

A. Project Summary

The purpose of this project is to develop and test a methodology for characterizing forest ecosystems that will allow direct comparison of fossil and modern forest stands. We use the form and venation of angiosperm leaves as proxy metrics for ecology because comparable leaf-architectural data are easily obtainable from imperfectly preserved or published fossil and modern floras, and because a strong *prima facie* case has been made for the presence of an ecological signal in leaf morphology. Using a test set of eight floras, of which six are modern and two fossil, we have generated two sets of numerical data based on the two main published systems of leaf architectural analysis—Climate Leaf Analysis Multivariate Program (CLAMP), and Compendium Index Categories (CICs). Phenetic analysis of these numerical descriptions of floras reflects the actual ecological relationships among the modern floras, and allows ordination of the fossil floras within the multi-dimensional space defined by the modern floras. We are seeking support in order to code a database of approximately 150 modern leaf floras. This will provide us with a statistically significant training set of modern floras of known ecology against which the approximately 250 fossil floras already coded in the Compendium Index of North American Fossil Plants can be calibrated. Using the results of this calibration, we can provide a standard method of reporting paleoecological data on leaf litter assemblages. Thus, for the first time, both paleobotanists and ecologists will have a common metric for comparing the data of their respective fields. This will be significant for paleobotanists, who have hitherto lacked a readily visualized way of comparing forest ecosystems on a meso- or macroscopic scale, as well as for ecologists who are developing an increasing awareness of the importance of the fossil record in determining long-term evolutionary patterns and rates of global change.

C. Project Description

Introduction: the need for diachronic data in ecology

In the past forty years the major focus of Anglo-American ecology has been on mathematical modeling of short-term population dynamics; the changes taking place on a time scale of more than a few decades or centuries have not been emphasized until very recently. In the past decade, however, the concept of macroecology (Brown, 1995) and more generally, the realization that ecological dynamics may be of importance at long time-scales has been brought to the fore. In particular, the special issue of *Science* magazine dated July 27, 2001 was entitled ‘Ecology Through Time’ and contained a number of articles written by ecologists searching for ways of dealing with ecological change on the order of 10^3 through 10^7 years.

At the same time, the direction of paleontology has been heavily influenced by the diversity curves for the marine invertebrate fossil record published by Raup and Sepkoski (Sepkoski et al., 1981; Raup and Sepkoski, 1984) and the focus of the Chicago School on macroevolution at the longest possible time scales. Because of the nature of the terrestrial record, measurements of terrestrial diversity through time (e.g., Niklas et al. 1985) and patterns of evolutionary change have received less attention than the marine record, but there have been recent attempts (Behrensmeyer et al. 1992) to provide an overview of terrestrial ecosystems through time. There has also been an increasing awareness amongst paleontologists of the importance of ecology in consideration of these long-term patterns and trends. The field of ‘evolutionary paleoecology’ is acquiring adherents (Allmon and Bottjer, eds., 2001) and several papers in the past five years (Patzkowsky, 1999; Cohen, 1998; Jablonski and Sepkoski, 1996) have dealt with the importance of understanding ecological relationships at relatively long time scales. The attempt at synergy between ecological and paleoecological perspectives even appears in the 2002 mission statement of the incoming editors of the journal *Paleobiology*, which actively solicits ‘contributions from paleobiologists that address questions of importance in neobiology’ (Pandolfi and DiMichele 2002:172).

In the past, ecologists have studied only the distributions of ecosystems that currently appear on the planet and from this synchronic picture, essentially a single slice of geological time like the final frame of a movie, they attempt to explain patterns like the altitudinal and latitudinal gradients in diversity, ecosystem responses to climatic change, or the presence of recognizable communities. The data provided by paleontologists are frequently of such low quality that they cannot be directly compared with modern ecological data, but poor as their resolution may be, they provide a dimension that cannot be obtained except from the fossil record. Our proposal is intended to provide a showing of the movie reel itself, albeit at a lower resolution.

We intend to do this by characterizing forest ecosystems and representing them as numerical vectors (as will be described below) so that multivariate statistical and classification tools can be used to ordinate fossil floras in the n -dimensional space defined by a modern training set of floras. In addition, iconic representations of these numerical vectors will allow rapid assessment of patterns at a meso- to macroscopic scale. By meso- to macroscopic scale, we mean that we are concerned with temporal dynamics happening on an order of thousands to millions of years and spatial patterning ranging from an order of kilometers up to biogeographical (continental) scales. Because the average accumulation rate of a stratigraphic

bed in terrestrial environments provides our smallest general time unit and the forest stand (or fossil collection locality) are our units of spatial analysis, we are not concerned with shorter time-scales and smaller areas. This approach—quantitative description of locality data in order to reveal medium to large-scale patterns—is entirely general, but we have restricted ourselves to forest ecosystems, both modern and ancient for the following reasons.

First of all, whether their dominance is measured by biomass or diversity, dicot forests are the most important contemporary terrestrial plant ecosystems. Secondly, unlike other important ecosystems such as grasslands, they have a long fossil record (over 100 ma) and constitute the bulk of the angiosperm fossil record (Wing and Dimichele, 1992; Davies-Vollum and Wing, 1998). Thirdly, Burnham et al. (1992) and Johnson (1989) have shown experimentally that the leaf-litter beds that constitute the majority of the dicot fossil record quantitatively reflect the local species abundances in the communities from which they were obtained, indicating that the ecological signal in most fossil leaf assemblages should not be overwhelmed by taphonomic biases. Fourthly, the organ generally fossilized, the leaf, is one of the most environmentally plastic plant organs. This has been argued by Cronquist (1968) on theoretical grounds, and has been supported empirically by the work the Unit of Comparative Plant Ecology at Sheffield University (see Grime, et al. 1988), which has shown that the autecology of a plant and its evolutionary strategy are predictable (in statistical terms) from its functional attributes. This has been developed into a methodology called Functional Interpretation of Botanical Survey (FIBS) (Hodgson, 1991) that relates characteristics like leaf area to parameters of the climate and environment in which the plant is found. Many of the characters used by FIBS are not recoverable from fossil material, but the close relationship between ecological survey data (equivalent to sedimentological context of fossil assemblages) and functional attributes of the plant has been emphasized by the Sheffield plant ecologists. Additionally, Givnish (1986) has recovered the basic architecture of many dicot leaves from simple mathematical models of mechanical support, incident radiation, and transpiration. He argues that the leaf is particularly strongly selected for certain functional and structural attributes and since a central tenet of evolutionary biology is that the form of an individual will reflect adaptation of its functionally important characteristics to the environment, we can expect to obtain a strong environmental signal in the architecture of an assemblage of leaves. The assumption that the vegetative body of a plant is more liable to be environmentally plastic than its reproductive organs is also implicit in the traditional basis of angiosperm systematics towards floral structure. Fourthly, the wide application of leaf margin analysis (which will be discussed in greater detail below) provides strong experimental support for the covariance of leaf morphological and environmental variables.

The importance to global change of dicot forests, their good fossil record, and the demonstrated covariance of leaf structure with ecology make them an ideal group of organisms for our proposed research: since quantitative description of a temporal series of floras constitutes diachronic ecological data, we intend to develop an ecologically meaningful way of quantitatively describing floras that can be used to compare fossil and modern plant communities. Before discussing the details of how we will produce and analyze these quantitative descriptions of floras, we will briefly review other attempts to extract ecological information from fossil leaf assemblages.

