



Short Paper

Use of CART analysis to differentiate pollen of red pine (*Pinus resinosa*) and jack pine (*P. banksiana*) in New EnglandAndrew M. Barton^{a,*}, Andrea M. Nurse^b, Katelyn Michaud^a, Sarah W. Hardy^c^a Department of Biology, University of Maine at Farmington, Farmington, Maine 04938, USA^b Climate Change Institute, University of Maine, Orono, Maine 04469, USA^c Division of Mathematics and Computer Sciences, University of Maine at Farmington, Farmington, Maine 04938, USA

ARTICLE INFO

Article history:

Received 2 January 2010

Available online 29 October 2010

Keywords:

Biogeography

CART analysis

New England

Paleoecology

*Pinus banksiana**Pinus resinosa*

Palynology

Pollen

Species differentiation

ABSTRACT

The identification of fossil pollen at the generic rather than species level is hampering progress in understanding the biogeography and dynamics of paleo-vegetation. We used CART analysis to facilitate the differentiation of fossil pollen of *Pinus banksiana* and *Pinus resinosa*, which are morphologically similar and nearly always combined in paleoecological studies. The CART model, using four of the ten morphological traits measured, exhibited a high level of correct identification for pollen of each of the species and shows promise as a tool for increasing the detail of paleoecological records and inferences.

© 2010 University of Washington. Published by Elsevier Inc. All rights reserved.

Introduction

There has been remarkable recent progress in our understanding of the biogeography and dynamics of North American plant species, communities, and biomes during the late Pleistocene and Holocene (e.g., Prentice and Webb 1998; Shuman et al. 2004; Williams et al. 2004; Gonzales et al. 2008). The main tool unlocking this history is the analysis of fossil pollen, preserved in water-saturated substrates with low decomposition rates. Most palynological analyses differentiate pollen only to the generic level, for a variety of reasons, including the difficulty of distinguishing species in some taxa. Lack of species resolution is irrelevant for genera represented by only one species in a given area (e.g., *Liriodendron tulipifera*), but the majority of species in North America have multiple sympatric congeners (e.g., *Acer*, *Carya*, *Fraxinus*, *Quercus*, *Picea*, *Pinus*). Decades of ecological work have demonstrated the unique and individualistic responses of tree species to environmental variables, true even for species within the same genus (Burns and Honkala 1990; Lindbladh et al. 2003, 2007; Sawada et al. 2004; Finkelstein et al. 2006; Lindbladh et al. 2007). Our understanding of Quaternary vegetation history, therefore, is based on incomplete information and could be advanced by techniques that allow for improved pollen species differentiation.

Research aimed at distinguishing the pollen of *Pinus* species in North America has a long history (e.g., Cain 1940; Whitehead 1964; Hansen and Cushing 1973; Ammann 1977), but most studies continue to distinguish members of the genus at a subgeneric level or not at all (e.g., Williams et al. 2004; Whitmore et al. 2005). Our study focuses on *Pinus* species native to northern New England. *Pinus* pollen can be readily resolved into two subgenera, allowing straightforward identification of white pine (*Pinus strobus*), the only member of the haploxyton subgenus of the four common northeastern species. Pollen of pitch pine (*Pinus rigida*, diploxyton) is distinctive for its large size. However, pollen of the other two (diploxyton) species, jack pine (*Pinus banksiana*) and red pine (*Pinus resinosa*), are similar in shape and overlap in size. Consequently, paleoecological studies generally lump the two together (e.g., Whitehead 1964; Davis and Jacobson 1985; MacDonald et al. 1998; Williams et al. 2004; Whitmore et al. 2005). Documentation of the presence of a species in a given site can occasionally be assessed with macrofossil data (e.g., Despons and Payette 1993). Macrofossil analysis, however, requires an additional level of cost and effort, and these records do not typically represent all species present nor species relative abundance with the same degree of detail as fossil pollen.

Similar challenges have faced researchers working on other tree genera, for example, the fossil pollen of *Picea* species endemic to northeastern North America. Pollen of these three species (*Picea glauca*, *Picea mariana*, and *Picea rubens*) are very similar, and, until recently, studies identified *Picea* to only the generic level. Lindbladh et

* Corresponding author. Department of Biology, University of Maine at Farmington, 173 High Street, Preble Hall, Farmington, Maine 04938, USA. Fax: +1 207 778 7365.
E-mail address: Barton@maine.edu (A.M. Barton).

al. (2002) measured morphological traits of modern pollen of each of these species and developed a classification model using an effective and simple statistical approach: a classification and regression-tree (CART) analysis. Lindbladh et al. (2003) then applied the model to fossil *Picea* pollen to resolve the unique Holocene histories of individual species.

