

MORPHOMETRIC ANALYSIS OF POLLEN GRAINS FOR PALEOECOLOGICAL STUDIES: CLASSIFICATION OF *PICEA* FROM EASTERN NORTH AMERICA¹

MATTS LINDBLADH,^{2,4} RAYMOND O'CONNOR,³ AND
GEORGE L. JACOBSON, JR.²

²Institute for Quaternary and Climate Studies and Department of Biological Sciences, University of Maine, Orono, Maine 04469 USA; ³Department of Wildlife Ecology, University of Maine, Orono, Maine 04469 USA

Little is known about the paleoecological histories of the three spruce species (white spruce, *Picea glauca*; black spruce, *P. mariana*; and red spruce *P. rubens*) in eastern North America, largely because of the difficulty of separating the three species in the pollen record. We describe a novel and effective classification method of distinguishing pollen grains on the basis of quantitative analysis of grain attributes. The method is illustrated by an analysis of a large sample of modern pollen grains (522 grains from 38 collections) of the three *Picea* species, collected from the region where the three species co-occur today. For each species X we computed a binary regression tree that classified each grain either as X or as not-X; these three determinations for each grain were then combined as Hamming codes in an error/uncertainty detection procedure. The use of Hamming codes to link multiple binary trees for error detection allowed identification and exclusion of problematic specimens, with correspondingly greater classification certainty among the remaining grains. We measured 13 attributes of 419 reference grains of the three species to construct the regression trees and classified 103 other reference grains by testing. Species-specific accuracies among the reliably classified grains were 100, 77, and 76% for *P. glauca*, *P. mariana*, and *P. rubens*, respectively, and 21, 30, and 22% of the grains by species, respectively, were problematic. The method is applicable to any multi-species classification problem for which a large reference sample is available.

Key words: classification and regression-tree (CART) analysis; modern pollen grains; *Picea glauca*; *Picea mariana*; *Picea rubens*.

Identification of fossil pollen usually involves a combination of analyst experience and taxonomic keys (e.g., Faegri and Iversen, 1964; Moore, Webb, and Collins, 1991) that requires binary decisions based on the morphology of the grain in question. Such keys, together with modern reference material, usually allow reliable identification of most pollen taxa. The taxonomic precision of the identifications depends on the natural variability of the group involved and on the degree to which the taxonomic variability has been recognized by palynologists. In many cases, only small and subtle differences in pollen morphology exist between closely related taxa, making precise identifications difficult or even impossible, and in other difficult taxa variability is evident only through careful statistical analysis. In some cases, changes in statistical distributions through time can reveal shifts in abundance of taxa (e.g., species within a genus), even though the identification of individual grains is often problematic.

Many difficult groups exhibit interesting ecological vari-

ability that would make reliable identification of species valuable. *Pinus*, *Quercus*, Rosaceae, Poaceae, and *Picea* are examples of genera or families that fall into this category. In a few cases, numerical and statistical approaches have been used with notable success. Hansen and Cushing (1973), for example, gained useful paleoecological insights by applying numerical methods to differentiate pollen grains of five species of *Pinus* from southwestern North America.

Another interesting challenge involves *Picea* in northeastern North America. Given the difficulty of separating the three *Picea* species—*Picea glauca*, *P. mariana*, and *P. rubens* (white, black, and red spruce)—in the pollen record, little is known about their specific histories in eastern North America following deglaciation. Previous attempts to differentiate the pollen of these species were based on relatively limited reference collections and on collections from different parts of North America (Cain, 1948; Richard, 1970; Birks and Peglar, 1980). In this study, we present the results from analysis of a large sample of modern pollen grains from the three northeastern *Picea* species. We examined 522 pollen grains from 38 different collections and made seven quantitative and six qualitative measurements on each grain. All collections were from Maine and Maritime Canada, the only region where the three species co-occur today. Previous studies demonstrated clearly that a single character would not suffice for identification of the *Picea* species. Accordingly, Birks and Peglar (1980), Hansen and Engstrom (1985), and Brubaker, Graumlich, and Anderson (1987) used discriminant function analysis for separation of *P. glauca* and *P. mariana*. Although this method is widely available in modern statistical packages, and although the species assignments can be made with an individual probability estimate, the procedure lacks the transparency of the traditional taxonomic key. In this study, we applied a classification and regression-tree (CART) analysis to the data

¹ Manuscript received 6 December 2001; revision accepted 16 April 2002.