Previous Work

Before 1960, most descriptions of fossil leaf floras consisted of lists in which fossil leaves were identified on the basis of gross morphological similarity to modern species; ecological interpretations were based predominantly on analogies to modern taxa. During the 1970s, several systematic methods of describing fossil leaf architecture without relying on analogies to modern species were proposed, of which Hickey's (1973, 1979) has been most widely applied. Subsequently, Johnson (1989, 1996) and the Leaf Architecture Working Group (Ash et al. 1999) developed Hickey's (1979) architectural classification of leaves into a standard methodology for 'morphotyping' fossil floras. This methodology is now being applied to Cretaceous and Paleogene floras throughout the western USA (e.g. Gemmill and Johnson, 1997).

Morphotype descriptions emphasize the characteristics of leaves that are easily recognized and unambiguous so that data can be collected from poorly preserved material or under field conditions. Therefore, from the perspective of repeatability, morphotypes represent a great improvement over the subjective and hence generally untrustworthy identifications of the preceding century. Their disadvantage, however, is that they are not of use for comparison with the general botanical and paleobotanical literature and no methodology has been developed for comparing morphotyped floras across studies or outside of the region (the western U.S.A.) where the method has been generally applied.

A second development that took place during the 1970s was an increased concentration on the sedimentary context in which leaf assemblages were found and especially in the ecological inferences that could be drawn from sedimentology (Hickey and Doyle 1977, Hickey 1980, Scheihing and Pfefferkorn, 1984; Wing, 1984; Spicer and Wolfe, 1987; Wing 1987; Davies-Vollum and Wing, 1998). The particular advantage of this method of inferring paleoecology is that it does not rely on analogy to modern taxa and is therefore entirely independent of floral (taxonomic) identifications.

Based on these two developments (uniform criteria for morphotyping and paleoecological information derived from sedimentary context) there has been some synthetic work based on the small amount of high-quality data that has been collected so far. For instance, Wing et al. (1995) categorized and synthesized an estimated 60–80% of the published data on floral diversity from the Paleocene and Eocene from the western USA and this data base has made possible studies of long-term trends and patterns on a regional scale (e.g. Johnson and Hickey, 1990; Wing et al. 1995), which show a degree of temporal resolution and stratigraphic and evolutionary control that can be favorably compared with portions of the marine invertebrate record. But vastly more data have been collected than have been published or synthesized. One of the reasons that synthesis has lagged behind data collection is that taxonomic identifications and attributions to modern taxa of specimens from before the Oligocene epoch have been highly inconsistent and speculative. This means that formal description of the material by Linnean binomial—which because of the volume of material recovered would in any case be a Herculean task—is difficult; despite some careful taxonomic work by researchers such as Manchester (Crane, Manchester and Dilcher, 1990) and Johnson (1996), the availability of any overall taxonomic synthesis of fossil leaf floras, particularly of the Late Cretaceous and Early Tertiary, will be decades in coming.

Another challenge faced by this sort of synthesis is that there are few areas outside the western U.S.A. where these methods have been widely applied. In addition, the data collected

using these methods are not directly comparable with the traditional floral lists of the preceding century (which still constitute the bulk of the published literature) so that synthetic studies can not make use of all, or even a large proportion of the published data. Finally, despite the increasing amount of comparable data being collected for the western U.S.A. and more recently other areas, no framework has yet been developed for publishing these data in such a way that they can be used to address large-scale questions.

We view the current state of plant paleoecology as somewhat analogous to the state of invertebrate paleontology before the work of Raup and Sepkoski (1984; Sepkoski et al. 1981): a large body of data is available but has not yet been presented in a digestible form. Almost the only attempt that has been made to synthesize botanical data on a long time scale is the diversity curve of Niklas et al. (1985), which shows some long-term patterns of diversity in the same way that Raup and Sepkoski revealed marine invertebrate diversity patterns. Useful as these curves are, they reveal only the very longest-term patterns in diversity and have no potential for elucidating regional ecological patterns.

Hence this is a proposal to develop a methodology for integrating what are now regionally and methodologically isolated studies into the general botanical and paleontological literature. This is of significance to paleobotanists because it provides an opportunity to compare recent results to hitherto nearly useless monographs from the 19th and early 20th centuries and to ecologists because it will allow comparisons between modern and ancient floras, thereby helping to add to plant ecology a long-term temporal component whose importance has only begun to be appreciated.

Previously we mentioned two of the methods that have been employed by paleobotanists for extracting ecological data from fossil assemblages: analogy to modern taxa and sedimentary context of the fossil environment. The defect of the first is immediately evident: even if fossil leaves can be confidently referred to an extant taxon, inferring ecological information based on analogy with modern taxa (the 'floristic method' of ecological interpretation) is problematical because climatic tolerances and ecological roles are distributed irregularly across plant relationships. In other words, sister species can diverge dramatically in their ecological characteristics while unrelated species with similar gross morphologies can play the same ecological role or share climatic tolerances. The same difficulty applies to the 'nearest living relative' approach (Hickey, 1977; Tiffney, 1977), especially with older floras.

Sedimentary context provides less questionable paleoecological information but was seldom recorded in sufficient detail before the 1980s. Moreover, although sedimentology provides accurate and detailed data on autochthonous assemblages, it quickly declines in value if transport, mixing, and redeposition alter the sedimentary context in which fossil leaves are found from that in which they grew (Wing, 1987; Wing and DiMichele, 1992).

Methods of ecological reconstruction that rely on functional morphology are less subject to non-uniformitarian biases. For instance; a third method of extracting ecological (or rather climatic) information from leaf assemblages is the method of 'leaf physiognomy', including leaf-margin analysis (Wolfe 1979), multiple linear regression (Weimann et al. 1998), and the climate-leaf analysis multivariate program (CLAMP) developed by Wolfe (1993). Leaf physiognomy is an empirical method of estimating temperature and rainfall from leaf architectural properties of the woody dicot component of a flora (in particular, the proportion of entire-margined [toothless]

leaves in the flora). The general method dates back to the early twentieth century when Bailey and Sinnott (1915, 1916) observed that toothed leaves occurred preferentially in cool climates. This was later developed by Wolfe (1993) into an elaborate methodology (CLAMP) employing multivariate statistics to derive estimates of ancient climatic parameters. Since the climate estimates are based on regression of modern specimens, leaf physiognomy is potentially affected by the same problems of analogy between modern and pre-Oligocene floras, but in cases where it can be compared with other paleoclimate proxies it has been empirically justified as a technique (Wing and Greenwood, 1993, and references therein).

A recent paper by DiMichele et al. (2001) takes an approach more similar to ours: it describes a method for quantifying ‘ecomorphospace’ by coding particular taxa for a set of 22 ‘ecomorphic’ characters and then clustering them mathematically using principle components analysis and an agglomerative clustering algorithm. They recover clusters that are roughly congruent with the traditional Linnaean classes, thereby supporting our expectation that in general the Linnaean classes correlate with broadly construed ecological niches. They then use this observation to support the macroevolutionary theory that early radiations ‘fill’ niche space in a comparatively short time, leaving subsequent evolution certain niches to exploit under historical constraints just as morphological evolution fixes the body-plan early in history and canalizes subsequent modifications.