In this article, we examine the possibility of using CART analysis to distinguish the pollen of *P. banksiana* and *P. resinosa* from northern New England, using measurements of four quantitative and six qualitative traits.

Methods

Pollen

Because of possible regional genetic variation, reference pollen sets should be collected from within the region of paleoecologic interest (Lindbladh et al. 2002). We used *P. banksiana* and *P. resinosa* pollen from two sources: (1) pollen collected in Maine and New Hampshire, residing in the reference collection at the Paleocology Research Laboratory, Climate Change Institute, University of Maine, Orono, Maine, USA, and (2) pollen we collected directly from strobili on trees in populations of both species from across Maine in May and June of 2007. All collections were from geographically separated stands. We examined a total of 420 grains of *P. banksiana* from 9 populations (20–60 grains per population) and 560 grains of *P. resinosa* from 11 populations (40–80 grains per population).

All pollen reference sets were prepared at the University of Maine Paleocology Research Laboratory using standard procedures for paleoecological pollen samples. Pollen samples were soaked in 10% potassium hydroxide for 20 min at 90°C, treated with glacial acetic acid, and acetolyzed with acetic anhydride and sulfuric acid. Samples were then dehydrated in tertiary butyl alcohol and stored in silicone oil. Samples were not stained. Prepared slides were sealed with paraffin.

Our morphological analysis of *Pinus* pollen closely followed methodologies established for *Pinus* by Hansen and Cushing (1973) and for *Picea* by Lindbladh et al. (2002). We used only grains that were unbroken, symmetrical, and in equatorial view. An attached camera produced a digital image for each grain, and quantitative variables were measured using SPOT software. SPOT measurements were verified with an ocularmeter calibrated to a stage micrometer at 400×. A Nikon light microscope at 400× magnification was employed to evaluate qualitative variables. We evaluated four quantitative and six qualitative traits (Figs. 1 and 2). Grain length (X_1), corpus breadth (X_2), bladder width (X_3), and bladder intersection length (X_4) were measured in micrometers. The following qualitative traits were assigned ordinal values. Furrow-to-corpus transition (X_5) and cap marginal frill (X_6) were scored from one to three, with one indicating a sharp or pronounced feature and three an indistinct or gradual feature. Cap infrastructure (X_7) was scored from one to three: knotty (verrucate), reticulate, or indistinct (Fig. 1B). Furrow membrane sculpture (X_8) was scored from one to three: verrucate, scabrate, and psilate (Fig. 1B). McAndrews et al. (1973) and Moore et al. (1991) refer to verrucate sculpture of the furrow membrane as distal membrane verrucae or “belly warts.” Cap-to-bladder indentation (X_9) was scored as one (indented) to three (not indented). Bladder shape (X_{10}) was scored as one to three: spherical, hemispherical, and irregular. In less than 5% of cases, grain morphology was intermediate between full scores, and half-fractional score (i.e., 1.5 and 2.5) were given. Although we distinguished these features at 400×, magnification of 1000× may facilitate determination of some characters, especially for cap infrastructure (X_7). Examples of qualitative characters are provided in Figure 2.

Statistical analysis

Classification and regression-tree (CART) analysis provides a simple, nonparametric method for recursively classifying levels of a