The authors thank the following for their kind and constructive help with this project: Ed Cushing (Minnesota), Hartmund Kunze (JD Irving Ltd. New Brunswick) and Howard Frame (Nova Scotia Dept. of Natural Resources) for providing modern pollen grains of *Picea*; Isabelle Larocque (Sweden/Switzerland) for translation of some French text; to Ann Dieffenbacher-Krall (Maine) for considerable assistance in the laboratory and for helpful discussions; PM Ekö (Sweden) for his exceptional knowledge in pollen-related computer programming and statistical analysis; and Molly Schaufli and Shawn Fraver (Maine) for helpful discussions. The study was supported in part by grants to M. Lindbladh from The Swedish Foundation for International Cooperation in Research and Higher Education (STINT) and from the Fulbright Commission; by a grant to A. White, G. L. Jacobson Jr., and W. Krohn from The Nature Conservancy (U.S.); and by the University of Maine Institute for Quaternary and Climate Studies.

⁴ Author for reprint requests, current address: Southern Swedish Forest Research Centre, Swedish University of Agricultural Science (SLU), P.O. Box 49, S-230 53 Alnarp, Sweden (e-mail: matts.Lindbladh@ess.slu.se).

TABLE 1. The mean size, standard deviation, and 5% confidence interval for the quantitative measurements of the modern pollen grains of three *Picea* species. All measurements are in micrometers.

Variable	<i>P. glauca</i> (N = 173)			<i>P. mariana</i> (N = 176)			<i>P. rubens</i> (N = 173)		
	Mean	SD	CI	Mean	SD	CI	Mean	SD	CI
Total grain size (X_1)	98.87	8.01	1.19	84.60	9.91	1.46	89.69	8.03	1.20
Corpus breadth (X_2)	70.90	6.68	1.00	61.49	7.43	1.10	60.46	7.82	1.17
Corpus height (X_3)	54.97	7.30	1.09	46.61	6.54	0.97	46.55	5.56	0.83
Saccus height (X_4)	33.22	4.96	0.74	28.35	4.53	0.67	29.17	4.25	0.63
Saccus width at base (X_5)	48.77	6.53	0.97	39.86	5.56	0.82	40.17	5.63	0.84
Cap thickness (X_7)	2.45	0.61	0.09	2.53	0.54	0.08	2.95	0.66	0.10
Saccus attachment (X_8)	18.64	3.58	0.53	17.53	2.95	0.44	16.14	3.81	0.57

in order to distinguish pollen grains of all three *Picea* species. We wished to determine whether a classification method that combines the transparency of the traditional key with an objective and optimized numerical analysis of the data could be used to differentiate the reference grains.

METHODS

Pollen—We used pollen available from reference collections housed at the University of Maine Orono, Maine, USA, and the University of Minnesota, Minnesota, USA. In addition, we received pollen from two separate tree nurseries: Sussex Tree Nursery, New Brunswick, and Department of Natural Resources, Nova Scotia, Canada. All collections were from different individual trees from northeastern North America, where ranges of the three species overlap.

The pollen grains were washed in potassium hydroxide. They were treated with glacial acetic acid and then acetolysed with acetic anhydride and sulfuric acid for 2 min at 90°C. Silicon oil was added after the samples were dehydrated with tert-butyl alcohol. We avoided glycerin, as studies have shown that pollen may swell and to some extent change in morphological character when stored in this medium (Andersen, 1960). The samples were not stained. Prepared slides were sealed with paraffin. We added small amounts of sand grains (63–74 μm) to the slides in order to avoid flattened pollen grains because of cover-slip pressure (Cushing, 1961; Hansen and Engstrom, 1985). This was done in all but eight collections equally distributed among the three species. The mean grain size for the non-sand slides was slightly, but not significantly, bigger in *P. glauca* (Student's *t* test, $P = 0.33$) but was nonsignificantly smaller in *P. marina* ($P = 0.97$) and *P. rubens* ($P = 0.64$).

We examined a total 522 grains from 38 different collections. At least 20 grains were counted from each collection. Only grains that were unbroken,

symmetrical, and fully expanded in equatorial view were used. The same person using a Leitz light microscope with 400 \times magnification and an ocularmeter that was calibrated to a stage micrometer did all the counting and measuring.

All except one of the seven quantitative variables measured by Birks and Peglar (1980) were measured in this study (Table 1, Fig. 1). We also tested all the qualitative variables suggested by other authors (Richard, 1970; Birks and Peglar, 1980; Hansen and Engstrom, 1985) as possibly useful for distinguishing *Picea* species (Table 2). Following a preliminary analysis, we excluded three of these as unimportant, either because there were no significant differences among the species or because the character was regarded as unclear or difficult to judge. The three excluded were: "thinning exine at the end of capus," "ratio of the saccus/corpus area," and "saccus shape."