How does our proposal differ from these other attempts to derive ecological information from fossil assemblages? First of all, leaf physiognomy is a method of climate reconstruction, not a step towards characterization of forest communities: it is essentially about estimating climatic parameters, not facilitating ecological pattern recognition. No doubt many of the patterns we hope to identify will be highly correlated with climatic parameters that are identified by leaf physiognomy, but the statistical tools used by CLAMP and related methods take complex and multivariate inputs and return a few variables. Our approach is to reformat the complex relationships between ecosystems using leaf architecture as a proxy measurement so as to reveal hitherto unknown patterns in the history of plant evolution. For instance, an unanticipated pattern that appeared (see Figure 6) in our sample data set was the diagnostic presence of many entire-margined pinnately compound leaves in tropical dry forest (*sensu* Murphy and Lugo 1986) communities. Naturally with a sample of only eight floras, this pattern is not yet well supported, but it already raises the question of whether the strong signal is a result of phylogenetic contingency (the dominance of leguminous trees in such forests) or whether the pinnately compound leaf is a convergent adaptation to the hot, seasonally dry climate. We can potentially answer this question by looking for ecosystems with the same architectural signal provided by non-leguminous plants.

Although DiMichele et al. (2001) do address the question of ecological patterns through evolutionary time, their paper is not intended as a general methodology, but rather as a way of answering a specific question about the ecological influences on plant evolution. Our approach is more pattern-oriented: instead of formulating a particular one-off hypothesis and then finding a specific way to test it, we plan to develop a general way of displaying and publishing floral data so that fossil floras can be fit into the framework in which we analyze modern floras. It is known that particular communities are characterized by particular patterns of leaf architecture; our proposal is to systematize data on plant associations using leaf morphology as a proxy in the way that the early plant synecologists like Warming and Raunkiaer used the ‘physiognomic method of

ecological characterization' (i.e. ecological characterization by growth form, see Whittaker, 1962; not to be confused with leaf physiognomy) to classify communities into formations.

Details of proposed methodology

There are a number of ways to describe a leaf assemblage or flora by a numerical vector that reflects the shape and venation of the leaves in the flora of which two have already been extensively applied. The first and most broadly applied is CLAMP, which has been discussed above, and the second is the set of Compendium Index Categories (CICs) used to organize the Compendium Index of North American Fossil Plants at Yale University. The Compendium Index was the first systematic catalogue of fossil plants, started by Dorf (1940) as an aid to taxonomic identification of fossils plants. It is cross-indexed taxonomically, stratigraphically, and alphabetically, and organized according to simple categories of leaf shape and venation whose idealized forms are shown in Figure 1. Definition of these categories is based on petiole attachment, primary venation pattern, leaf-shape, and tooth type; the combinations of these variables produce 65 numerically labeled categories into which broad-leaved dicot leaves can be sorted. Though the Compendium Index was primarily intended for cataloguing North American fossils, similar or derivative schemes have been applied to other modern floras over the past two decades, including Australian (Christophel and Hyland, 1993) and Chinese (Yu and Chen, 1991) floras. These applications of the same principles of classification to very different floras shows the robustness of the method and indicate that the current Compendium Index base can easily be extended to incorporate floras worldwide. Furthermore, because the Compendium Index was designed for classification of fossils, the characteristics needed in order to assign a CIC to a specimen are easily recognized in fossil material or from imperfect illustrations in old publications. For convenience, we have used the CIC numbers (100-164) as variable names, but these numbers are not themselves significant: they are merely bins into which leaves can be sorted based on their morphology and structure.

For the purposes of this proposal, we coded eight floras using both CLAMP variables and CICs in order to compare two different procedures for producing quantitative descriptions of leaf floras. Ideally, the coding procedure used should be applicable to fragmentary fossil material and inadequate publications, easy and quick to perform and repeatable by different researchers, comparable when applied to modern and fossil assemblages, and ecologically informative. Naturally neither of these procedures fulfils all these desiderata, so our final method of coding may be a compromise between these established methodologies or may choose one or the other based on its desirable characteristics. The CICs were designed to allow easy (taxonomic) identification of fossil leaves, therefore by construction they supply the first two desiderata mentioned above: wide applicability and ease of application. CLAMP variables take much longer to code, but were chosen for their functional importance so they are more likely to provide ecologically meaningful descriptions. Below we describe the strong evidence provided by our sample data set that the ecological signals we are trying to extract are so strong that it does not matter which set of variables is ultimately chosen for the analysis.

BASIC ANGIOSPERM LEAF MORPHOTYPES

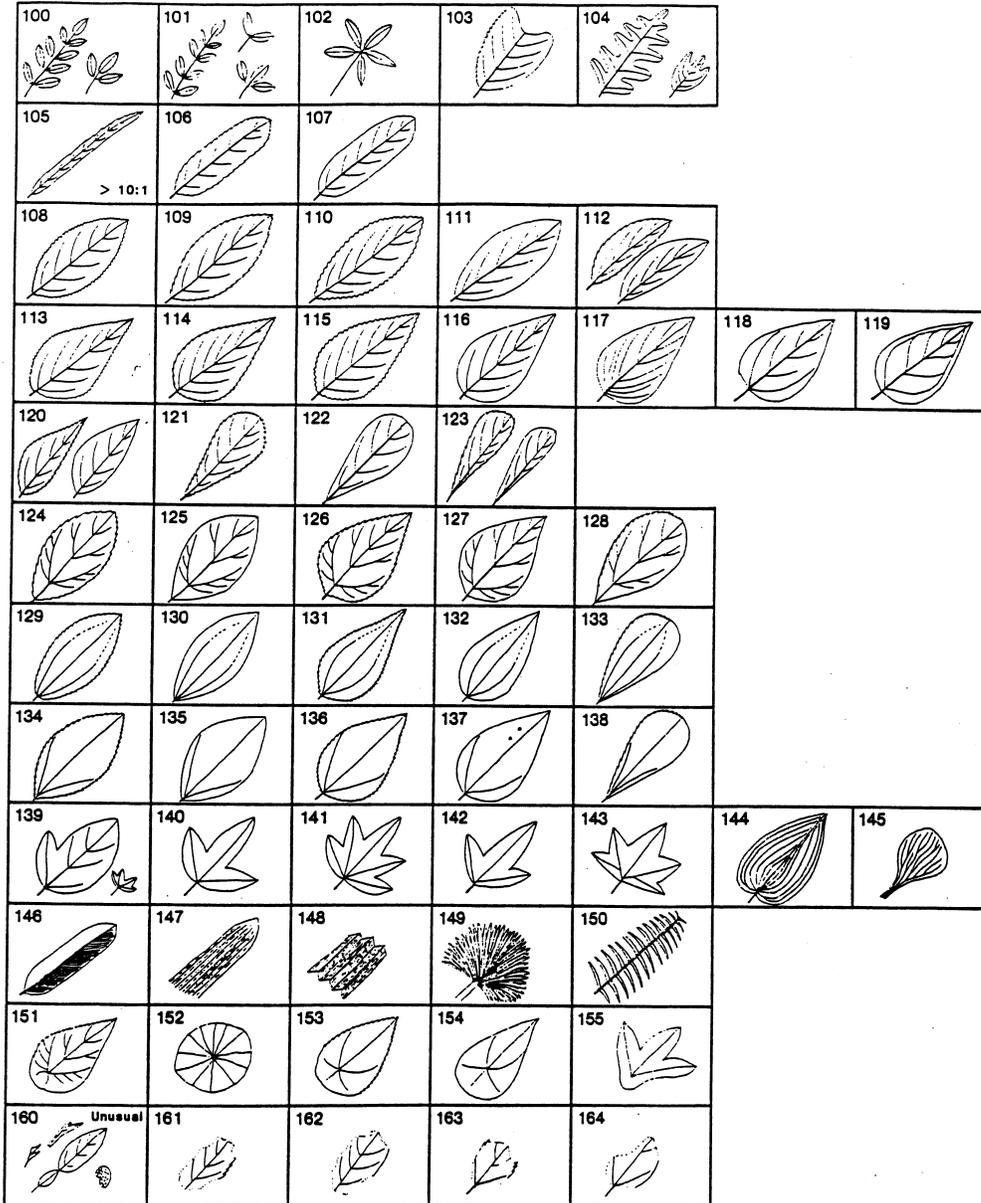


Figure 1

Figure 1: Idealized illustrations of Compendium Index Categories for dicot leaves (Ash et al. 1999)

In order to indicate exactly how such an ecological signal can be extracted and presented, we coded and analyzed a sample of eight floral lists (six modern and two fossil) drawn from the literature. The first three floras, A, B, and C were divided into canopy, understory, and total (hence, **LOWLAND NORTH CAROLINA TOTAL**, **LOWLAND NORTH CAROLINA CANOPY**, **LOWLAND NORTH CAROLINA UNDERSTORY**, etc.) and the Puryear Flora, from the Eocene of the Mississippi Embayment, was subsampled (hence, †**PURYEAR TOTAL**, †**PURYEAR SAMPLE**) so as to give us fifteen communities, whose names are shown in **SMALL CAPS**. The six modern floral lists were derived from forest stand censuses while the fossil floras (marked by the dagger symbol), are from single sedimentary localities.