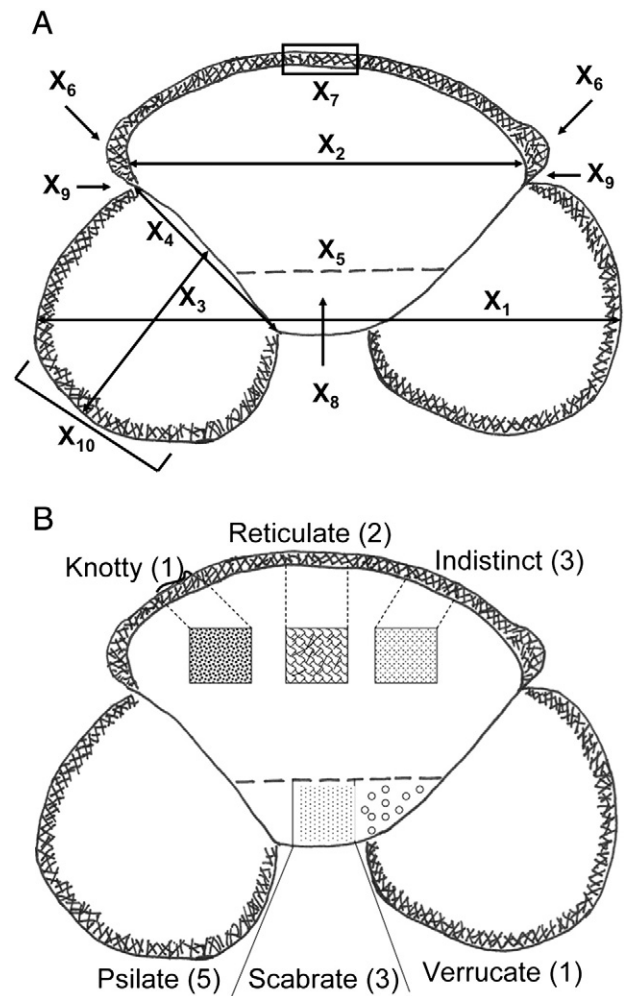


Figure 1. (A) Four quantitative and six qualitative traits analyzed: grain length (X_1), corpus breadth excluding marginal frill (X_2), bladder width (X_3), bladder intersection length (X_4), furrow-to-corpus transition (X_5), cap marginal frill (X_6), cap infrastructure (X_7), furrow membrane sculpture (X_8), cap-to-bladder indentation (X_9), and overall bladder shape (X_{10}). (B) Graphic depiction and numeric classification of traits X_7 (cap infrastructure) and X_8 (furrow membrane sculpture). Trait analyses were adapted from Hansen and Cushing (1973).

dependent variable using a set of independent variables (Brieman et al. 1984). In this case, the two *Pinus* species are classified based on pollen morphological traits. In each step of the CART analysis, the splitting criterion that minimizes the probability of misclassification, measured by the Gini index, is chosen. A 10-fold cross-validation estimate of predictive error is computed for each split. The tree is “grown” recursively, adding a new split in each step, until reaching a specified level of complexity, after which the tree is pruned back to the number of splits that minimizes the cross-validation error. The analysis, therefore, produces a classification or decision tree, much like a standard taxonomic key, allowing robust identification of pollen grains using a small set of morphological measurements (see Lindbladh et al. (2002) for more details). In order to test the accuracy of the CART tree, we reserved a test set of 100 randomly selected pollen grains for each species. The model was then built using the remaining grains (780), and the accuracy of the model was tested on the reserve data set. The R programming environment (R Development Core Team 2009) was used to carry out the CART analysis (Mairdonald and Braun 2003; Therneau et al. 2009).

In order to exclude *P. strobus* and *P. rigida*, we also carried out preliminary analyses of collected pollen from these two species (unpublished data). Pollen of haploxyton species *P. strobus* is readily distinguished from diploxyton subgenus species by the gradual, and

