CLAMP characters they suffer from many of the same problems.

Recently, Huff et al. (2003) and Royer et al. (2005) developed a new procedure, called digital leaf physiognomy, which has three major advantages over CLAMP and the univariate approaches. First, it minimizes the ambiguity of CLAMP scoring because computer algorithms process most of the measurements. Second, it uses mostly continuous variables, such as tooth number and size, not categorical characters. Thus, for example, digital leaf physiognomy can discern between a leaf with one and 100 teeth, but CLAMP and leaf-margin analysis do not (Royer et al., 2005, 2008). Third, digital leaf physiognomy incorporates more traits that have a functional and/or physiological connection to climate, such as tooth number, tooth size, leaf area and degree of leaf dissection (see earlier discussion). Importantly, the traits used in digital leaf physiognomy can display some degree of phenotypic plasticity (Royer et al., 2009b), suggesting they can respond quickly to climate change even in the absence of evolutionary responses.

Using digital leaf physiognomy, Huff et al. (2003) and Royer et al. (2005) observed that leaves from cold climates are more likely to be highly dissected and to have many, large teeth; importantly, these correlations are consistent with the ecophysiological principles outlined earlier. Royer et al. (2005) also developed a preliminary, multiple linear regression model for predicting MAT that was considerably more accurate than leaf-margin analysis and CLAMP. A limitation of the study, however, is that it was based on 17 sites from eastern North America and Panama that spanned a limited biogeographic and climatic range (Fig. 1).

Here, we investigate correlations between leaf physiognomy and climate across 92 globally distributed sites from the biomes where fossil leaves are most likely to be preserved (Fig. 1). A major goal of the study was to assess global correlations of MAT and MAP to functionally linked leaf traits using a phylogenetically and climatically diverse data set of extant vegetation (Fig. 1). In addition, we quantitatively tested the importance of two potential confounding factors on these correlations: the evergreen effect (i.e. are woody dicot evergreens less likely to be toothed?) and the freshwater-margin effect (i.e. do freshwater-margin habitats contain a higher percentage of toothed species?). We also compared leaf–climate correlations between extant northern and southern hemisphere floras; however, it is beyond the scope of the present study to employ more formal phylogenetic tests (e.g. Little et al., 2010). Third, we developed multiple linear regression equations derived from the extant vegetation to estimate MAT and MAP. To gauge the accuracy of the equations, we estimated the climate of each extant site using a jackknife-type approach. We then applied the equations to nine, well-studied fossil floras and compared the climate reconstructions to other climate proxies, including leaf-margin analysis and leaf-area analysis.

Materials and Methods

Calibration sites

We photographed leaves of native, woody dicots from 92 geographically and climatically diverse extant sites (Fig. 1) (n = 6525 leaves and 3033 species-site pairs). This data set expands on the 17 calibration sites of Royer et al. (2005).
The majority of new sites \((n = 42)\) come from the CLAMP collection (Wolfe, 1993; Spicer, 2009), whose voucher specimens are housed in the Department of Paleobiology, National Museum of Natural History, Smithsonian Institution, Washington, DC, USA. Sampling was generally restricted to outer, exposed leaves in the canopy or tree crown (see the Supporting Information for detailed collection protocols). To test the potential of herbs as climate indicators, a collection of 34 herbaceous dicot species was made from north of Reed Gap in Wallingford, Connecticut (see Royer et al., 2010 for sampling details).

Mean annual temperature of our sites ranged from 0.1 to 27.7°C and MAP from 189 to 4694 mm (see the Supporting Information, Table S1). Mean monthly climate data were extracted from a global, interpolated 1 km spatial resolution climate model (WORLDCLIM, Hijmans et al., 2005). Where available, WORLDCLIM matches local climate station data at all but five sites for MAT \((\pm 0.3°C)\) and three sites for MAP \((\pm 22 \text{ mm})\). For the seven sites where the model deviated strongly from station data \((> \pm 2.0°C \text{ or } \pm 100 \text{ mm})\), we relied on the latter. We defined the growing season as the period during which the...
mean monthly minimum temperature exceeded 0°C and precipitation exceeded 20% of the maximum monthly precipitation, and growing-degree-days as the number of degree-days in a year when the average temperature exceeded 10°C (Table S2) (e.g. Johnson et al., 2000). The functional basis of leaf physiognomy (see the Introduction) may imply that physiognomic traits are more closely linked to growing-season variables such as growing-season precipitation, growing-season mean temperature and mean annual range in temperature (warmest month mean minus coldest month mean); however, we focus here on MAT and MAP because correlations of leaf physiognomy to annual and growing-season climate variables were very similar (Table S3).