Note that the fossil floras are included to demonstrate that data collected from modern and fossil communities are comparable, but this test data set consists primarily of modern floras of known ecology because we are initially interested in demonstrating that our analysis correlates with known ecological conditions. In the full study that we propose, we intend to begin with a larger set of modern floras (about 150) in order to calibrate the method with communities of known ecology and then begin coding fossil floras as they can be obtained. As soon as modern calibration is complete, the Compendium Index will provide about 250 fossil floras that have already been coded and merely require ordination with respect to our modern training set. Finally, we propose coding about 50 critical fossil floras of known sedimentology, which have been described in the past 20 years and are of particular importance in the evolution and diversification of dicot-dominated forests.

The eight floras, with subdivisions giving fifteen descriptive vectors, that make up our test set of data are:

A. Lowland (post-climax) forest of the Piedmont of North Carolina Oosting (1942) in the Deciduous Forest Formation, Oak-Pine Forest Region of Braun (1950:266) 24 spp. canopy and understory.

Three vectors: **LOWLAND NORTH CAROLINA TOTAL**, **LOWLAND NORTH CAROLINA CANOPY**, **LOWLAND NORTH CAROLINA UNDERSTORY**

B. Upland (climatic climax) forest of the Piedmont of North Carolina (*Q. alba* dominant) Oosting (1942) in the Deciduous Forest Formation, Oak-Pine Forest Region of Braun (1950:264) 21 spp. canopy and understory.

Three vectors: **UPLAND NORTH CAROLINA TOTAL**, **UPLAND NORTH CAROLINA CANOPY**, **UPLAND NORTH CAROLINA UNDERSTORY**

C. Beech-magnolia forest of Louisiana in the Deciduous Forest Formation, Southeastern Evergreen Forest Region of Braun (1950:264) 21 spp. canopy and understory.

Three vectors: **LOUISIANA TOTAL**, **LOUISIANA CANOPY**, **LOUISIANA UNDERSTORY**

D. Beech forest from Maycock (1994:374) in Mixed Mesophytic Region of Braun (1950) 21 spp. undifferentiated position in canopy

One vector: **BUCK BRIDGE KENTUCKY TOTAL**

E. Beech forest from Maycock (1994:375) in Mixed Mesophytic Region of Braun (1950) 16 spp. undifferentiated position in canopy

One vector: **CARTER CAVES KENTUCKY TOTAL**

F. Lowland forests on calcareous soils; Rio Grande valley, Cuba from Smith (1954:60) 20 spp. undifferentiated position in canopy

One vector: **CUBA TOTAL**

G. Puryear Flora, Lower Eocene of Tennessee, from Berry (1916) 238 specimens, undifferentiated position in canopy

Two vectors: †**PURYEAR TOTAL**, †**PURYEAR SAMPLE**

H. Kilgore Flora, Late Miocene of Northern Nebraska, MacGinitie (1962) 33 specimens, undifferentiated position in canopy

One vector: †**KILGORE TOTAL**

These floras were coded using the CLAMP variables in strict accordance with Wolfe's (1993) protocols, producing 29-dimensional descriptive vectors for each of the 15 communities. The only irregularity was in the case of the Puryear flora, which has over 200 species. It would have been prohibitively time-consuming to code all of these, so we instead coded a random subsample (†**PURYEAR SAMPLE**) and the CLAMP length of the vector for the whole flora (†**PURYEAR TOTAL**) was set to zero. Using the CICs, however, it was possible to code both the total flora and a random subset.

Because there is not an established methodology for using the CICs to produce a vector description of a community, we had more freedom in how exactly to obtain a numerical vector. For each of the 65 CICs, we tallied the number of species (or morphotypes) whose leaves fell into that shape category. This produced a raw 65-dimensional vector, most of whose entries were zero. This vector was then normalized to a unit length in order to control for the presence of different numbers of species in each flora.

In the case of some of the floras, percent abundance data were also available, so we experimented with calculating a leaf-shape metric equal to the raw vector scaled by the percent coverage of each species. Since, however, acceptable results were obtained with the presence/absence data, we restrict the following discussion to the normalized vectors based on the raw tallies. Standardization of the data to zero mean and unit variance also did not affect the results and therefore is ignored.

Given what we know about the ecologies of these floras, we would expect the two Kentucky beech forests (**BUCK BRIDGE KENTUCKY TOTAL** and **CARTER CAVES KENTUCKY TOTAL**) to be paired; we would expect **CUBA TOTAL** and †**PURYEAR TOTAL** to be paired (so long as they are normalized for number of specimens; otherwise †**PURYEAR TOTAL** should be the most different from the rest.) We may not have a good idea of the relationships between the North Carolina and Louisiana communities but we would expect the total floras to be close to their subsets (canopy and understory). All other things being equal, we would expect the upland and lowland North Carolina forests to be more similar to each other than either is to the Louisiana forest.

The question is: can we recover these relationships from the 15 by 65 matrices of numbers that encode our floras? In order to explore this question, we followed two statistical approaches. First, we applied a range of simple mathematical clustering algorithms to the 15

variables to obtain ‘natural’ (i.e. arbitrary mathematical) groupings in the data. Figures 2 and 3 show the trees from normalized CIC and the CLAMP matrices obtained by divisive clustering (DIANA) using the default settings in the open-source statistics package R (Kaufman and Rousseeuw, 1990).

