

Climate, leaves, and the legacy of two giants

Paleobotanical novices have no simple script to follow for using angiosperm fossil floras to test climatic hypotheses in the geological record. Many methods and approaches have been used, none of which can be verified with iron-clad independent methods. This is the current nature of the field. In this issue of *New Phytologist*, Peppe *et al.* (pp. 724–739) advance the latest cycle of recalibration of taxon-free leaf-climate methodology, using an expanded set of modern sites. The research is energized by nearly 100 calibration sites and the adoption of ‘digital leaf physiognomy’. Thus, three upgrades have occurred: an analytic method for data capture, a geographic expansion, and a move to multiple linear regression. The global calibration, compared to previous applications of digital leaf physiognomy (Royer *et al.*, 2005), was achieved by adding some of Jack A. Wolfe’s (1993) worldwide sites to earlier datasets, plus *c.* 30 new sites collected by the authors.

‘Had communication between these two relentless, empirical botanists occurred, what discussions might have ensued?’

For almost a century, the most common means for reconstructing Cenozoic paleotemperature was via leaf fossils from clastic sediments. The simple correlation between the proportion of species with nonentire (toothed/serrated) margins in a fossil flora and the mean annual temperature (MAT), was used as an estimate of paleo-MAT. Angiosperm paleobotanists were in the driver’s seat with respect to estimates of paleoaltitude (via temperature lapse rates with elevation), paleogeographic reconstructions, and predicted response to climate change. However, as methods of deriving temperatures from oxygen isotopes, mammal and reptile climatic limitations, paleo-circulation patterns, and even phytoplankton distributions were refined and re-calibrated, discrepancies showed up between the MAT predicted by paleobotanists and other, leaf-margin independent, evidence. Previously complacent to use established techniques to reconstruct paleoclimate, paleobotanists sought to reconcile the discrepancies by refining the paleobotanical methods.

Cycles of improved approximations by calibration → complacency → reinvestigation → recalibration, etc. have been in motion since Wolfe (1978) first quantified the

method of taxon-free temperature estimation in the 1970s. In fact, his first data were based on those from a half-century earlier: those of Bailey & Sinnott (1915, 1916). Initially, Wolfe’s intention was to free paleoclimate estimates from a dependency on taxonomic identification of each leaf, which could be only as accurate as the individual making the taxonomic determination. Motivated by frustration with small samples and single characters, Wolfe (1993, 1995) subsequently created the first multi-site, multi-character database for evaluating morphological correlation of leaves with climate. Both database and analysis are referred to as CLAMP (Climate Leaf Analysis Multivariate Program). The assumption was that, because leaves are under intense selection to perform optimally, all dicotyledonous angiosperm species from a similar climate regime will tend to converge on a unified solution to the triple problems of capturing sunlight, regulating temperature, and conserving water.

Wolfe acknowledged that leaves of some species may not be easily molded by their environment (Hickey & Wolfe, 1975; Wolfe, 1993). Instead, he specified that large numbers of woody dicot species (> 20) be used to calibrate models and to reconstruct paleotemperatures (Wolfe, 1995). However, the method was so alluringly simple, that some researchers, including Wolfe, were tempted to reconstruct paleotemperature based on small numbers of species: the species available in fossil floras were all they had! Controversy continued over whether the best reconstructions of paleoclimate could be achieved with a combination of leaf characters, or whether a univariate method was sufficient. The large number of subjectively scored characters in the CLAMP database was proposed as one reason that the simple character of nonentire vs entire leaf margins predicted MAT with greater accuracy than the combination of characters of CLAMP. In addition, when global CLAMP data were tested against regional data, be they local, hemispheric, or continental, the regional correlations were more precise than global correlations (e.g. Teodoridis *et al.*, 2011). In this issue of *New Phytologist*, Peppe *et al.* attribute high regional correlations to phylogenetic history, although this might well be due to ecological similarity of available habitats. Few researchers have sought to change the basic approach: correlation of leaf characters in modern climate space is used to predict paleotemperature using fossil leaves (Spicer *et al.*, 2009). When estimates from fossil floras do not agree entirely, all estimates are reported (e.g. Roth-Nebelsick *et al.*, 2004).