Typically, at least two leaves or leaflets per species at each site were used. More than two leaves were used if there was a large variation in leaf form (e.g. compound leaves, species with and without lobes or teeth). Computerized resampling indicates that this level of sampling is sufficient for detecting site-level patterns (Royer et al., 2005). All leaf images used in this study are available from Dryad (http://dx.doi.org/10.5061/dryad.8101) and the personal websites of DJP and DLR. Leaves were manipulated in Adobe Photoshop (Adobe Systems, San Jose, CA, USA) to separate the petiole and teeth (if present) from the blade following the protocols of Royer et al. (2005). Most physiognomic characters were calculated using IMAGEJ (http://rsbweb.nih.gov/ij/); presence of teeth and number of teeth were determined visually (see Table S4 for all physiognomic data). Definitions of characters follow Royer et al. (2005) (see also Table S2). Site means (Table S1) were calculated from species means. For variables involving teeth, untoothed species were excluded in order to maintain normal distributions (Huff et al., 2003). Because climate impacts leaf physiognomy, we plot climate as the independent variable and leaf traits as the dependent variables. Site-mean data were correlated to climate with single and multiple linear regression (SPSS 17; SPSS Science, Chicago, IL, USA) and with CCA (CANOCO 4.5; Microcomputer Power, Ithaca, NY, USA). Using leaf traits as the independent variables and climate as the dependent variable, we developed predictive multiple linear regression models for MAT and MAP. The variables shape factor (perimeter²/blade area), compactness (4π × blade area/perimeter²), number of teeth, tooth area and perimeter/area cannot be calculated in any meaningful way for fragmentary fossils and were excluded from our models (Royer et al., 2005). However, these traits may be useful for studying extant leaf–climate relationships (Royer et al., 2005, 2008; see Table S5 for most significant MAT and MAP models derived using all variables). Models were considered only if: the model and all individual variables in the model were significant at the α = 0.05 level, and variables did not show a high degree of co-linearity with the other predictor variables (variance inflation factor < 10; Sokal & Rohlf, 1995). We used the ordinary least squares regression module in the program SMATR (http://www.bio.mq.edu.au/ecology/SMATR/; Warton et al., 2006) to test for slope and intercept differences between regression lines. We define accuracy as the extent to which a given MAT or MAP estimate agrees with other independent lines of evidence. Precision is defined as uncertainty of an estimate derived from a regression model (i.e. the standard error).

Fossil sites

We applied the digital leaf physiognomy MAT and MAP models, as well as leaf-margin analysis and leaf-area analysis, to 10 fossil floras from the latest Cretaceous and early Paleogene (c. 66 to c. 47.0 million years ago (Ma)) of North and South America (Fig. 1, Table 1). All floras are well-studied and represent a broad range of interpreted biomes and phylogenetic histories. For each site, we processed 1–48 specimens of each species or morphotype (median = 3; see Tables S6, S7 for all fossil physiognomic data). As fossil specimens are in rock matrix and often fragmentary, additional processing protocols were necessary (Cariglino, 2007; see Methods S1). Because it is possible to determine the margin type (toothed, untoothed) of specimens that cannot be digitally processed, we calculated the percentage of untoothed species based on all species, not just the digitally-processed species.

The Fox Hills flora is from the Linton Member of the Fox Hills Formation and is late Maastrichtian in age (c. 66 Ma; Peppe, 2003; Peppe et al., 2007; Table 1). Specimens are stored at the North Dakota Heritage Center in Bismarck, North Dakota, USA, and at St Lawrence University in Canton, New York, USA.

The Fort Union Formation floras (Williston Basin I, II, and III) are from the Fort Union Formation in the Williston Basin of southwestern North Dakota, USA (65.5 to c. 58.5 Ma; Peppe, 2009, 2010; Table 1). We grouped these taxa by floral zone following Peppe (2009, 2010). Specimens used in this study are housed at the Yale Peabody Museum in New Haven, Connecticut, USA.

The Palacio de los Loros flora (P. Loros), first described in Berry (1937), is from the westernmost exposures of the Salamancan Formation in southern Chubut Province, Argentina, and is early Paleocene in age (c. 61.7 Ma; Iglesias et al., 2007; Table 1). Specimens used in this study come from two outcrops representing the same general depositional environment that are geographically and stratigraphically close to each other (Iglesias et al., 2007). The specimens are reported by Iglesias et al. (2007) and are housed at the Museo Paleontológico Egidio Feruglio in Trelew, Argentina.

The Cerrejón flora is from the middle Late Paleocene (c. 58 Ma) Cerrejón Formation of Colombia reported by Wing et al. (2009b; Table 1). Specimens are housed at INGEOMINAS in Bogotá, Colombia.
Table 1  Age, paleolatitude, number of species, and provisional mean annual temperature and mean annual precipitation estimates for fossil floras