Each qualitative variable were given a score from 1 to 5, with 1 corresponding to a positive judgment of the variable in question, 3 an intermediate, and 5 a negative judgment. For example, a grain with a very high "degree of verrucation of the exine in the sinus area of the corpus" was given a score of 1.

Of the 13 remaining variables tested in this study, only four quantitative and two qualitative characters were ultimately determined to be useful in distinguishing the species. The four quantitative characters used were "total grain size" (X_1), "corpus breadth" (X_2), "cap thickness" (X_7), and "position of attachment of saccus to corpus" (X_8) (hereafter termed "saccus attachment") (Table 1). The two qualitative characters were the "degree of undulating margin of the corpus cap" (Y_2) (hereafter: "cap undulation") and the "degree of verrucation of the exine in the sinus area of the corpus" (Y_6) (hereafter "exine verrucation") (Table 2).

Measurements of these quantitative characters are straightforward. Of the qualitative characters used, "cap undulation" is relatively easy to detect (Fig. 2). The "exine verrucation" is a somewhat more subtle character, but consult Fig. 2, the drawing in Birks and Peglar (1980), or photographs in Richard (1970). (Those wishing to apply the method presented here will find that examination of reference grains helps enormously in recognizing the within- and between-species variability of the two qualitative characters in question.)

Here we briefly mention one of the inherent assumptions in the procedure of applying the results to fossil *Picea* grains; for a more thorough discussion of this matter consult Birks and Peglar (1980). The assumption is that the fossil grains originate only from the tree species of *P. glauca*, *P. mariana*, or *P. rubens*. We consider this to be a reasonable assumption for eastern North America because no other *Picea* species occurs today east of the Rocky Mountains. To the best of our knowledge, no macrofossils from Holocene sediments contradict this assumption. We cannot exclude the possibility of occasional long-distance transport of single grains from other species, including fossil pollen from the extinct species *Picea critchfieldii*, which once lived just north of the Gulf of Mexico (Jackson and Weng, 1999).

Classification and regression-tree (CART) analysis—The CART models recursively partition data sets on the basis of a set of independent variables. Here such a classification analysis has been applied to the morphometrics of the three *Picea* species (Fig. 3). The CART procedure first tested each morphometric variable to find the best combination of variable and split threshold that separated the entire sample into two groups that were internally as ho-

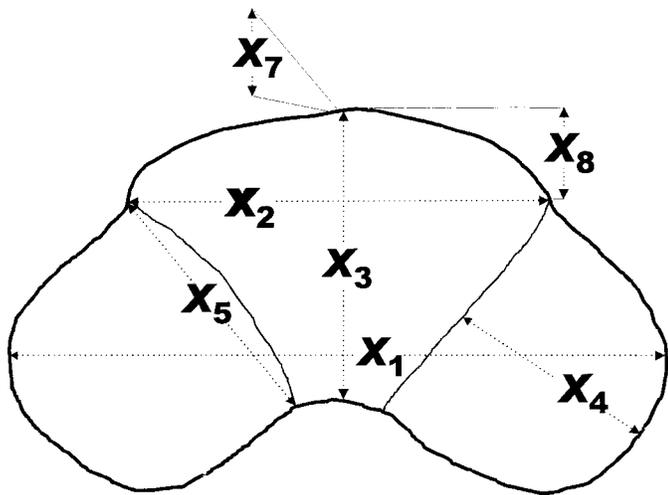


Fig. 1. The seven quantitative variables measured in this study. The variable designation (X_1 , etc.) are the same as in Birks and Peglar (1980).

TABLE 2. Frequency distribution by species for each qualitative variable of three *Picea* species ($N = 173$ for *P. glauca* and *P. rubens*, $N = 176$ for *P. mariana*).

Variable	Species	Score				
		1	2	3	4	5
Distal saccus attachment to the corpus (Y_1)	<i>P. glauca</i>	11	50	42	46	24
	<i>P. mariana</i>	16	68	61	29	2
	<i>P. rubens</i>	12	46	43	64	8
An undulating margin of the body cap (Cap undulation) (Y_2)	<i>P. glauca</i>	6	44	39	49	35
	<i>P. mariana</i>	3	27	42	76	28
	<i>P. rubens</i>	35	79	21	31	7
Large lumina of the internal reticulate structure of the saccus (Y_3)	<i>P. glauca</i>	8	52	24	69	20
	<i>P. mariana</i>	0	17	15	72	72
	<i>P. rubens</i>	1	24	13	79	56
Irregular lumina of the internal reticulate structure of the saccus (Y_4)	<i>P. glauca</i>	8	55	33	58	19
	<i>P. mariana</i>	0	23	28	55	70
	<i>P. rubens</i>	1	28	21	68	55
Constriction at the saccus attachment to the corpus (Y_5)	<i>P. glauca</i>	13	97	32	23	8
	<i>P. mariana</i>	17	119	24	15	1
	<i>P. rubens</i>	31	102	24	15	1
Verrucation of the exine in the sinus area of the body (Exine verrucation) (Y_6)	<i>P. glauca</i>	11	83	26	41	12
	<i>P. mariana</i>	53	85	24	11	3
	<i>P. rubens</i>	73	80	14	6	0