A novel approach was employed to solve the problem of subjectively-scored character states (Huff *et al.*, 2003; Royer *et al.*, 2005). By digitally capturing the shape and size of leaves from geographic areas distinct from those originally used by Wolfe, algorithms of shape capture programs could be tested and applied to characters more objectively. ‘Digital leaf physiognomy’ was capable of placing different

leaf tooth attributes on a continuum, rather than leaving them to the scoring decisions that a weary researcher might make. Further, this approach highlighted the presumed physiological advantages that specific traits conferred and included only those traits. For example subtle, irregularly-spaced teeth may serve a leaf less effectively under transpirational dynamics than large, regular teeth, and accordingly they would be scored differently. Reducing the total number of characters scored, combined with standardizing the measurements, should have made it possible to create credible paleotemperature estimates. However, the differences between nonleaf generated paleotemperatures (Markwick, 1998; Zachos *et al.*, 2001; Fricke & Wing, 2004) and those from leaves have not been erased.

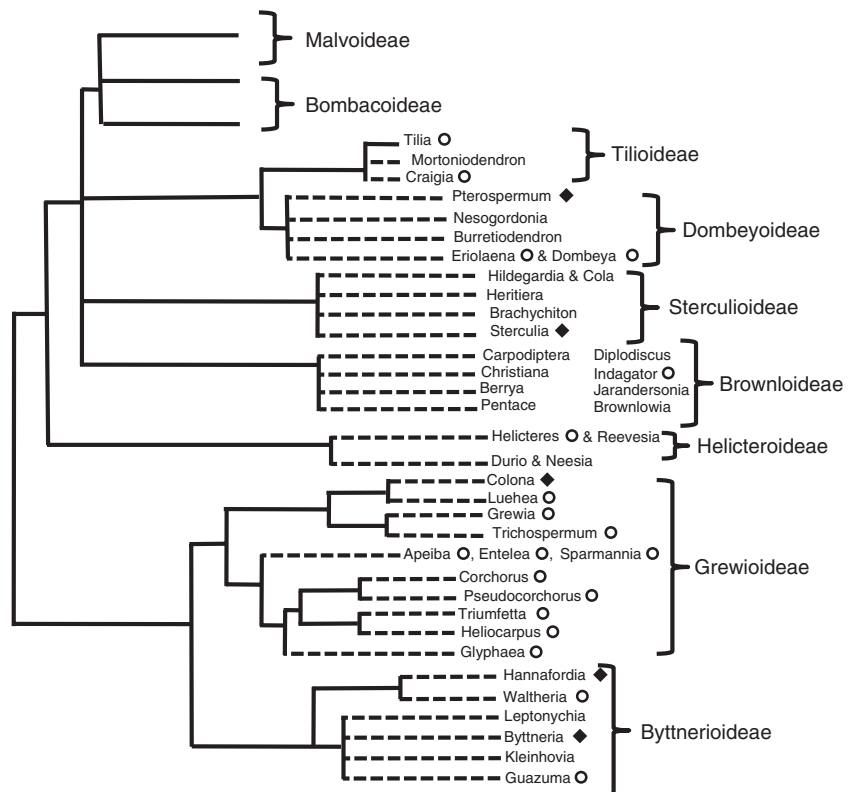
Similar efforts to reconstruct precipitation from fossil leaves have run parallel to MAT methods (e.g. Wilf, 1997). Predictions of precipitation from fossil leaves have been more readily acknowledged as approximations, perhaps because of the clear tradeoffs for leaves in investments among photosynthesis, herbivore defense, desiccation resistance, and leaf construction. Models to approach the tradeoffs through the leaf economics spectrum have been proposed (Royer *et al.*, 2007), with a surprising correlation between the petiole width of a leaf and the leaf mass per area (M_A). This however did not solve the precipitation conundrum.