<table>
<thead>
<tr>
<th>Site</th>
<th>Age (Ma)</th>
<th>Paleolatitude</th>
<th>Number of woody dicotyledonous angiosperm species in flora</th>
<th>Number of woody dicotyledonous angiosperm species processed</th>
<th>Digital leaf physiognomy MAT estimate (°C)</th>
<th>Leaf-margin analysis MAT estimate (°C)</th>
<th>Regional digital leaf physiognomy MAT estimate (°C)</th>
<th>Digital leaf physiognomy MAP estimate (cm²)</th>
<th>Leaf-area analysis MAP estimate (cm²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fox Hills</td>
<td>66.5</td>
<td>49.7</td>
<td>34</td>
<td>25</td>
<td>21.6</td>
<td>14.8</td>
<td>17.1</td>
<td>141 (+116, –64)</td>
<td>152 (+125, –68)</td>
</tr>
<tr>
<td>Williston Basin I</td>
<td>65.5–64.0</td>
<td>50.8</td>
<td>26</td>
<td>20</td>
<td>15.7</td>
<td>10.9</td>
<td>12.6</td>
<td>175 (+144, –79)</td>
<td>157 (+129, –71)</td>
</tr>
<tr>
<td>Williston Basin II</td>
<td>64.0–63.0</td>
<td>50.8</td>
<td>31</td>
<td>23</td>
<td>15.0</td>
<td>10.2</td>
<td>12.4</td>
<td>148 (+122, –69)</td>
<td>156 (+129, –71)</td>
</tr>
<tr>
<td>Williston Basin III</td>
<td>61.0–58.5</td>
<td>50.8</td>
<td>19</td>
<td>18</td>
<td>16.3</td>
<td>9.4</td>
<td>11.6</td>
<td>152 (+125, –68)</td>
<td>157 (+129, –71)</td>
</tr>
<tr>
<td>Palacio de los Loros</td>
<td>61.7</td>
<td>–54.7</td>
<td>36</td>
<td>33</td>
<td>12.8</td>
<td>12.9</td>
<td>12.8</td>
<td>125 (+103, –56)</td>
<td>144 (+119, –65)</td>
</tr>
<tr>
<td>Cerrejón</td>
<td>58</td>
<td>7.4</td>
<td>48</td>
<td>48</td>
<td>23.6</td>
<td>20.5</td>
<td>14.0</td>
<td>264 (+217, –119)</td>
<td>212 (+174, –96)</td>
</tr>
<tr>
<td>Hubble Bubble</td>
<td>55.8</td>
<td>47.6</td>
<td>29</td>
<td>16</td>
<td>20.3</td>
<td>17.9</td>
<td>20.2</td>
<td>147 (+121, –66)</td>
<td>146 (+120, –66)</td>
</tr>
<tr>
<td>Laguna del Hunco</td>
<td>51.9</td>
<td>–49.0</td>
<td>132</td>
<td>119</td>
<td>10.9</td>
<td>14.1</td>
<td>16.9</td>
<td>127 (+103, –57)</td>
<td>142 (+117, –64)</td>
</tr>
<tr>
<td>Republic</td>
<td>49.4</td>
<td>50.9</td>
<td>45</td>
<td>41</td>
<td>9.0</td>
<td>9.2</td>
<td>8.9</td>
<td>134 (+110, –60)</td>
<td>135 (+111, –61)</td>
</tr>
<tr>
<td>Bonanza</td>
<td>47.3</td>
<td>40.4</td>
<td>28</td>
<td>24</td>
<td>14.8</td>
<td></td>
<td></td>
<td>110 (+90, –50)</td>
<td></td>
</tr>
</tbody>
</table>

Ma, million years ago; MAT, mean annual air temperature; MAP, mean annual precipitation.

Paleolatitude reconstruction based on Torsvik et al. (2008).

Standard Error (SE) is ± 4.0°C.

Standard Error is ± 4.8°C. Independent proxy evidence suggests that most of these MAT estimates are considerable underestimates (see text).

Regional digital leaf physiognomy models were created for North America and South America. The North American model ($r^2 = 0.81$, SE = ± 3.3°C) used the variables percent untoothed and number of teeth : internal perimeter. The North American model was based on all extant sites in our calibration from North America, Central America, and Asia and was applied to all fossil sites from North America. The South American model ($r^2 = 0.96$, SE = ± 1.7°C) used the percent untoothed and Feret’s diameter ratio. The model was based on all extant sites in our calibration from South America and was applied to all fossil sites from South America.

Standard errors are asymmetrical because they were converted from logarithmic units.

MAT and MAP for Bonanza were not reconstructed (see discussion in text).

The Hubble Bubble flora (USNM locality 42384) is from the Willwood Formation in the Bighorn Basin, Wyoming, USA, and dates to within the Paleocene–Eocene thermal maximum (PETM, c. 55.8 Ma, Currano et al., 2008, 2010; Wing et al., 2009a; Table 1). Specimens are housed in the Department of Paleobiology, National Museum of Natural History, Smithsonian Institution in Washington, DC, USA.

The Early Eocene Laguna del Hunco flora, which was first described by Berry (1925), is 51.91 ± 0.22 Ma and comes from the Tufolitas Laguna del Hunco, a lacustrine unit in the Chubut River volcanoclastic complex in the northwestern Chubut Province in Patagonia, Argentina (Wilf et al., 2003a, 2005a; Table 1). Specimens are stored at the Museo Paleontológico Egidio Feruglio in Trelew, Argentina.

The Bonanza flora, first described by MacGinitie (1969), is from the uppermost Parachute Creek Member of the Green River Formation in northeastern Utah, USA, and is early Middle Eocene in age (c. 47.3 Ma, Smith et al., 2008; Table 1). Specimens studied here are a subset of those reported in Wilf et al. (2005b). Both collections are housed at the Denver Museum of Nature and Science in Denver, Colorado, USA.

Results and Discussion

Physiognomic correlation with climate

The site means of many leaf physiognomic characters correlate strongly with temperature and precipitation (Figs 2, 3, Table S3). Notably, MAT correlates significantly to tooth-related characters, including percent of untoothed species ($r^2 = 0.58$, $P < 0.001$), number of teeth ($r^2 = 0.23$, $P < 0.001$), tooth area : internal perimeter ($r^2 = 0.11$, $P = 0.001$; internal perimeter is the leaf perimeter after teeth are removed), and number of teeth : internal perimeter ($r^2 = 0.35$, $P < 0.001$), as well as leaf dissection variables such as perimeter ratio ($r^2 = 0.37$, $P < 0.001$; blade perimeter divided by internal perimeter) and shape factor ($r^2 = 0.22$, $P < 0.001$) (Fig. 2). In warmer climates, leaves generally have fewer, smaller teeth and are less dissected, as previously observed by Royer et al. (2005).