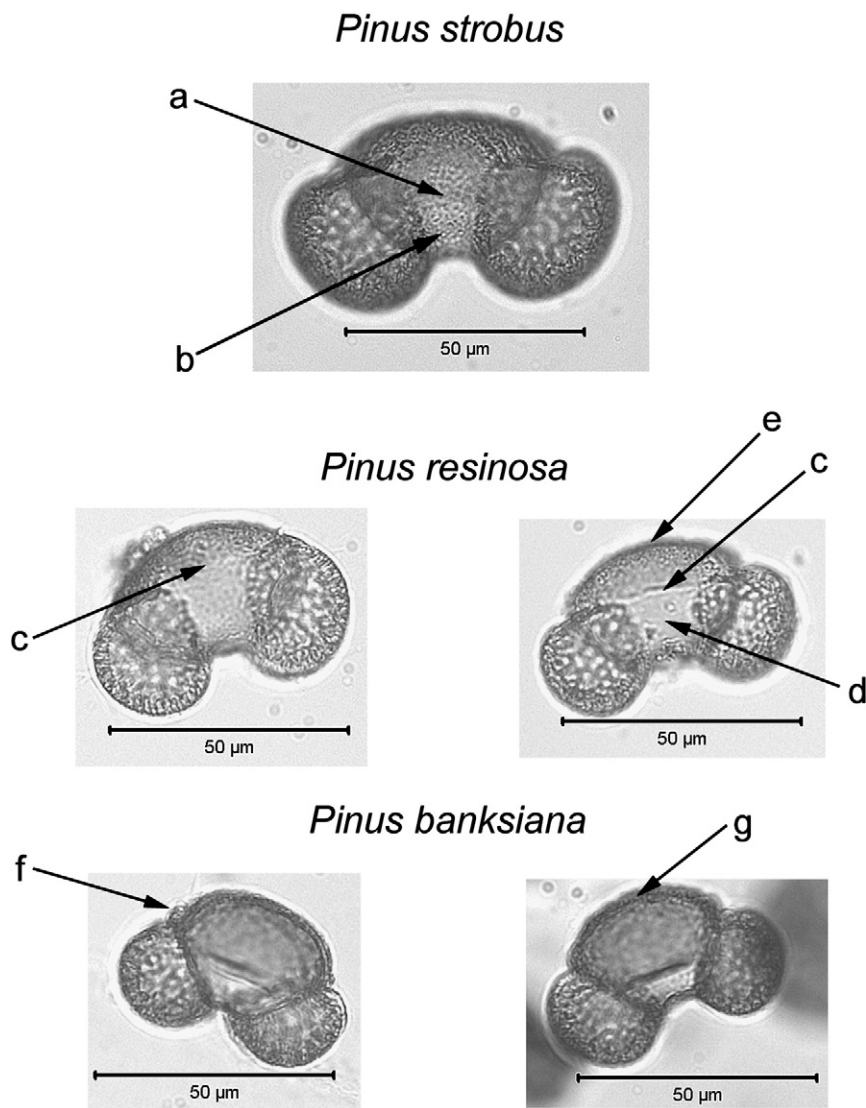


Figure 2. Photographs (400 \times) of pollen of *P. strobus*, *P. resinosa*, and *P. banksiana*, demonstrating morphological traits: (a) gradual transition between furrow membrane and cell corpus, (b) verrucate furrow membrane, (c) sharp furrow-to-corpus transition, (d) scabrata to psilate furrow membrane, (e) indistinct cap infrastructure, (f) knotty cap infrastructure, and (g) pronounced marginal frill.

often undefined, transition between the cell body (corpus) and the furrow membrane (Fig. 2a) and the presence of verrucae in the furrow membrane (Fig. 2c and d). In contrast, the pollen of the three diploxyton species exhibits a sharp transition between corpus and furrow membrane, with furrow membrane sculpture grading from psilate to weakly scabrata (Fig. 2c and d). The diploxyton species *P. rigida* was distinguished from *P. banksiana* and *P. resinosa* by its prominently larger size ($\geq 70 \mu\text{m}$). Researchers wishing to apply the methods reported here to fossil pollen samples will find the drawings in Hansen and Cushing (1973) helpful and examination of reference grains essential in evaluating within-species and between-species qualitative characters.

Results

Pollen grains of *P. resinosa* were significantly larger than those of *P. banksiana* for all of the measured dimensional variables: grain size, corpus width, bladder width, and bladder intersection (Table 1). The two species differed significantly in their frequency distributions for every qualitative trait except for furrow-to-corpus transition (Fig. 3).

The classification tree produced by CART analysis used four morphological variables to distinguish the pollen of *P. banksiana* and

P. resinosa (Fig. 4). All pollen with a large grain length ($>53.5 \mu\text{m}$) were identified by the CART analysis as *P. resinosa*. Smaller grains ($<53.5 \mu\text{m}$) with a cap infrastructure score of >2.25 , cap marginal frill score of >2.75 , and furrow membrane sculpture score of >2.25 or with a cap infrastructure of <2.25 and bladder shape of >2.75 were also identified as *P. resinosa* grains. Smaller grains ($<53.5 \mu\text{m}$) were

Table 1

Means, standard deviations, margins of error for 95% confidence intervals, ranges, and sample sizes for quantitative morphological traits for pollen of *P. banksiana* and *P. resinosa*. *T*-value is from Welch's Two Sample *t*-test ($***p < 0.001$).