The oldest method for climate reconstruction from leaves is the 'nearest-living relative' method and it has also been

upgraded and quantified to create comparable terminology and methods (Mosbrugger & Utescher, 1997; Roth-Nebelsick *et al.*, 2004). One drawback is that the method is slow and difficult because each fossil species must be confidently identified, which is tricky without attachment of leaves to stems, details of surface features like glands or hairs, or presence of reproductive parts. Another is that the method is rooted in the concept that species of closely related plant lineages share climatic tolerances. While this might be true, it is also true that speciation often occurs by the very separation in space that a climatic barrier would provide. Most angiosperm paleobotanists are, in practice, working with a combination of methodologies, using one to illuminate the others. For example, the freezing tolerances of palm species are well-defined, with only a few species capable of surviving freezing temperatures (Larcher & Winter, 1981). The presence of a palm in a fossil flora immediately suggests a relatively high mean annual temperature and a relatively long growing season (Walther *et al.*, 2007). This type of observation is added to a taxon-free analysis to exclude extremes.

These methods stand on the shoulders of Jack A. Wolfe, an individual with the foresight to voucher his intensive worldwide collections, at a time when such vouchering was only conceivable for herbarium-quality specimens. At the same time that Wolfe gathered data on the relationship between leaf form and climate, Alwyn H. Gentry was

Fig. 1 Malvaceae phylogeny (after Alverson *et al.*, 1999), coded showing margin condition and climatic distribution, demonstrates that, in this family, leaf margin condition is not related to the mean annual temperature (MAT) in the species range. However, many species are found in dry tropical areas. Dashed lines leading to terminal taxa indicate tropical distributions (no frost), with higher MAT. Solid line to terminal taxa indicates temperate distribution (only *Tilia*). Open circles following genus indicates nonentire leaf margins; closed diamonds following genus indicates both nonentire and entire leaf margins within genus; no symbol indicates genus composed of species with entire margins. Coincidence of dashed line and circle following generic name show genera that depart from the assumptions underlying the leaf-margin climate estimation methods.



amassing data on woody plant density and species composition worldwide. Gentry censused and collected modern plant biodiversity data from 226 sites in temperate and tropical forests using 0.1-ha transects. Had communication between these two relentless, empirical botanists occurred, what discussions might have ensued? To our knowledge, they were not in communication, even though they were both known to haunt the world's herbaria. Gentry's phenomenal drive to capture the world's plant biodiversity ran parallel to Wolfe's search for the perfect correlation between leaf form and climate. Both were deeply involved in angiosperm phylogenetics, although neither made phylogeny the centerpiece of their work (Hickey & Wolfe, 1975; Gentry, 1990). Three sites from Gentry's data made it into the data set used by Peppe *et al.*, an addition that seems long overdue. However, the data used by Peppe *et al.* include 17 sites with 20 or fewer species, which could have been exchanged for more of the sites from Gentry's data (Phillips & Miller, 2002).

Another splash was made recently by Little *et al.* (2010), who demonstrated that climate-leaf analysis can be influenced by 'nonrandom phylogenetic signal'. Their work, using explicitly phylogenetic models, demonstrates that the ancestry of plant species influences the presence and abundance of teeth. Similarly, significant phylogenetic effects on leaf veins were found by Walls (2011), almost simultaneously. If true, then the great precision and accuracy sought by Peppe *et al.* using taxon-free methods may not be possible. For example, if species of Sapindaceae bear foliar serrations, it may not be due to an extended winter, rather to phylogenetic history. In an overview of clades of the Malvaceae (s.l.), we found that 78% of 571 species surveyed bear nonentire leaf margins, and yet 91.2% have tropical distributions. Speciose clades, like the Grewioideae, largely encompass species with nonentire-margined leaves, whose distributions are almost entirely tropical and subtropical (Fig. 1). Expectations based on adaptive convergence would have predicted exactly the opposite pattern. Notable among this particular example is the frequent presence of nonentire margined clades in both moist tropical and dry tropical climates (not temperate). This detailed survey of a single large subfamily reiterates the conclusions offered by Little *et al.* (2010): early branching angiosperm clades bear high proportions of nonentire leaf margins, a pattern that will certainly bias interpretation of paleoclimate.

The advances and limitations of the current state of leaf-based climate reconstruction are clearly evident in Peppe *et al.* and Little *et al.* (2010). The cycle of small improvements and testing should be interrupted at this point with a thorough evaluation of the influence of phylogenetic history on calibration and interpretation of paleoclimate models. This would best include all characters employed thus far, including those for which a specific physiological interpretation has been lacking. We anticipate that the next

improvement in leaf-climate correlations will expand the sites used to evaluate the phylogenetic signal, but more importantly will suggest specific means for reducing or removing the nonenvironmental signal.