Leaf-margin analysis models are currently calibrated with woody dicots because the physiognomy of herbaceous angiosperms is considered to be less sensitive to climate (Bailey & Sinnott, 1916). However, we found that the
percentage of untoothed herbaceous dicot species from a central Connecticut site (35.4%) was almost identical to that of woody dicots species from four nearby sites (mean = 34.5%; see Table S1). There may be potential for including herbaceous taxa in leaf-climate proxies, but further work is needed.

Moisture variables also significantly correlate with several physiognomic characters (Fig. 3, Table S3). Correlations are stronger with loge(MAP) than with untransformed MAP, probably owing to the non-normal distribution of MAP across sites (Fig. 1) and/or a non-linear relationship between MAP and water stress. As expected, leaf area positively correlates with loge(MAP) ($r^2 = 0.23$, $P < 0.001$; Fig. 3). Tooth area/blade area inversely correlates with loge(MAP) ($r^2 = 0.18$, $P < 0.001$), indicating that tooth area normalized to leaf area declines as precipitation increases (Fig. 3b). Although the functional significance of the relationship between precipitation and tooth area/blade area is unclear, it is consistent with a field study of *Acer rubrum* (Royer et al., 2008).

Water availability is a major control on leaf size, but temperature is also important (see the Introduction). In our calibration, MAT weakly correlates with leaf area ($r^2 = 0.09$, $P = 0.003$; Fig. S1). However, the relationship is

---

**Fig. 2** Relationship between site mean of physiognomic variables and mean annual temperature for the 92 calibration sites. Standard errors of the means for each site are plotted. Linear regression fits and associated $r^2$ and $P$ values are given in each panel (see also the Supporting Information, Table S3). For comparison, the regression of Wolfe (1979) is plotted in panel (a) (dashed line, $r^2 = 0.98$, $P < 0.001$). Standard errors for percent untoothed character are calculated using Eqn 3 in Miller et al. (2006). Internal perimeter is the blade perimeter after teeth are removed, perimeter ratio is blade perimeter divided by internal perimeter, and shape factor is $4\pi \times$ blade area/perimeter$^2$; see Table S2 for definitions of all variables.
sites south of 34°S. South America = southern South America and encompasses all sites north of 34°S latitude; NZ = New Zealand.

Potential confounding factors

Freshwater-margin effect  Sites with shallow water tables often have a higher percentage of species with teeth (c. 10–15%) than nearby drier sites (e.g. Burnham et al., 2001; Kowalski & Dilcher, 2003; Greenwood, 2005; Royer et al., 2009a). When using leaf-margin analysis to estimate MAT, this freshwater-margin effect could lead to an underestimation of up to 4°C (Burnham et al., 2001; Kowalski & Dilcher, 2003; Greenwood, 2005; Royer et al., 2005, 2009a). Further, the effect may be more severe (up to 10°C) at warmer temperatures (Kowalski & Dilcher, 2003). To test for this bias, we compared the slope of the regression fit between MAT and the percentage of untoothed species in the entire CLAMP data set (Wolfe, 1993) with that of the edaphically wet sites from Kowalski & Dilcher (2003) and Wolfe (1993), and found no statistical difference (P = 0.12; Fig. S2). By contrast, the y-intercept of a regression fit for edaphically dry CLAMP sites is shifted towards a higher percentage of untoothed species than that for edaphically wet sites (P < 0.001, Fig. S2). Thus, while we detected the freshwater effect, it probably does not strongly affect most paleo-MAT reconstructions because enough calibration sites contain a sufficient proportion of edaphically wet vegetation. The freshwater-margin effect reported by Kowalski & Dilcher (2003) is not representative; instead, a bias of up to 4°C is more plausible (Burnham et al., 2001; Fig. S2). Critically, the additional characters used in digital leaf physiognomy (e.g. number of teeth) generally show less sensitivity to the freshwater-margin effect than does percent of untoothed species (Fig. S3).

Effect of leaf habit and phylogeny  Are woody dicots with teeth more likely to be deciduous than evergreen at a given temperature (e.g. Bailey & Sinnott, 1916; Wolfe, 1993; Jacobs, 2002)? We selected sites from our calibration and from the CLAMP calibration that each contained > 15% evergreen and > 15% deciduous species (n = 29 sites). At individual sites, deciduous species are more likely to be toothed than evergreen species (P < 0.001); at warm temperatures, this discrepancy diminishes such that above 16°C MAT there is no significant effect (P = 0.18; Fig. 4). The slope of the relationship between the proportion of toothed deciduous species in a flora and MAT is significantly steeper than that of evergreen species (P = 0.04), indicating the presence of a leaf-habit effect. The evergreen effect is also present in many of the digital leaf physiognomy variables (Fig. 4). Evergreen species usually have evergreen and > 15% deciduous species (n = 29 sites). At individual sites, deciduous species are more likely to be toothed than evergreen species (P < 0.001); at warm temperatures, this discrepancy diminishes such that above 16°C MAT there is no significant effect (P = 0.18; Fig. 4). The slope of the relationship between the proportion of toothed deciduous species in a flora and MAT is significantly steeper than that of evergreen species (P = 0.04), indicating the presence of a leaf-habit effect. The evergreen effect is also present in many of the digital leaf physiognomy variables (Fig. 4). Evergreen species usually have