mogenous as possible with respect to species composition. In this case (Fig. 3), “total grain size” was the best such variable, with a size threshold of 88.5 μm . Relatively few grains of or below this size were *P. glauca* and most of those above this size were *P. glauca*. Each of these two subsets was then partitioned in turn, producing a substantial separation of *P. rubens* from *P. mariana* on the basis of the score for “cap undulation” with *P. rubens* typically scoring 1 or 2 and *P. mariana* scoring 3, 4, or 5. In the subset of larger grains, on the other hand, “corpus breadth” was the best discriminator variable, with the broader grains typically being *P. glauca*. Note how different criteria can apply within the two sibling subsets, allowing the types of contingent rule typical of taxonomic keys to emerge in a way not possible with multiple regression or discriminant function analysis. (There are other ways of achieving this with statistical methods, but they are either not straightforward or they require prior knowledge not needed with the CART procedure.) The process is repeated recursively through descendent nodes so that a decision tree is grown until certain stopping rules are encountered. At this point the resulting tree was pruned back to an optimally fitting version determined by cross-validation. The result is a decision tree that specifies a hierarchically organized suite of correlates of species identity.

Within each end node, all sample members present share common attributes, i.e., those that satisfy all of the decision criteria from root node to that end point. The same species can be classified successfully along more than one path from root to end node, i.e., by virtue of different chains of decisions. Note that the probabilistic nature of the classification achieved can lead to some splitting criteria distinguishing not between species but between probability of being a species: in Fig. 3, for example, the two rightmost end nodes are split on the basis of “Exine verrucation,” yet both nodes classify the grains falling into them as *P. glauca*. Where they differ is that grains reaching the rightmost node have a 96% probability of being *P. glauca* while those with lower “Exine verrucation” scores have only a 55% probability of being *P. glauca*. Finally, the order of entry of variables into CART models is normally unimportant, because it is ultimately the combination of attributes that uniquely characterizes a species (as in standard taxonomic keys).

The classification accuracy of the classification tree constructed in this way can be assessed in the usual way, as the proportion of the sample of known identity that are correctly classified by the classification tree. However, this proportion is likely to be inflated because the tree is constructed (via the exhaustive search procedure) so as to maximize the classification accuracy. This bias is controlled via the cross-validation pruning, but this involves a trade-off coefficient that weights classification accuracy against tree complexity (Breiman et al., 1984; Clark and Pregibon, 1992). Hence, a permissive choice of this coefficient allows optimistic results, akin to the use of alpha

levels of 0.10 or 0.15 in admitting terms to a step-up, multiple linear regression. An independent assessment of the efficacy of the pruned tree is therefore best obtained by making predictions from the model against a reserved sample of data (test set) not used in the derivation of the initial CART model (in this study we used a training set of 419 grains and a test set of 103 grains).

RESULTS

Table 1 summarizes the distribution of the quantitative variables. *Picea glauca* was markedly larger in all metrics except “cap thickness,” while *P. mariana* and *P. rubens* were relatively similar and *P. mariana* had smaller “total grain size” and “cap thickness” and larger “saccus attachment” than *P. rubens*. Nevertheless, despite these significant differences in the quantitative characters the considerable overlaps preclude differentiation of the three taxa on the basis of any one character alone. The mean sizes for *P. glauca* and *P. mariana* for all variables were considerably larger in this study compared to the study of Birks and Peglar (1980). For example, “total grain size” for *P. glauca* was 98.9 μm in this study compared to 89.7 μm in the Birks and Peglar study; for *P. mariana* the measurements were 84.6 and 74.7 μm , respectively. These differences are so large they cannot be explained by the difference in sample size (i.e., number of collections) alone and probably reflect differences in collection locality: all collections in the Birks and Peglar study were from the Midwest, while in this study all were from eastern North America. The collections of *P. rubens* in both studies were from eastern North America, and the means of each variable were similar in the two studies. (The mean for “total grain size” reported in Birks and Peglar [1980] is actually slightly larger [91.2 μm compared to 89.7 μm]).