Species		Grain length	Corpus width	Bladder width	Bladder intersection
<i>P. banksiana</i> (n = 420)	Mean	49.72	33.34	20.16	16.94
	SD	3.96	3.26	2.55	2.43
	95% CI	0.37	0.31	0.24	0.23
	Range	37–62	22–45	12–26.5	11–24.5
<i>P. resinosa</i> (n = 560)	Mean	58.47	37.98	23.36	18.51
	SD	5.89	4.38	3.00	2.84
	95% CI	0.49	0.36	0.25	0.23
	Range	41–77	24–53	14–37.5	10.5–31
<i>t</i>		−27.8***	−19.0***	−18.0***	−9.3***

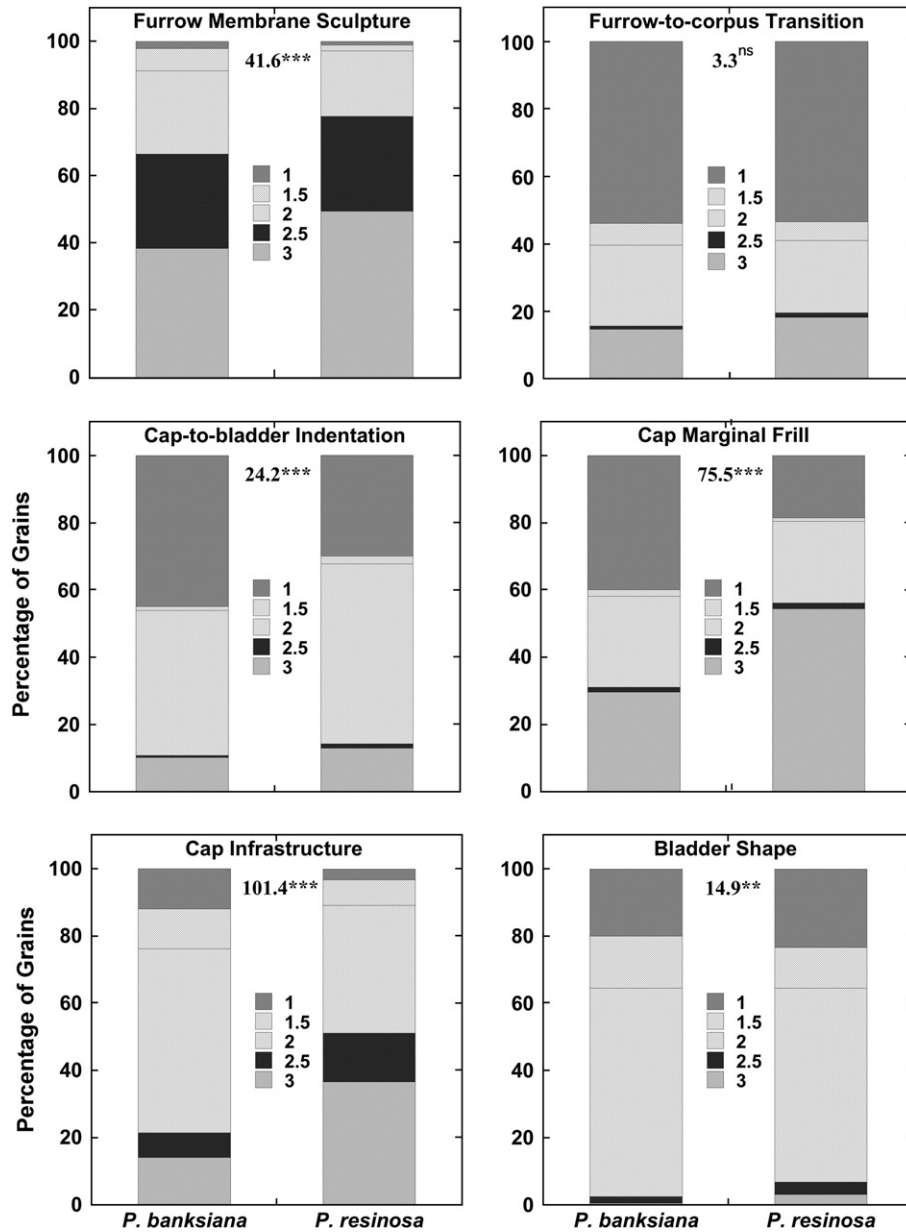


Figure 3. Frequency distribution for qualitative morphological traits for pollen of *P. banksiana* ($n = 420$) and *P. resinosa* ($n = 560$). Values are percentages. χ^2 -values, given below trait names, are from Chi-squared tests of independence (ns, $p > 0.05$; ** $p < 0.01$; *** $p < 0.001$). See **Methods** for description of morphological traits associated with scores.