Even weaker after accounting for the covariation between MAT and MAP (Fig. 1b) with partial correlation (r² = 0.07, P = 0.01). As noted by Webb (1968), the leaf size–MAT relationship is strong within the Australia/New Zealand subset (Fig. S1; r² = 0.34, P = 0.002 for partial correlation). These observations raise two points. First, paleoclimate reconstructions based on leaf physiognomy should consider the interactive MAT–MAP control on physiognomy (discussed later). Second, regional differences in leaf–climate correlations exist (see also Figs 2–3 and the Introduction) and understanding their root causes, whether related to phylogeny, ecology or other factors, will improve paleoclimate reconstructions. Next, we discuss some of these biases.
fewer teeth ($P < 0.001$), smaller teeth ($P = 0.005$), and smaller teeth relative to their leaf area ($P = 0.03$) than do deciduous species at the same site. Evergreen species also have lower Feret’s diameter ratio (diameter of a circle with the same area as the leaf divided by the leaf’s longest axis; $P = 0.009$; Wolfe, 1993; Greenwood and Basinger, 1994). Furthermore, for these traits, either the slope of the regression between the trait and MAT in deciduous taxa is significantly steeper than for evergreen taxa (Feret’s diameter ratio: $P = 0.04$), or there is a significant difference in the $y$-intercept between deciduous and evergreen regressions (number of teeth, $P < 0.001$; tooth area, $P < 0.001$; number of teeth/blade area, $P < 0.001$). As with percent of untoothed species, the effect diminishes at warmer temperatures. We posit that evergreen species are less toothed because leaves in many evergreen taxa flush throughout the growing season, and thus any tooth-driven pulse in sap flow is more muted relative to neighboring deciduous taxa with a more synchronized leaf flush.

The physiognomy of evergreen taxa therefore responds differently to climate than that of deciduous taxa. Across all sites in our calibration and the CLAMP calibration, 11% of the variance in the relationship between MAT and the percent of untoothed species can be explained by the percentage of evergreen species. This leaf-habit effect can contribute to physiognomic differences both within and across sites (Figs 2, 4, 5). It may even provide a simple explanation for the higher percentage of untoothed species in southern hemisphere floras compared with northern hemisphere floras (Greenwood et al., 2004; Fig. 5) because southern hemisphere floras are typically dominated by evergreen taxa (mean = 98% vs 25% in our sites). However, as discussed in the Introduction, differing evolutionary or environmental histories of the floras may also contribute to differences.

Estimating climate from leaf physiognomy

**A global approach** Our models include all 92 calibration sites. The most commonly applied leaf-margin analysis model is based on 34 sites from eastern Asia (Wolfe, 1979; Wing & Greenwood, 1993). Because the correlation between MAT and the percent of untoothed species is remarkably strong in this data set ($r^2 = 0.98$), the standard errors quoted in the paleobotanical literature are typically $c. \pm 2^\circ C$ (Wilf, 1997). However, these errors are too low because factors associated with sample size and over-dispersion in the binary data set will inflate them (Miller et al., 2006). Weaker, but similar correlations to those of Wolfe (1979) are found in other regional studies (Wilf, 1997; Jacobs, 1999, 2002; Gregory-Wodzicki, 2000; Kennedy et al., 2002; Kowalski, 2002; Greenwood et al., 2004; Traiset et al., 2005; Miller et al., 2006; Adams et al., 2008; Aizen & Ezcurra, 2008; Hinojosa et al., 2010; Su et al., 2010). The leaf-margin analysis regression using our calibration, which is more climatically, geographically, and
Aizen & Ezcurra (2008), Su
Greenwood (1995), Wilf (1997), Burnham P
2011 The Authors

Percent untoothed
Mean annual temperature (°C)

Fig. 5 Relationship between the percentage of untoothed species in a flora and mean annual temperature for 535 globally distributed sites. Linear regression fit, \( r^2 \) and \( P \) values are given. For comparison, the regression of Wolfe (1979) is plotted (dashed line, \( r^2 = 0.98 \), \( P < 0.001 \)). Sources include Wolfe (1979, 1993), Midgley et al. (1995), Wilf (1997), Burham et al. (2001), Jacobs (1999, 2002), Gregory-Wodzicki (2000), Kennedy (1998), Kowalski (2002), Greenwood et al. (2004), Royer et al. (2005), Hinojosa et al. (2006), Aizen & Ezcurra (2008), Su et al. (2010), and this study.

phylogenetically diverse than any single regional data set (Fig. 1), is considerably weaker than most regional equations (Fig. 2a; \( r^2 = 0.58 \); standard error (SE) = ± 4.8°C). A larger compilation from the literature (\( n = 535 \) sites) is consistent with this finding (\( r^2 = 0.64 \); SE = ± 4.1°C; Fig. 5). This suggests that the error associated with a globally derived leaf-margin analysis equation is at least ± 4°C.

The calibration data for leaf-area analysis (Wilf et al., 1998; Jacobs, 1999, 2002; Gregory-Wodzicki, 2000) are primarily from low-latitude in Central America, South America, Asia and Africa. A compilation of these calibration sites suggests a strong univariate correlation between \( \log_2(MAP) \) and \( \log_2(leaf \ area) \) (\( r^2 = 0.71 \); Jacobs, 2002). Similar to leaf-margin analysis, our more global calibration indicates a much weaker correlation (\( r^2 = 0.23 \), Fig. 3).