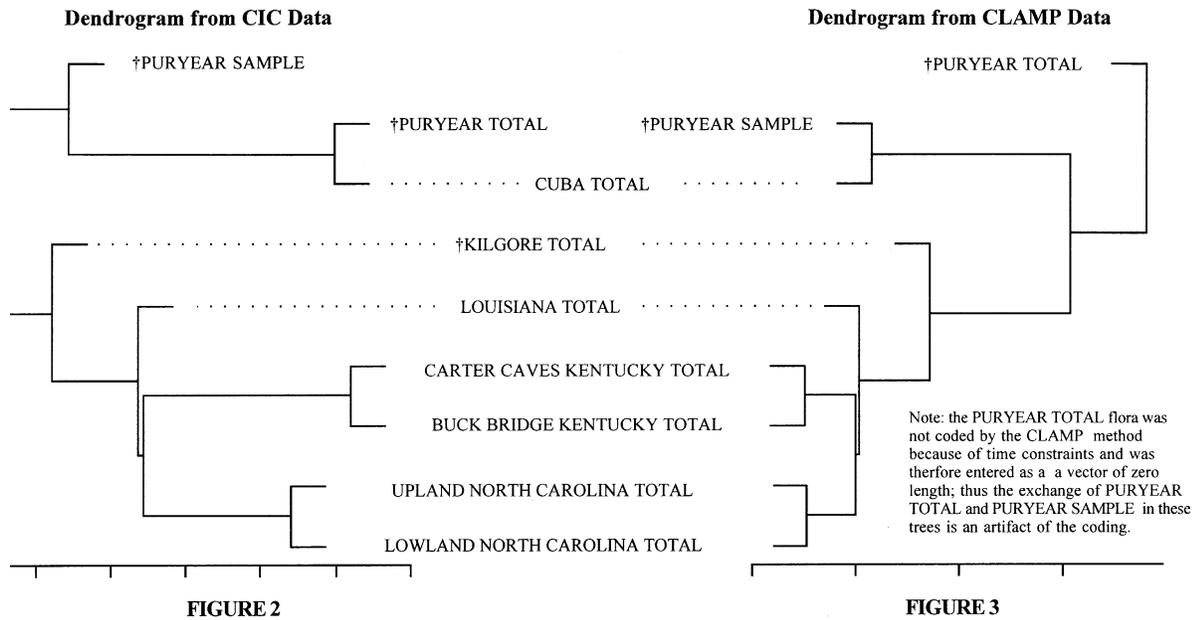


Figure 2: Dendrogram produced by a divisive clustering algorithm (DIANA) with default settings applied to normalized CIC vectors describing the eight test floras.

Figure 3: Dendrogram produced by a divisive clustering algorithm (DIANA) with default settings applied to CLAMP vectors describing the eight test floras.

Note that these trees are almost identical (the single difference is due to the subsampling in the CLAMP data), and correctly reflect the known degrees of similarity between the floras. This is a reassuring result because it shows that there are some gross mathematical relationships between our quantitative descriptions of the floras that reflect our heuristic understanding of the relationships between them. Hierarchical clustering algorithms are notoriously sensitive when applied to complex data. Other subsets of the data, however, also produce reasonable trees (Figures 4 and 5) and in order to obtain a tree that does not approximate our assumptions about the relationships between the floras, it was necessary to use the raw CIC data (not normalized to vectors of unit length). There are some statistical solutions to the sensitivity of hierarchical clustering: jackknifing and application of different algorithms to the same data can provide an estimate of the stability of the groups shown under permutation of the floras, and a larger number of floras will stabilize the axes along which variation occurs.

Dendrograms from CIC Data

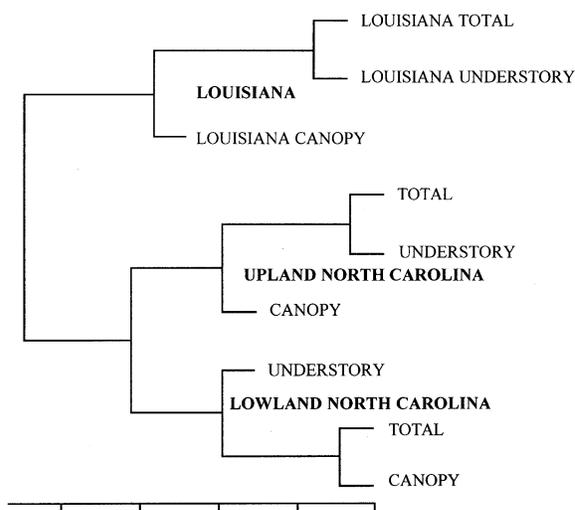


FIGURE 4

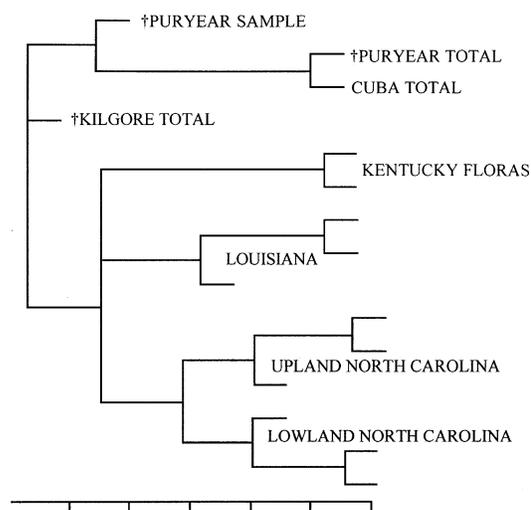


FIGURE 5

Figure 4: Dendrogram produced in the same way as Figure 2 (divisive clustering with default settings applied to normalized CIC data), but showing the relationships of the North Carolina and Louisiana floras to their subsets. Note that the relationships among the three floras are just as expected. Application to the CLAMP data produces the same tree.

Figure 5: Dendrogram produced in the same way as Figure 2, but showing the relationships of all the floras and their subsets. Relationships remain stable. Application to the CLAMP data produces a tree with minor differences. Note the stability of the method of phenetic classification under permutation of the sample.

Hierarchical clustering, however, is best used as an exploratory tool (Everitt, 1974; Hartigan, 1975) not a method of proof: it has been called ‘the data analyst’s fox hound’— frequently imprecise, but usually indicative of a general direction. More important, a clustering algorithm only measures similarity via a single arbitrary mathematical metric, it cannot do justice to the complex relationships amongst even our small set of floras. For the same reason, canonical correspondence analysis or principal components analysis would provide reduction of the variation to a few significant variables (climate, as we know, being the most significant of them). We, however, are concerned with discovering the other, biotic variables that affect ecosystem development and the way in which they interact. The techniques that statisticians have developed for addressing complicated multi-faceted relationships like these are largely visual in nature. The best-known example of this strategy is the suggestion of the statistician Herman Chernoff (1973) that such data be coded onto small face-like icons so that human pattern recognition (being applied to a familiar object, the human face), would be best able to appreciate the subtle and complex patterns in the data. Subsequently this strategy has been developed and reviewed by Kleiner and Hartigan (1981). We do not yet have access to the necessary software to construct Chernoff’s ‘faces’ or Kleiner and Hartigan’s ‘trees and castles’, but Figure 6 shows a set of simpler icons (called ‘profiles’), which display, albeit imperfectly, our data. Despite the imperfections of profiles (which can be read much like spectra) a brief examination of these icons reveals the same pattern of relationships as appeared under the clustering algorithms and which is confirmed by our knowledge of the true relationships between the floras. In this iconic representation, however, we can also see a greater degree of detail and we can begin to appreciate

the multifarious nature of the relationships between the coded floras. Most importantly, we can also read certain trends or spectral signatures directly from the representation (Figure 6): the ‘lobed oak’ line appears in most of the temperate floras while the ‘compound leguminous’ line unites the two tropical floras even though one is fossil and the other modern. These are the most obvious signatures, but many more subtle relationships can be discerned when the icons are examined closely. Moreover, because the icons are actually representations of the raw data, no complexity or dimensionality is lost in the analysis.

This kind of analysis can be applied with equal ease to field censuses, museum drawer censuses, and any publication of floral data that provides rudimentary pictures of the leaves of the species recorded. The significance of this presentation of the data is the scale at which the relationships between the floras can be appreciated: standard publication of stand or ‘relevé’ censuses crowd the exact census data from, for instance, ten floras onto a page in such a way as to make appreciation of the differences between them difficult to perceive. This is microscopic detail that is of relevance to foresters studying the exact composition of several stands in a forest but is not effective for discerning long-term spatial and temporal patterning. At the other extreme are representations like the diversity curve published by Niklas et al. (1985), which obscures all but the very grossest relationships between geological stages. The method we suggest is intended to parallel instead the scale at which, for instance, pollen diagrams represent ecological changes through time.