Robyn J. Burnham^{1*} and Gayle S. Tonkovich²

¹Museum of Paleontology, University of Michigan,
Ann Arbor, MI 48109-1079, USA;

²Department of Botany, The Field Museum, 1400 S. Lake
Shore Dr, Chicago, IL 60605-2496, USA

(*Author for correspondence: tel +1 734 647 2585;
email rburnham@umich.edu)

References

- Alverson WS, Whitlock BA, Nyffeler R, Bayer C, Baum DA. 1999. Phylogeny of the core Malvales: evidence from ndhF sequence data. *American Journal of Botany* **86**: 1474–1486.
- Bailey IW, Sinnott EW. 1915. A botanical index of Cretaceous and Tertiary climates. *Science* **41**: 831–834.
- Bailey IW, Sinnott EW. 1916. The climatic distribution of certain types of angiosperm leaves. *American Journal of Botany* **3**: 24–39.
- Fricke HC, Wing SL. 2004. Oxygen isotope and paleobotanical estimates of temperature and $\delta^{18}\text{O}$ -latitude gradients over North America during the Early Eocene. *American Journal of Science* **304**: 612–635.
- Gentry AH. 1990. Floristic similarities and differences between southern central America and upper and central Amazonia. In: Gentry AH, ed. *Four neotropical rainforests*. New Haven, CT, USA: Yale University Press, 141–160.
- Hickey LJ, Wolfe JA. 1975. The bases of angiosperm phylogeny: vegetative morphology. *Annals of the Missouri Botanical Garden* **62**: 538–589.
- Huff PM, Wilf P, Azumah EJ. 2003. Digital future for paleoclimate estimation from fossil leaves? Preliminary results. *Palaios* **18**: 266–274.
- Larcher W, Winter A. 1981. Frost susceptibility of palms: experimental data and their interpretation. *Principes* **25**: 143–152.
- Little SA, Kembel SW, Wilf P. 2010. Paleotemperature proxies from leaf fossils reinterpreted in light of evolutionary history. *PLoS ONE* **5**: e15161.
- Markwick PJ. 1998. Fossil Crocodylians as indicators of Late Cretaceous and Cenozoic climates: implications for using palaeontological data in reconstructing palaeoclimate. *Palaeogeography, Palaeoclimatology, Palaeoecology* **137**: 205–271.
- Mosbrugger V, Utescher T. 1997. The coexistence approach – a method for quantitative reconstructions of Tertiary terrestrial palaeoclimate data using plant fossils. *Palaeogeography, Palaeoclimatology, Palaeoecology* **134**: 61–86.
- Peppe DJ, Royer DL, Cariglino B, Oliver SY, Newman S, Leight E, Enikolopov G, Fernandez-Burgos M, Herrera F, Adams JM *et al.* 2011. Sensitivity of leaf size and shape to climate: global patterns and paleoclimatic applications. *New Phytologist* **190**: 724–739.
- Phillips O, Miller JS. 2002. *Global patterns of plant diversity: Alwyn H. Gentry forest transect data set*. St. Louis, MO, USA: Missouri Botanical Garden Press.
- Roth-Nebelsick A, Utescher T, Mosbrugger V, Diester-Haass L, Walther H. 2004. Changes in atmospheric CO₂ concentrations and climate from the Late Eocene to Early Miocene: palaeobotanical reconstruction based on fossil floras from Saxony, Germany. *Palaeogeography, Palaeoclimatology, Palaeoecology* **205**: 43–67.