Together, these results raise the obvious question: Why use a global model when regional calibrations are usually more precise (i.e. smaller standard errors)? On one hand, regional models capture the current relationship between leaf physiognomy and climate, which may be appropriate for specific floras. On the other hand, regional models capture a narrower slice of biological and ecological information (see the Introduction), which is not appropriate for fossil floras with a taxonomic composition or environmental setting different from the modern. For example, if the distinct leaf–climate character of Australian vegetation is related to nutrient-poor soils, lack of frost tolerance, evergreen leaf habit, and/or phylogenetic isolation (Jordan, 1997; Greenwood et al., 2004), any fossils that use an Australia-specific calibration must fit within this relatively narrow phylogenetic and ecological space. We find with our fossil floras that application of regional calibrations typically leads to cooler MAT estimates than the global calibration (Table 1), and that these are more at odds with independent evidence (see ‘Application of digital leaf physiognomy to fossil record’ section). The regional-based estimates are thus more precise, but may be less accurate.

An advantage of a global calibration for fossil applications is that it increases the likelihood that the appropriate biological and ecological information has been captured, although it may also lead to the incorporation of information not applicable to some fossil floras. For example, the biggest difference between the Jacobs (2002) compilation and our calibration of leaf area is at wet sites. Our data show a much wider range in site-mean leaf area at high MAP, regardless of temperature. That is, some of the warmest, wettest sites have comparatively small leaves (e.g. sites from Colombia, Australia, and Hawaii and Florida, USA, circled in Fig. 3), demonstrating that small leaves at wet sites are not always driven by the confounding influence of cool temperature.

There are two possible reasons for the discrepancy between our calibration and the Jacobs (2002) compilation. First, our data contain many sites that are both wetter and drier than the compilation of Jacobs (2002). Second, although the Jacobs (2002) compilation includes sites from Africa, Asia, and Central and South America, many of the sites are from a few discrete areas (e.g. 35% of sites are from Costa Rica and Bolivia). Our calibration includes a greater phylogenetic, geographic and climatic diversity of sites, and probably better reflects the global range of leaf size.

The trade-off with a global calibration is that any single regional signal, which could be important in a fossil application, is diluted through the inclusion of extra-regional sites. Clearly, if sufficient phylogenetic and ecological information is available, approaches that take this information into account would be preferred. We consider our global models to be important, but conservative, first steps for digital leaf physiognomy because a global approach captures the widest range of information and accounts for floras with mixed phylogenetic histories, such as extinct species that are related to extant taxa living in both the northern and southern hemisphere.

Digital leaf physiognomy models The standard error of the best MAT multiple linear regression model that can be applied to fragmentary fossil leaves is ± 4.0°C (\( r^2 = 0.70 \), \( P = 10^{-23} \)) (Table 2). Compared with the leaf-margin analysis equation derived from the same 92 sites (± 4.8°C), our model represents a moderate improvement in precision. The multivariate MAT model incorporates the percentage of untoothed species, the number of teeth : internal perimeter,
and Feret’s diameter ratio. Both of the tooth variables are probably related functionally to accelerating growth early in the growing season in cooler climates (Royer & Wilf, 2006). Feret’s diameter ratio decreases in warmer climates; that is, leaves typically become longer than they are wide as MAT increases. This negative correlation most likely allows leaves to better shed heat in warm climates (e.g. Givnish, 1984). None of these three variables significantly correlate with MAP, even after accounting for the covariation of MAT (Table S3).

The multiple linear regression MAP model \( (r^2 = 0.27, P = 10^{-6}; \text{SE} = 0.60) \) is somewhat more precise than the univariate leaf-area analysis MAP model \( (r^2 = 0.23, P = 10^{-6}; \text{SE} = 0.61; \text{Table 2}) \). For example, the error for the Fox Hills fossil flora using digital leaf physiognomy is +116/−64 cm but with leaf-area analysis is +125/−68 cm (Table 1); the errors are asymmetric because both methods estimate \( \log_e(\text{MAP}) \). The multivariate MAP model incorporates \( \log_e(\text{leaf area, mm}^2) \), \( \log_e(\text{number of teeth : internal perimeter}) \) and \( \log_e(\text{perimeter ratio}) \). Both perimeter ratio and number of teeth : internal perimeter negatively correlate with MAT (i.e. leaves are less toothy at higher rainfalls); the functional basis for this response is not known (see the Introduction). Leaf area increases with MAT, a leaf trait that is functionally related to water loss (Parkhurst & Loucks, 1972). Of the three variables in our MAP model, two also correlate significantly with MAT after controlling for MAP with partial correlation \( (\log_e(\text{perimeter ratio}): r^2 = 0.36, P < 0.001; \log_e(\text{number of teeth : internal perimeter}): r^2 = 0.24, P < 0.01; \text{Table S3}) \). This raises the possibility that our paleo-MAP estimates are affected by the confounding influence of MAT.

To gauge the accuracy of our models, MAT and MAP were estimated at each site using the regression based on the other 91 sites (i.e. a jackknife-type approach). For MAT, the standard error of the estimates was smaller for the multivariate model than for leaf-margin analysis (4.0 vs 4.8°C). Furthermore, a paired sample \( t \)-test indicates that the absolute values of the deleted residuals are significantly smaller in the multivariate model \( (P = 0.02) \). Our multivariate MAT model is thus more accurate and precise than a similarly-calibrated leaf-margin analysis equation. The patterns for MAP are less convincing. The standard error of the estimates is marginally smaller for the multivariate model than for leaf-area analysis (0.60 vs 0.61), and a paired sample \( t \)-test indicates that the absolute value of the deleted residuals are smaller in the multivariate model, but not significantly so \( (P = 0.10) \). Thus, our MAP model is somewhat more precise, but not significantly more accurate than the univariate leaf-area analysis; further, two of the variables are confounded by the influence of MAT. For these reasons, it is not clear whether our MAP model is worth the additional processing effort relative to leaf-area analysis. In summary, neither our model nor leaf-area analysis are particularly good at estimating MAP.