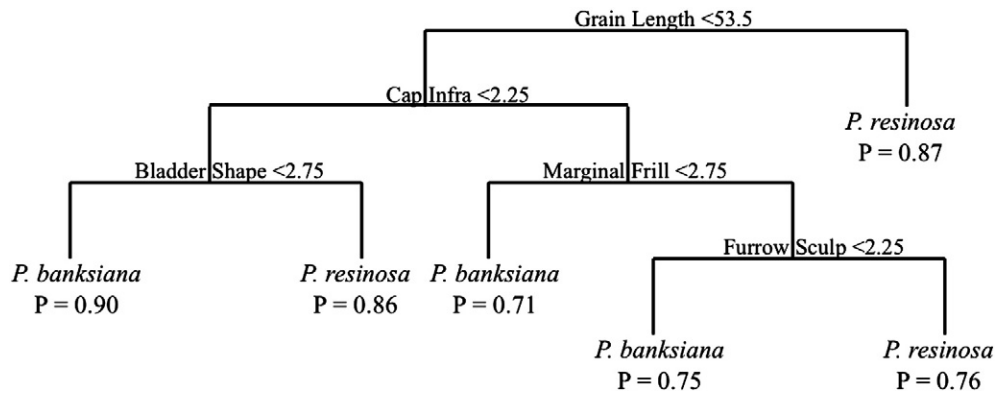


Figure 4. Classification tree produced by CART analysis of *P. banksiana* ($n = 320$) and *P. resinosa* ($n = 460$), after 100 randomly selected grains of each species were removed for a test set. P is the probability that a grain meeting the criteria of an end node is actually from the species indicated at the node.

identified as *P. banksiana* with any of three further subdivisions: cap infrastructure score of <2.25 and a bladder shape of <2.75; a cap infrastructure score of >2.25 and cap marginal frill score of <2.75; or a cap infrastructure score of >2.25, cap marginal frill score of >2.75, and furrow membrane sculpture score of <2.25. The probability that an end node (i.e., the combination of traits leading to that node) produces the predicted species varied from 71% to 90% (Fig. 4).

The developed model, using all of the data except a reserve set of 100 grains for each species, correctly predicted the identity of 86% of the pollen grains overall, 79.1% for *P. banksiana*, and 90.9% for *P. resinosa*. When tested on the reserve data set, the model exhibited an accuracy of 85% for the two species combined, 83% for *P. banksiana* and 87% for *P. resinosa* (Table 2).

Discussion

P. banksiana and *P. resinosa* migrated considerable distances during the late Pleistocene and Holocene to their present, primarily boreal, distributions (Davis 1981; Rudolf 1990; Rudolph and Laidly 1990). Consequently, the two species are of great interest to ecologists focusing on the responses of tree populations to climate change. Although the two species are similar ecologically, their current geographic distributions differ (with *P. banksiana* more northerly), and they often respond differently to environmental variables (compare Rudolph and Laidly 1990 with Rudolf 1990). A deep understanding of their species-specific responses to past climate change, however, has been hampered by the similarity of their pollen. Attempts have been made to distinguish the pollen of these species based on morphological traits. For example, Ammann (1977) found significant differences between the two species in the number and size of bladder nodules on the pollen surface. However, as Ammann (1977) and Klaus (1977) reported, potassium hydroxide, a necessary concentration step in the highly organic and clay-rich samples from northeastern North America, destroys the bladder nodules, rendering that character trait unavailable for species comparison.

While we have shown that the pollen of *P. banksiana* and *P. resinosa* differ for a number of morphological traits, no single trait allows an observer to discriminate easily between the two species, because of substantial overlap in the distributions for each of these pollen characteristics. Hansen and Cushing (1973) used eleven morphological traits to develop a numerical method for identifying *Pinus* in southwestern USA. They calculated the probability of species identity of a given pollen grain given its similarity to a reference collection of all possible pines. This was a major step beyond standard pollen taxonomic keys. The drawback of their statistical approach, however, was that it was less than transparent in operation, it required measurement of many traits, and the model successfully identified less than half of the grains examined.

Our results suggest that CART analysis may provide a solution to these challenges by combining relatively few morphological pollen traits into a straightforward classification tree that successfully identifies most grains. Our final CART classification employs four

easily assessed pollen characteristics – one quantitative and three qualitative – in a transparent model that is easy to apply to pollen identification. By following the classification tree to any of its nodes, the chance of correctly identifying any given pollen grain ranged from 79.1% to 90.9%, somewhat higher than that found for a CART model for *Picea* pollen (Lindbladh et al. 2002). Our results suggest that future studies could apply this classification tree to pollen grains in a paleoecological study to identify these two pine species with a known degree of probability, thereby facilitating species-specific analyses of their migrations and dynamics during the Holocene. Lindbladh et al. (2007) employed a similar CART model in this manner to reveal a late-glacial transition from *Picea glauca* to *P. mariana*, and the near absence of *P. rubens* during that time—results that were not possible using previous paleoecological approaches (see also Lindbladh et al. 2003).