- Royer DL, Wilf P, Janesko DA, Kowalski EA, Dilcher DL. 2005. Correlations of climate and plant ecology to leaf size and shape: potential proxies for the fossil record. *American Journal of Botany* **92**: 1141–1151.
- Royer DL, Sack L, Wilf P, Lusk CH, Jordan GJ, Niinemets U, Wright IJ, Westoby M, Cariglino B, Coley PD *et al.* 2007. Fossil leaf economics quantified: calibration, Eocene case study, and implications. *Paleobiology* **33**: 574–589.
- Spicer RA, Valdes PJ, Spicer TEV, Craggs HJ, Srivastava G, Mehrotra RC, Yang J. 2009. New developments in CLAMP: calibration using global gridded meteorological data. *Palaeogeography, Palaeoclimatology, Palaeoecology* **283**: 91–98.
- Theodoridis V, Mazouch P, Spicer RA, Uhl D. 2011. Refining CLAMP – investigations towards improving the Climate Leaf Analysis Multivariate Program. *Palaeogeography, Palaeoclimatology, Palaeoecology* **299**: 39–48.
- Walls RL. 2011. Angiosperm leaf vein patterns are linked to leaf functions in a global-scale data set. *American Journal of Botany* **98**: 244–253.
- Walther G-R, Gritti ES, Berger S, Hickler T, Tang Z, Sykes MT. 2007. Palms tracking climate change. *Global Ecology and Biogeography* **16**: 801–809.
- Wilf P. 1997. When are leaves good thermometers? A new case for Leaf Margin Analysis. *Palaeobiology* **23**: 373–390.
- Wolfe JA. 1978. Paleobotanical interpretation of Tertiary climates in Northern Hemisphere. *American Scientist* **66**: 694–703.
- Wolfe JA. 1993. A method of obtaining climatic parameters from leaf assemblages. *United States Geological Survey Bulletin* **2040**: 1–71.
- Wolfe JA. 1995. Paleoclimate estimation from Tertiary leaf assemblages. *Annual Review of Earth and Planetary Sciences* **23**: 119–142.
- Zachos J, Pagani M, Sloan L, Thomas E, Billups K. 2001. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* **292**: 686–693.

Key words: climate, leaf margins, mean annual temperature, nonentire, paleobotany, phylogeny.

Letters

Can publication bias affect ecological research? A case study on soil respiration under elevated CO₂

Literature surveys are a prerequisite to enhance scientific knowledge as they allow us to separate systemic from idiosyncratic mechanisms and processes, and thus provide insight at a higher level than can be gained from individual studies. In ecological research, statistical synthesis of literature surveys, using meta-analysis, has become a powerful tool to quantify global mean responses to a changing climate (Curtis & Wang, 1998; Medlyn *et al.*, 1999; Rustad *et al.*, 2001; Treseder, 2004, 2008; Knorr *et al.*, 2005; de Graaff *et al.*, 2006; Luo *et al.*, 2006; Janssens *et al.*, 2010). Such analyses have greatly improved our understanding of ecosystem functioning and the parameterization of models. Unfortunately, there is also a downside to quantitative review methods such as meta-analysis, as they can easily be affected by publication bias (Møller & Jennions, 2001). Publication bias can be defined as the selective publication of articles showing certain types of results over those showing other types of results. The most commonly suspected publication bias is the tendency for authors and journals to only publish studies with statistically significant results, which has been termed the ‘file-drawer problem’ (Rosenthal, 1979). Moreover, researchers are under increas-

ing pressure to publish frequently, and it is much easier to publish results that can easily be explained or support widely accepted hypotheses (Jarvis *et al.*, 2001), than having to fight a time-consuming battle with conservative and suspicious referees (as they perhaps should be).

A clear example of such a publication bias occurred in a European network of CO₂-enrichment experiments on trees (ECOCRAFT, 1999). From the 19 experiments that measured the response of soil respiration to elevated CO₂, only one-third were published. Across these 19 experiments the mean stimulation of soil respiration was 9% (Fig. 1a). In elevated CO₂, more C is allocated belowground to fine roots and mycorrhizae (Ceulemans *et al.*, 1999; Pendall *et al.*, 2004), and therefore an increase in soil respiration is the expected response (Zak *et al.*, 2000). However, we found that elevated atmospheric CO₂ concentrations enhanced soil respiration only in 11 out of 19 experiments (Fig. 1a), despite enlarged root systems in 15 experiments. Surprisingly, among the six published studies on soil respiration in this network, five showed that the expected increase in soil respiration and the mean overall response across these six experiments was a statistically significant 45% increase in soil respiration (Fig. 1b). Of the eight experiments with a negative response, only one was published in a conference proceedings book (Le Dantec *et al.*, 1997). The other negative responses were either never written up and submitted for publication because the researchers had difficulties believing or explaining their results, or were rejected by reviewers who did not accept such unexplainable, negative responses. Therefore, it is likely that many similar, negative responses from outside