### Application of digital leaf physiognomy to fossil record

We applied our multivariate models to 10, well-studied, latest Cretaceous to Eocene fossil floras (Table 1). We emphasize that the climate estimates presented here are provisional until the potential confounding effects already discussed (especially phylogeny and leaf habit) are more fully accounted for. Nonetheless, we feel an initial application of this new approach is warranted and demonstrates its promise.

First, we used CCA as an initial quality check for our fossils. If a fossil site plotted outside the range of the calibration data, then it occupies uncalibrated physiognomic space; we did not attempt to reconstruct climate from such sites. All fossil sites plotted within our calibrated space except Bonanza (Fig. S4). Bonanza may be an outlier because it mixes two habitats, a lowland lake margin and an upland distal to the lake margin (MacGinitie, 1969). Also,
among fossil sites, Bonanza has the highest estimated mean leaf mass per area, suggesting a mix of both evergreen and deciduous species (Royer et al., 2007), whereas the other sites were likely composed of a higher percentage of deciduous taxa (Fig. S5). As discussed earlier, leaf habit may influence leaf–climate correlations. For these reasons, we currently do not advocate using leaf physiognomy to reconstruct paleoclimate at Bonanza.

Mean annual temperature estimates made using our leaf-margin analysis equation for the Williston Basin floras are c. 10°C (± 4.8°C; Table 1), which is cooler than expected for three reasons. First, high-latitude deep-sea temperatures were c. 10°C at this time (Zachos et al., 2001) and are incompatible with low-elevation, mid-latitude MATs of c. 10°C. Second, the presence of palm fossils in floral zone Williston Basin 1 (Peppe, 2009, 2010) suggests a MAT > 10°C (Larcher & Winter, 1981; Sakai & Larcher, 1987; Wing & Greenwood, 1993; Greenwood & Wing, 1995). Third, crocodilian fossils are present throughout the Paleocene sequence in the Williston Basin and across the Western Interior of North America, implying a MAT of ≥ 14°C (Markwick, 1998). The digital leaf physiognomy estimates for the three Williston Basin floral zones are, on average, 5.5°C warmer than leaf-margin analysis estimates (Table 1). These estimates, which are all ≥ 15°C (± 4.0°C), are in better agreement with the independent evidence cited above.

The warmer temperatures with digital leaf physiognomy are mostly driven by the low teeth : internal perimeter values, which negatively correlate with MAT (Fig. S6). The percentage of toothed species at these three sites is quite high (c. 75%), which accounts for the cool MAT estimates with leaf-margin analysis, but most of the toothed species have small and few teeth. Thus, these floras demonstrate the usefulness of incorporating climatically meaningful physiognomic variables and provide strong support for the digital leaf physiognomy approach.

The MAT estimate for the Fox Hills flora using digital leaf physiognomy is over 6°C warmer than with leaf-margin analysis (21.6, ± 4.0°C vs 14.8, ± 4.8°C; Table 1), and is more compatible with independent MAT estimates based on δ18O of shallow-water marine invertebrates from the adjacent, contemporaneous Fox Hills Seaway (18.0°C; Carpenter et al., 2003). As with the Williston Basin floras, the warmer estimate is largely driven by a low teeth : internal perimeter ratio (Fig. S6).

In the case of the Hubble Bubble flora from the PETM in the Big Horn Basin, independent evidence from the basin suggests a warming (Koch et al., 2003; Fricke & Wing, 2004; Wing et al., 2005; Secord et al., 2010) and drying (Kraus & Riggins, 2007; Smith et al., 2009a) during the PETM. Digital leaf physiognomy produces an MAT estimate that is 2.4°C warmer than leaf-margin analysis (Table 1), and thus is in slightly better agreement with the expected temperature increase during the PETM (Fricke & Wing, 2004). The warmer estimate is again driven primarily by the flora’s low teeth : internal perimeter (Fig. S6).

Several lines of evidence are consistent with the Cerrejón flora being a tropical rainforest, including the presence of a large-bodied snake (Head et al., 2009) and soft-shelled turtles (Cadena et al., 2010), as well as the climatic affinities of the nearest living relatives of several Cerrejón plant taxa (Doria et al., 2008; Herrera et al., 2008; Gómez-Navarro et al., 2009; Wing et al., 2009b). The digital leaf physiognomy estimates of MAT and MAP support a tropical rainforest interpretation and are wetter and considerably warmer than estimates from the univariate approaches (Table 1, Fig. S4). We note that our MAP estimate is somewhat drier than the leaf-area analysis estimate of Wing et al. (2009b) (264, ±217/–119 cm vs 324, +140/–98 cm), but this is because they used the more regional leaf-area analysis regression of Wilf et al. (1998).