Pollen identification at the generic rather than species level imposes three important limitations on paleoecological reconstructions (Williams et al. 2004). First, our current understanding of the stability of taxon associations (i.e., communities) and plant formations (e.g., biomes) is based in part on generic-level data that very likely hide changes in species composition over time. Second, given that vegetation responses to environmental change occur at the species level, results for genera may mask the true rates and extent of these responses. Finally, our notion of the sizes and nature of the fundamental niches of Quaternary tree species is again based partly on the behavior of genera, which undoubtedly exhibit broader niche dimensions than for individual species. Two studies provide compelling evidence that these limitations might be considerable. Lindbladh et al. (2003) for *Picea* and Finkelstein et al. (2006) for *Acer*, *Fraxinus*, and *Juglans* revealed strong heterogeneity among congeners in their responses to changing environments in late-glacial and Holocene times, which were masked by previous analyses at the generic level (Sawada et al. 2004). These are, of course, important considerations for the development of basic theory in paleoecology and biogeography, but it is difficult to imagine that these limitations are not also hampering our ability to develop reliable models informed by paleoecology that project the impacts of climate change at ecological scales ranging from individual taxa to biomes (Williams et al. 2004).

Assumptions implicit in the current study include the stability of *Pinus* pollen morphology from the time of deglaciation to the present. Additionally, the classification has not been tested outside of northern New England. Lindbladh et al. (2002) argued that, because of regional variation in morphological traits of *Picea* species, researchers in each region might need to develop their own models. A similar situation might hold for the two *Pinus* species examined here. Nevertheless, our study confirms the potential of CART analysis on multiple morphological traits for identification of fossil tree pollen to the species level, in this case for a genus for which such identification has been problematic.

Acknowledgments

We greatly appreciate financial support from The Nature Conservancy in Maine and the University of Maine at Farmington and helpful suggestions on the manuscript by Ann Dieffenbacher-Krall and two anonymous reviewers.

References

- Ammann, B.R., 1977. A pollen morphological distinction between *Pinus banksiana* Lamb. and *P. resinosa* Ait. *Pollen et Spores* 14, 521–529.
- Brieman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. *Classification and Regression Trees*. Wadsworth International Group, Belmont, CA.
- Burns, R.M., Honkala, B.H. (Eds.), 1990. *Silvics of North America*. U.S.D.A. Forest Service Agriculture Handbook 654. Washington, D.C.
- Cain, S.A., 1940. The identification of species in fossil pollen of *Pinus* by size–frequency determination. *American Journal of Botany* 27, 301–308.
- Davis, M.B., 1981. Quaternary history and the stability of forest communities. In: Botkin, D.B. (Ed.), *Forest Succession*. Springer, New York, New York, USA, pp. 132–177.

Table 2

Percentage identification accuracy of the CART classification tree for the pollen grains used to build the model and for a reserve sample to which the model was applied. Model data set was $n = 320$ for *P. banksiana* and $n = 460$ for *P. resinosa*. Test set was 100 grains for each species.

		%Identified as <i>P. banksiana</i>	%Identified as <i>P. resinosa</i>
CART model	Actual <i>P. banksiana</i>	79.1	20.1
	Actual <i>P. resinosa</i>	10.0	90.0
Test set	Actual <i>P. banksiana</i>	83.0	17.0
	Actual <i>P. resinosa</i>	13.0	87.0