Logistics and Details of Proposed Work

Year 1: Calibration of analysis method using modern floral data.

PI and CPI select floras to be coded, prepare data capture forms; CPI selects publications; PI trains data collectors in coding leaf characters; checks accuracy of coding.

Two undergraduate data-entry operators code approximately 100 floras at 2 floras per work-day (average). This includes training time. One month of full-time work equivalent, each.

CPI travels to Smithsonian Institution for drawer censuses of critical floras, 3 weeks.

Field census of a target flora by PI and CPI, 1 week

Preliminary analysis of modern data by PI and CPI, 3 months

Year 2: Coding fossil floras

PI and CPI select floras to be coded, CPI selects publications; PI trains data collectors (if they are new) in coding leaf characters; checks accuracy of coding.

Two undergraduate data-entry operators code approximately 100 floras at 2 floras per work-day (average). This includes training time. One month of full-time work equivalent, each.

CPI travels to Florida State Museum and Denver Museum of Natural History for drawer censuses and data capture of already censused floras, 3 weeks.

Full data analysis under way by PI and CPI.

Year 3: Conclusion of data analysis

Full scale data analysis, PI: 1 month, CPI: 9 months, full time. Preparation of figures and text for publication. Submission of manuscript(s).

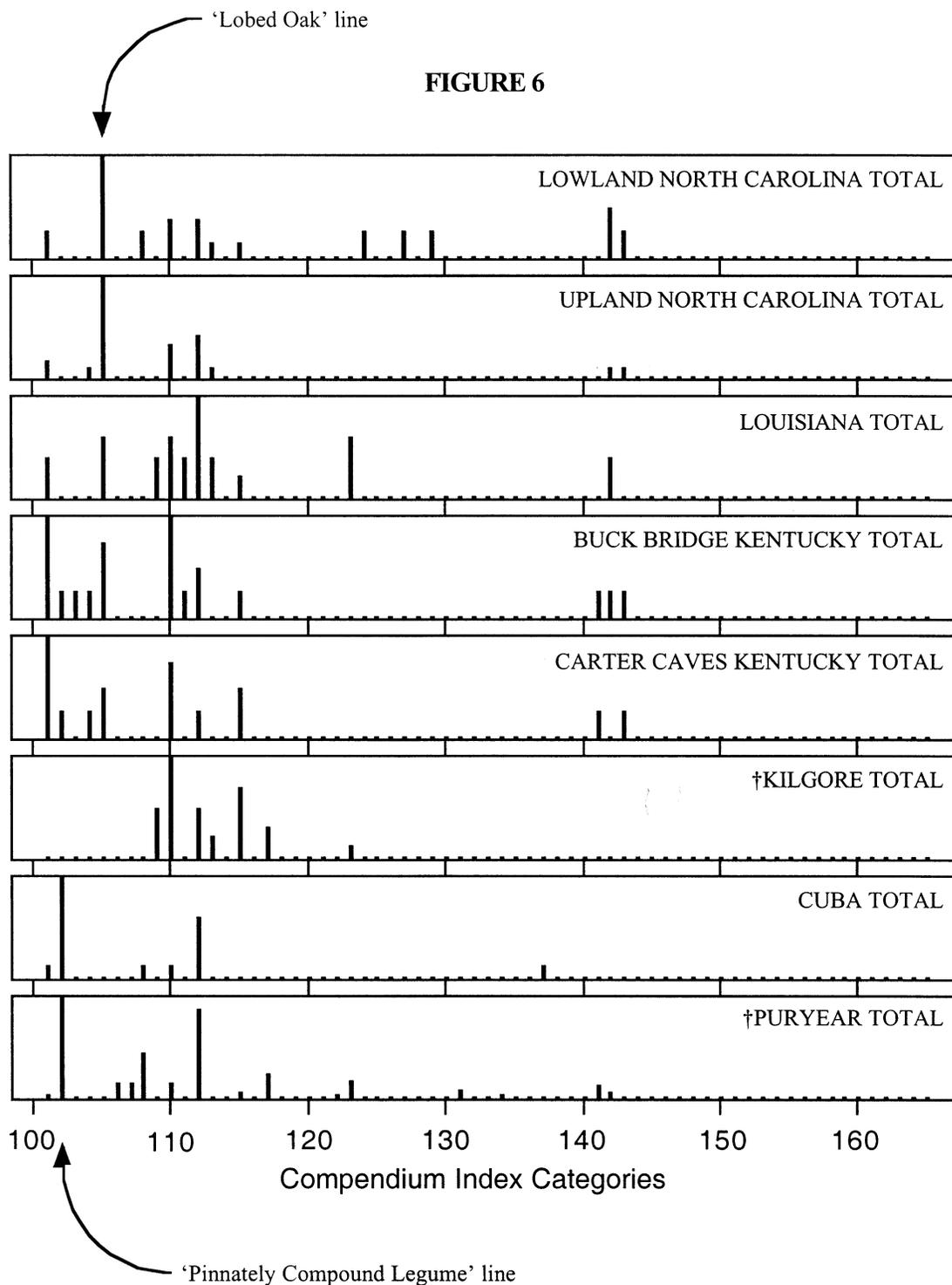


Figure 6: Profiles showing the normalized CIC data for the eight test floras. These profiles can be read as spectra, with the spectral lines representing leaf shapes that predominate in each community. Note the lines for 'lobed oaks' and for 'pinnately compound legumes'.

Broader Significance and Conclusion

We propose to produce a phenetic classification of forest stands based on leaf architectural data whereby ancient floras can be ordinated with respect to modern ones. The power of ordination as a predictive tool has been demonstrated by some of the most important works in plant community ecology of the past half century, including Grime's (1979) description of plant strategies and Whittaker's (1967) use of gradient analysis. In addition to an ordination, we will provide readily visualized representations of the data we collect that will allow recovery of complex spatial and temporal patterns that do not appear in a phenetic classification. Our method will be important to plant ecologists and to paleobotanists for different but complementary reasons, which are elaborated below, and most importantly it has the potential to draw plant ecology and paleoecology together by allowing the direct comparison of modern and ancient floras.

For plant ecologists, the initial stage of the project in which we plan to collect and analyze a modern calibration set may be of importance for its own sake because it will classify ecosystems using a set of criteria (from leaf architecture) that is particularly rich in ecological information but is not considered in existing ecosystem classifications. Therefore concordance between the classification we plan to produce and previous vegetation classifications would provide an independent test of current systems of classification. More important, however, is the opportunity for direct comparison between fossil and modern plant communities. Unlike the current standard methods of reporting paleoecological data, our method will allow direct comparison with modern floras on functional grounds that bypass taxonomic ambiguity in the fossil record and therefore it will provide plant ecology with a hitherto unobtainable diachronic set of data. In certain places, like the Late Cretaceous Meeteetse Formation of Wyoming, which contains about 80 fossil floras in a nearly continuous stratigraphic sequence of 40 ma, the fossil record provides a real temporal series of floras that can show the change in plant communities through deep time. In addition, our method will connect the microscale of site census data with the macroscale of diversity curves through geological time by emphasizing the variation in the data that is significant to ecological characterization at mesoscopic scales. It has the potential to facilitate identification equally of temporal and of spatial macroecological patterns and should allow researchers to investigate the correlation between plant community evolution and global change.