The digital leaf physiognomy estimates for Republic are similar to the univariate model estimates (Table 1), which broadly agree with some independent evidence (Wolfe & Wehr, 1987) but are cooler than estimates based on the species composition of the flora (c. 12–13°C, Greenwood et al., 2005). The MAT and MAP estimates for the two southern hemisphere floras, P. Loros and Laguna del Hunco, are similar to estimates from univariate approaches, but are cooler and drier than expected (Table 1; see also Fig. S4). For example, the presence of a species of Papuacedrus in the Laguna del Hunco flora (P. prechilensis) suggests that the flora was fairly warm and wet (Wilf et al., 2009). This discrepancy may be due to the phylogenetic histories of the floras (see earlier discussions). Because we have few sites from southern South America in our calibration, we may not have fully characterized the physiognomy–climate space of this region.

Implications and future directions

Our study demonstrates the promise of using leaf–climate correlations in a multivariate context for reconstructing MAT and MAP from fossil floras. Digital leaf physiognomy has three major advantages over the traditional univariate and multivariate methods. First, the physiognomic variables are mostly continuous, highly reproducible, and are functionally linked to climate. Second, digital leaf physiognomy is somewhat more precise than global univariate approaches, offering the potential for more refined climate reconstructions. Third, and perhaps most importantly, climate estimates for fossil floras made using digital leaf physiognomy are typically warmer and wetter, and much closer to independent climate evidence than other leaf–climate approaches. Digital leaf physiognomy thus offers the potential for better understanding ancient greenhouse climates. However, there is room for improvement; in particular, more calibration sites from Europe, Africa, southern South America, Oceania, and the tropics are needed to
increase phylogenetic diversity. Most critically, a quantitative assessment of the impacts of leaf habit and phylogeny (and their interaction) on leaf physiognomy is required so that ecologically and phylogenetically informed calibrations can be developed.

Acknowledgements

Work at Wesleyan was supported primarily by the National Science Foundation (NSF) (grant EAR-0742637 to DLR). Funding for the Patagonia fossil collections (Laguna del Hunco and P. Loros) was supported by NSF and the National Geographic Society (grants DEB-0345750, DEB-0919071, and NGS 7337-02 to Peter Wilf and others). We thank Wesleyan students C. Ariori, A. Bobman, C. Coleman, G. Doria, S. Kim, O. Korol, E. Mendelsohn, M. Moody, J. Schroder, S. Schwarz and S. Wicaksono for help with photography and image processing, N. Cáneo, P. Wilf, P. Puerta, L. Canessa, M. Caffa, E. Ruigomez, R. Horwitt, K. Rega, E. Perkins for assistance with the Patagonian fossils, S. Gunter for help with photography, K. Wilson, I. Schönberger, J. Cruickshank and L. van Essen for help pulling herbarium sheets, D. Warton for helpful discussions about statistics, P. Resor for GIS help, R. Spicer for information about CLAMP sites, M. Lyon for leaf images, L. Hickey, S. Hu and P. Sweeney for help collecting and identifying herbs, K. Saleh for assistance identifying specimens from Malaysia, the Nahueltripay family for land access to Laguna del Hunco, the Brown, Clark, Davis, Hanson, Krutzfeld, Van Daele, Walser and Weinreiss families, the Horse Creek Grazing Association and the United States Forest Service for land access to the Williston Basin localities, the North Dakota Department of Transportation for permission to excavate the Fox Hills locality, the Stonerose Interpretive Center for access to the Republic locality, the Bureau of Land Management for access to the Bonanza site, D. Greenwood, an anonymous reviewer and D. Ackerly for comments that improved this manuscript, and especially P. Wilf for his intellectual support during early phases of the project and for comments on manuscript drafts.

References


Peppe DJ. 2003. Fox Hills I, a new Upper Maastrichtian megafossil zone within the Williston Basin of North Dakota. BS thesis, St Lawrence University, Canton, NY, USA.


Wing SL, Herrera F, Jaramillo CA, Gómez-Navarro C, Wilf P, Labandeira CC. 2009b. Late Paleocene fossils from the Cerrejón Formation, Colombia, are the earliest record of neotropical rainforest. *Proceedings of the National Academy of Sciences, USA* 106: 18627–18632.


**Supporting Information**

Additional supporting information may be found in the online version of this article.

---

**Fig. S1** Relationship between site-mean leaf area and mean annual temperature for all calibration sites.

**Fig. S2** Correlation between the percent of untoothed species in a flora and mean annual temperature for edaphically wet and dry sites.

**Fig. S3** Relationship between site-mean leaf physiognomy and mean annual temperature for sites along local water availability gradients.

**Fig. S4** Canonical correspondence analysis plot of all calibration and fossil sites based on leaf physiognomy.

**Fig. S5** Box-and-whisker plots of leaf mass per area for nine of the 10 fossil floras used in study.

**Fig. S6** Box-and-whisker plots of number of teeth to internal perimeter and Feret’s diameter ratio for all fossil floras, except Bonanza.

**Table S1** Climate variables and site-mean physiognomic variables for all calibration sites.

**Table S2** Definitions of physiognomic and climatic variables used in study.

**Table S3** Correlations between site-mean physiognomic variables and climate variables.

**Table S4** Physiognomic scores for specimens from calibration sites.

**Table S5** Regression models for predicting mean annual temperature and precipitation for 92 calibration sites using all physiognomic variables.

**Table S6** Site-mean physiognomic variables for fossil sites.

**Table S7** Physiognomic scores for specimens from fossil sites.

**Notes S1** Geography and sampling methods for calibration sites.

**Methods S1** Protocol for digitally processing fossil leaves.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.