- Davis, R.B., Jacobson Jr., G.L., 1985. Late glacial and early Holocene landscapes in northern New England and adjacent areas of Canada. *Quaternary Research* 23, 341–368.
- Despons, M., Payette, S., 1993. The Holocene dynamics of jack pine at its northern range limit in Quebec. *Journal of Ecology* 81, 719–727.
- Finkelstein, S.A., Gajewski, K., Viau, A.E., 2006. Improved resolution of pollen taxonomy allows better biogeographical interpretation of post-glacial forest development: analyses from the North American Pollen Database. *Journal of Ecology* 94, 415–430.
- Gonzales, L.M., Williams, J.W., Kaplan, J.O., 2008. Variations in leaf area index in northern and eastern North America over the past 21, 000 years: a data-model comparison. *Quaternary Science Reviews* 27, 1453–1466.
- Hansen, B.S., Cushing, E.J., 1973. Identification of pine pollen of Late Quaternary age from the Chuska Mountains, New Mexico. *Geological Society of America Bulletin* 84, 1181–1200.
- Klaus, W., 1977. Forschungsschwerpunkte der Paläobotanik und Palynologie der Universität Wien. *Review of Palaeobotany and Palynology* 23, 303–330.
- Lindbladh, M., O'Connor, R., Jacobson Jr., G.L., 2002. Morphometric analysis of pollen grains for paleoecological studies: classification of *Picea* from eastern North America. *American Journal of Botany* 89, 1459–1467.
- Lindbladh, M., Jacobson Jr., G.L., Schauffler, M., 2003. The postglacial history of three *Picea* species in New England, USA. *Quaternary Research* 59, 61–69.
- Lindbladh, M., Oswald, W., Foster, D., Faison, E., Hou, J., Huang, Y., 2007. A late-glacial transition from *Picea glauca* to *Picea mariana* in southern New England. *Quaternary Research* 67, 502–508.
- McAndrews, J.H., Berti, A.A., Norris, G., 1973. Key to the Quaternary pollen of the Great Lakes Region. Royal Ontario Museum Miscellaneous Publication, Toronto.
- MacDonald, G.M., Cwynar, L.C., Whitlock, C., 1998. The late quaternary dynamics of pines in northern North America. In: Richardson, D.M. (Ed.), *The Ecology and Biogeography of Pinus*. Cambridge University Press, Cambridge, England, pp. 122–136.
- Maindonald, J., Braun, W.J., 2003. *Data Analysis and Graphics Using R—an Example-based Approach*. Cambridge University Press, Cambridge, England.
- Moore, P.D., Webb, J.A., Collinson, M.E., 1991. *Pollen Analysis*, second ed. Blackwell Scientific Publications, Oxford.
- Prentice, I.C., Webb III, T., 1998. BIOME 6000: reconstructing global mid-Holocene vegetation patterns from palaeoecological records. *Journal of Biogeography* 25, 997–1005.
- R Development Core Team, 2009. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. (URL <http://www.R-project.org>).
- Rudolf, P.O., 1990. *Pinus resinosa* (Lamb.), Red Pine. In: Burns, R.M., Honkala, B.H. (Eds.), *Silvics of North America*, Vol. 1, Conifers. U.S.D.A. Forest Service Agriculture Handbook 654. Washington, D.C, pp. 442–455.
- Rudolph, T.D., Laidly, P.R., 1990. *Pinus banksiana* (Lamb.), Jack Pine. In: Burns, R.M., Honkala, B.H. (Eds.), *Silvics of North America*, Vol. 1, Conifers. U.S.D.A. Forest Service Agriculture Handbook 654. Washington, D.C, pp. 280–293.
- Sawada, M., Viau, A.E., Vettoretti, G., Peltier, W.R., Gajewski, K., 2004. Comparison of North American pollen-based temperature and global lake-status with CCCma AGCM2 output at 6 ka. *Quaternary Science Reviews* 23, 225–244.
- Shuman, B., Newby, P., Huang, Y., Webb III, T., 2004. Evidence for the close climatic control of New England vegetation history. *Ecology* 85, 1297–1310.
- Therneau, T.M., Atkinson, B., R port by Ripley B, 2009. Rpart: Recursive Partitioning. R package. Version 3.1-45. (<http://CRAN.R-project.org/package=rpart>).
- Whitehead, D.R., 1964. Fossil pine pollen and full-glacial vegetation in southeastern North Carolina. *Ecology* 45, 767–777.
- Whitmore, J., Gajewski, K., Sawada, M., Williams, J.W., Shuman, B., Bartlein, P.J., Minckley, T., Viau, A.E., Webb III, T., Shafer, S., Anderson, P., Brubaker, L., 2005. Modern pollen data from North America and Greenland for multi-scale paleoenvironmental applications. *Quaternary Science Reviews* 24, 1828–1848.
- Williams, J.W., Shuman, B.N., Webb III, T., Bartlein, P.J., Leduc, P.L., 2004. Late-Quaternary vegetation dynamics in North America: scaling from taxa to biomes. *Ecological Monographs* 74, 309–334.