For plant paleoecologists, the value of this proposal depends more on the predictions that can be made about ancient floras based on calibration by modern floras. In the case of our sample data set, the argument that the Puryear flora was equivalent to a tropical dry forest (Dilcher, 1973) like the current lowland Caribbean forests is supported by our analysis. Within paleoecology, the need for large-scale comparison of ecosystems has been articulated very clearly: for instance, the practice of describing the ecology of fossil floras without applying these descriptions to evolutionarily interesting questions has been described as 'me-too-ecology' (Kitchell 1985), in other words, the practice of merely illustrating another example of a phenomenon that has already been described. The success of palynology in revealing regional-scale ecological patterns on a scale of centuries and millennia owes a great deal to the way in which pollen diagrams can be visually compared and correlated to stratigraphy. Our proposal has the potential to open megafloral (leaf litter) assemblages to the same sort of analysis that made palynology one of the critical tools that led to the development of modern ecology in the 1940s and 50s. Megafloral assemblages even have the advantage over pollen preparations that they are

more liable to be autochthonous and therefore inherently capable of providing higher spatial resolution. Also, there is an enormous and untapped body of published data, which would provide a return in evolutionary synthesis out of proportion to the effort that needs to be expended on primary description. Thus, we would argue that this proposal has value as a way of representing patterns that are already known in a useful way even if we do not discover patterns that are unknown.

Finally, we consider the potential of this proposal for contributing to the growing synergy between ecology and paleoecology. Even though detailed distributional and metabolic data may never be obtainable from the fossil record, time depth is increasingly being recognized as important in answering major ecological questions. We may thus have to accept the reduced resolution obtainable from fossil data and rely on signals that are so robust that they are not lost in taphonomic noise. Leaf architecture may seem to be a poor substitute for the kind of autecological knowledge that can be obtained by modern field ecologists and plant physiologists. But even an inexperienced botanist, if given a basket of leaves from a modern forest floor, would be able to say a fair amount about the community from which they were obtained. This proposed research provides a way to systematize and quantify that sort of knowledge, and our analysis of a sample data set has convinced us that the ecological signal in leaf architecture is strong enough to be significant if we examine a large enough set of floras. For instance, a question that has aroused some curiosity is the fate of the Cretaceous ‘lobed-leaf flora’, as, for instance whether or not it performed a functionally similar role to the lobed oaks and maples in modern northern temperate forests. Our approach to analyzing the leaf morphologies found in a forest can in a sense be considered a sort of functional morphology: patterns of presence and absence of certain leaf characteristics can be correlated to geographic, climatic, altitudinal, and successional variables in the way that presence of a long lever arm in a vertebrate joint can be correlated to the presence of behaviors that require high torque on the joint. In the same way, we propose to exploit the evolutionary plasticity of leaves to study the behavior of forests through time.

Arguably the most influential American plant ecologist since Clements, Robert Whittaker, believed that “no aspect of synecological science...has had a more crucial role in the evolution of ecological schools than the classification of natural communities.” (1962:2) This is as true of fossil as of extant ecosystems, and we hope that our proposed method of classifying forests on the basis of leaf architecture will both contribute to the classification of extant plant communities and begin the ordination of fossil communities amongst them.

Results of Prior NSF Support

NSF Grant DEB 9987475: Compactorization, Reorganization, and Electronic Cataloguing of the Peabody Museum Paleobotany Collections

Funding Period: November 15, 2000 – October 31, 2003; Funding Amount: \$365,347.00

Principal Investigator: Leo J. Hickey

To date, all equipment including rails, compactors, storage cases, drawers and shelving has been installed. The entire collection has been moved into the Peabody Museum’s Environmental Science Center, and over 150,000 specimens are now stored in 4,000 new drawers in mobile open drawer units and on mobile oversized shelving or fixed oversized pallet shelving. The reorganization and electronic cataloguing of the collection is underway and the project will be completed by October 31, 2003.

D. References Cited

- Allmon, W.D. and D.R. Bottjer. 2001 [eds.]. *Evolutionary Paleoecology*. New York: Columbia University Press, 357 pp.
- Ash, A., B. Ellis, L.J. Hickey, K. Johnson, P. Wilf, and S. Wing. 1999. *Manual of Leaf Architecture: Morphological Description of Dicotyledonous and Net-Veined Monocotyledonous Angiosperms*. Privately published and distributed. Smithsonian Institution, 1999, 65 pp. with CD.
- Bailey, I.W. and E.W. Sinnott. 1915. A botanical index of Cretaceous and Tertiary climates. *Science* 41(1066):831-834.
- Bailey, I.W. and E.W. Sinnott. 1916. The climatic distribution of certain types of angiosperm leaves. *Am. J. Bot.* 34:24-39.
- Behrensmeyer, A.K., J.D. Damuth, W.A. DiMichele, R. Potts, H. Sues, and S.L. Wing [eds.]. 1992. *Terrestrial Ecosystems Through Time*. Chicago: University of Chicago Press, 568 pp.
- Berry, E.W. 1916. *The Lower Eocene Floras of Southeastern North America*. USGS Professional Paper 91, 481 pp.
- Braun, L.E. 1950. *Deciduous Forests of Eastern North America*. New York: Hafner Publishing Co., 596 pp.
- Brown, J. 1995. *Macroecology*. Chicago: University of Chicago Press, 296 pp.
- Burnham, R.L., S.L. Wing, and G.G. Parker. 1992. The reflection of deciduous forest communities in leaf litter: Implications for autochthonous litter assemblages from the fossil record. *Paleobiology* 18(1):30-49.
- Christophel, D.C. and B.P.M. Hyland. 1993. *Leaf Atlas of Australian Tropical Rain Forest Trees*. East Melbourne: CSIRO, 260 pp.
- Chernoff, H. 1973. The use of faces to represent points in K-dimensional space graphically. *J. Am. Stat. Assoc.* 68(342):361-368.
- Cohen, A.S. 1998. Reflections on community ecology and the community of ecology. *Palaios* 13:603-605.
- Crane, P.R., S.R. Manchester, and D.L. Dilcher. 1990. A preliminary survey of fossil leaves and well-preserved reproductive structures from the Sentinel Butte Formation (Paleocene) near Almont, North Dakota.. *Fieldiana Geology* 20(31 December): 1-63.
- Cronquist, A. 1968. *The Evolution and Classification of Flowering Plants* Boston: Houghton Mifflin Co., 396 pp.

- Davies-Vollum, K.S. and S.L. Wing. 1998. Sedimentological, taphonomic, and climatic aspects of Eocene swamp deposits (Willwood Formation, Bighorn Basin, Wyoming). *Palaios* 13: 28-40.
- Dilcher, D.L. 1973. A paleoclimatic interpretation of the Eocene floras of southeastern North America. In A. Graham [ed.], *Vegetation and Vegetational History of Northern Latin America*. Amsterdam: Elsevier Scientific Publishing Company, pp. 39-59.
- DiMichele, W.A., W.E. Stein, and R.M. Bateman. 2001. Ecological sorting of vascular plant classes during the Paleozoic evolutionary radiation. In W.D. Allmon and D.J. Bottjer [eds.], *Evolutionary Paleocology*. New York: Columbia University Press, pp. 285-335.
- Dorf, E. 1940. An illustrated catalogue of Mesozoic and Early Cenozoic Plants of North America. *Science* 91(2368):478.
- Everitt, B. 1974. *Cluster Analysis*. London: Heinemann, 122 pp.
- Gemmill, C.E.C. and K.R. Johnson. 1997. Paleoecology of a Late Paleocene (Tiffanian) megaflora from the Northern Great Divide Basin, Wyoming. *Palaios* 12:439-448.
- Givnish, T.J. [ed.] 1986. *On the Economy of Plant Form and Function*. Cambridge: Cambridge University Press, 717 pp.
- Grime, J. P. 1979. *Plant Strategies and Vegetation Processes*. Chichester: Wiley, 222 pp.
- Grime, J.P., J.G. Hodgson, and R. Hunt. 1988. *Comparative Plant Ecology*. London: Unwin Hyman, 742 pp.
- Hartigan, J.A. 1975. *Clustering Algorithms*. New York: John Wiley and Sons, 351 pp.
- Hickey, L.J. 1973. Classification of the architecture of dicotyledonous leaves. *Am. J. Bot.* 60(1):17-33.
- . 1977. *Stratigraphy and Paleobotany of the Golden Valley Formation (Early Tertiary) of Western North Dakota*. Geological Society of America Memoir 150, 181 pp.
- . 1979. A revised classification of dicotyledonous leaves. In C.R. Metcalfe and L. Chalk [eds.], *Anatomy of the Dicots*. Oxford: Oxford University Press, pp. 25-39.
- . 1980. Paleocene stratigraphy and flora of the Clark's Fork Basin. In P.D. Gingerich [ed.], *Early Cenozoic Paleontology and Stratigraphy of the Bighorn Basin, Wyoming: 1880-1980*. Univ. Mich. Pap. Paleontol. No. 24, pp 33-50.
- and J.A. Doyle. 1977. Early Cretaceous fossil evidence for angiosperm evolution. *Botanical Review*: 43(1):3-102.

- Hodgson, J.G. 1991. The use of ecological theory and autecological datasets in studies of endangered plant and animal species and communities. *Pirineos*: 138:3-28.
- Jablonski, D. and J.J. Sepkoski. 1996. Paleobiology, community ecology, and scales of ecological pattern. *Ecology* 77(5):1367-1378.
- Johnson, K.R. 1989. A high-resolution megafloral biostratigraphy spanning the Cretaceous-Tertiary boundary in the northern Great Plains. Ph.D. Dissertation, Yale University, New Haven, CT, 556 pp. (2 vols.).
- Johnson, K.R. 1996. Description of seven common fossil leaf species from the Hell Creek Formation (Upper Cretaceous: Upper Maastrichtian), North Dakota, South Dakota, and Montana. *Proceedings of the Denver Museum of Natural History, Series 3, No. 12* (1 April): 47 pp.
- Johnson, K.R. and L.J. Hickey. 1990. Megafloral change across the Cretaceous/Tertiary boundary in the northern Great Plains and Rocky Mountains. U.S.A. *Geol. Soc. Spec. Pap.* 247:433-443.
- Kaufman, L. and P.J. Rousseeuw. 1990. *Finding Groups in Data: An Introduction to Cluster Analysis*. New York: Wiley, 342 pp..
- Kitchell, J.A. 1985. Evolutionary paleoecology: recent contributions to evolutionary theory. *Paleobiology* 11(1):91-104.
- Kleiner B. and J.A. Hartigan 1981. Representing points in many dimensions by trees and castles. *J. Am. Stat. Assoc.* 76(374):260-269.
- Maycock, P.F. 1994. The ecology of Beech (*Fagus grandifolia* Ehrh.) forests of the deciduous forests of southeastern North America, and a comparison with the Beech (*Fagus crenata*) forests of Japan pp. *In* A.K. Iwatsuki and M. M. Grandtner [eds.], *Vegetation in Eastern North America* Miyawaki. Tokyo: University of Tokyo Press, pp. 351-410.
- MacGinitie, H.D. 1962. *The Kilgore Flora* Berkeley: University of California Press, 157pp.
- Murphy, P.G. and A.E. Lugo. 1986. Ecology of tropical dry forests. *Ann. Rev. Ecol. Syst.* 17:67-88.
- Oosting, H.J. 1942. An ecological analysis of the plant communities of Piedmont, North Carolina. *Amer. Midl. Nat.* 28:1-126.
- Pandolfi, J.M. and W.A. Dimichele. 2002. From the incoming editors. *Paleobiology* 28(1):172.
- Patzkosky, M. 1999. A new agenda for evolutionary paleoecology—Or would you in the background please step forward. *Palaios* 14:195-197.

- Raup, D.M. and J.J. Sepkoski, Jr. 1984. Periodicity of extinctions in the geologic past. *Proc. Nat. Acad. Sci.* 81:801-805.
- Sepkoski, J.J., R.K. Bambach, D.M. Raup, and J.W. Valentine. 1981. Phanerozoic marine diversity and the fossil record. *Nature* 293(5832):435-437.
- Scheihing, M.H. and H.W. Pfefferkorn. 1984. The taphonomy of land plants in the Orinoco Delta: a model for the incorporation of plant parts in clastic sediments of Late Carboniferous age of Euramerica. *Rev. Palaeobot. Palynol.* 41: 205-240.
- Smith, E.E. 1954. *The Forests of Cuba* (Maria Moors Cabot Foundation Publication No. 2). Petersham, Massachusetts: The Harvard Forest, 98 pp.
- Spicer, R.A. and J.A. Wolfe. 1987. Plant taphonomy of late Holocene deposits in Trinity (Clair Engle) Lake, northern California. *Paleobiol.* 13(2):227-245.
- Tiffney, B.H. 1977. Contribution to a monograph of the fruit and seed flora of the Brandon lignite. Ph.D. Dissertation, Harvard University.
- Weimann, M.C., S.R. Manchester, D.L. Dilcher, L.F. Hinojosa, and E.A. Wheeler. 1998. Estimation of temperature and precipitation from morphological characters of dicotyledonous leaves. *Am. J. Bot.* 85(12):1796-1802.
- Whittaker, R.H. 1962. Classification of natural communities. *Botanical Review* 28(1):1-239.
- . 1967. Gradient analysis of vegetation. *Biological Reviews* 42:207-264.
- Wing, S.L. 1984. Relation of paleovegetation to geometry and cyclicity of some fluvial carbonaceous deposits. *Journal of Sedimentary Petrology* 54:52-66.
- . 1987. Depositional environments of plant bearing sediments. In DiMichele, W.A. and S.L. Wing [eds.]. *Methods and Applications of Plant Paleoecology: Notes for a Short Course*. Washington, DC.: Paleobotanical Section, Botanical Society of America, pp. 1-19.
- . 1987. Eocene and Oligocene floras and vegetation of the Rocky Mountains. *Ann. Missouri Bot. Gard.* 74:748-784.
- Wing, S.L., J. Alroy, and L.J. Hickey. 1995. Plant and mammal diversity in the Paleocene to Early Eocene of the Bighorn Basin. *Palaeogeography, Palaeoclimatology, Palaeoecology* 115:117-155.
- Wing, S.L. and W.A. DiMichele. 1992. Ecological characterization of fossil plants. In A.K. Behrensmeyer, J.D. Damuth, W.A. DiMichele, R. Potts, H.D. Sues, and S.L. Wing [eds.], *Terrestrial Ecosystems through Time*. University of Chicago Press, Chicago, pp.139-181.

Wing, S.L. and D.R. Greenwood. 1993. Fossils and fossil climate: the case for equable continental interiors in the Eocene. *Phil. Trans. R. Soc. Lond. B* 341:243-252.

Wolfe, J. 1979. *Temperature parameters of humid to mesic forests of eastern Asia and relation to forests of other regions of the northern hemisphere and Australasia*. U.S. Geol. Surv. Prof. Pap. 1106:1-37.

----. 1993. A method of obtaining climatic parameters from leaf assemblages. *U.S. Geol. Surv. Bull.* 2040:1-73.

Yu, C.-h. and Z.-l. Chen. 1991. *Leaf Architecture of the Woody Dicotyledons from Tropical and Subtropical China*. Oxford: Pergamon Press, 414 pp.