Only two qualitative variables exhibited consistently large differences among the species (Table 2). Pollen from *Picea rubens* had more often a “cap undulation” than pollen from *P. glauca* and *P. mariana*. On average 66% of the *P. rubens* pollen had this feature, but significant variation existed among collections (see Appendix, <http://ajbsupp.botany.org/v89>). The second variable with some differences was the “exine verrucation.” It was more commonly a character on *P. rubens* and

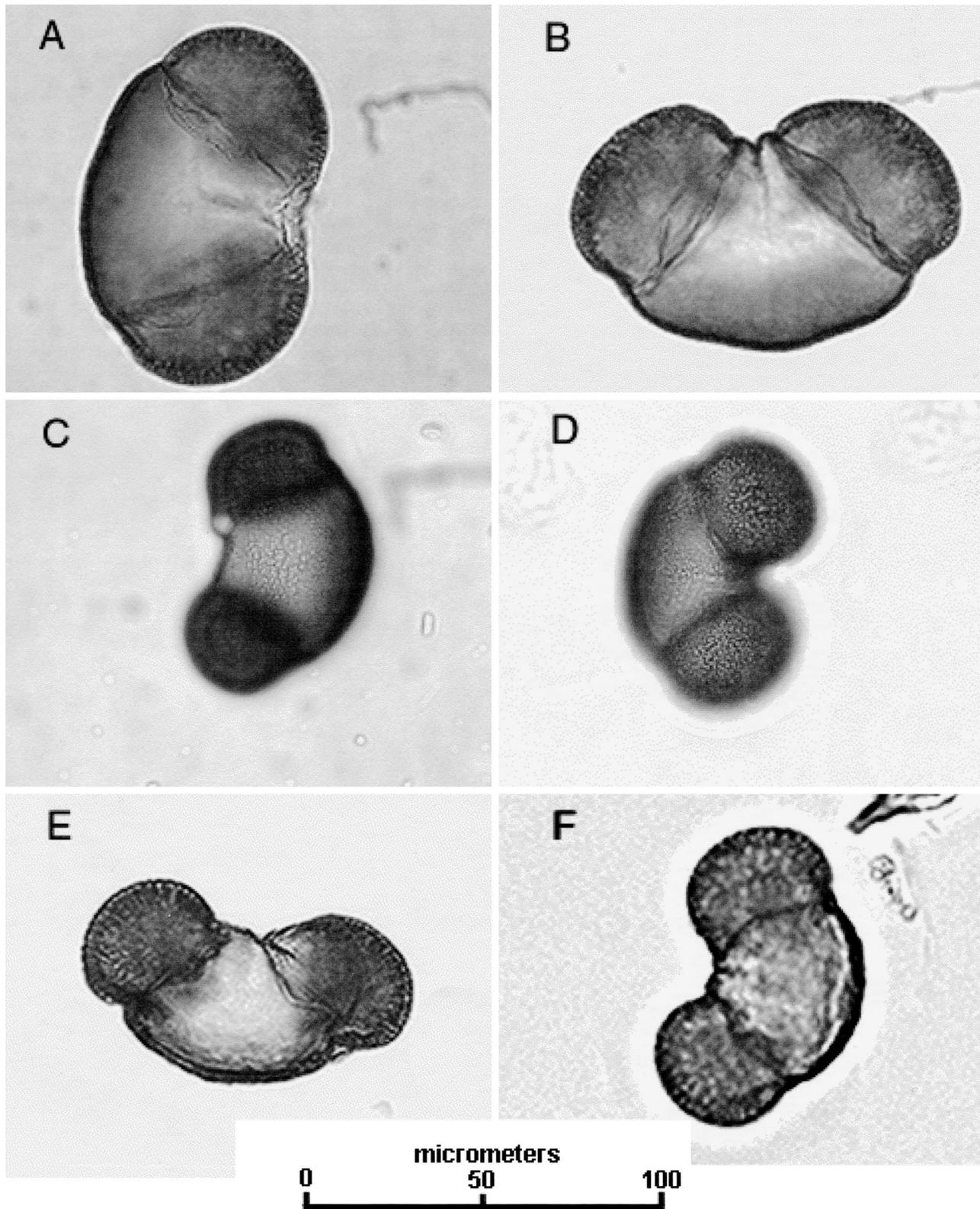


Fig. 2. Characteristic modern pollen grains of the three *Picea* species. (A–B) *Picea glauca*, (C–D) *Picea mariana*, and (E–F) *Picea rubens*.

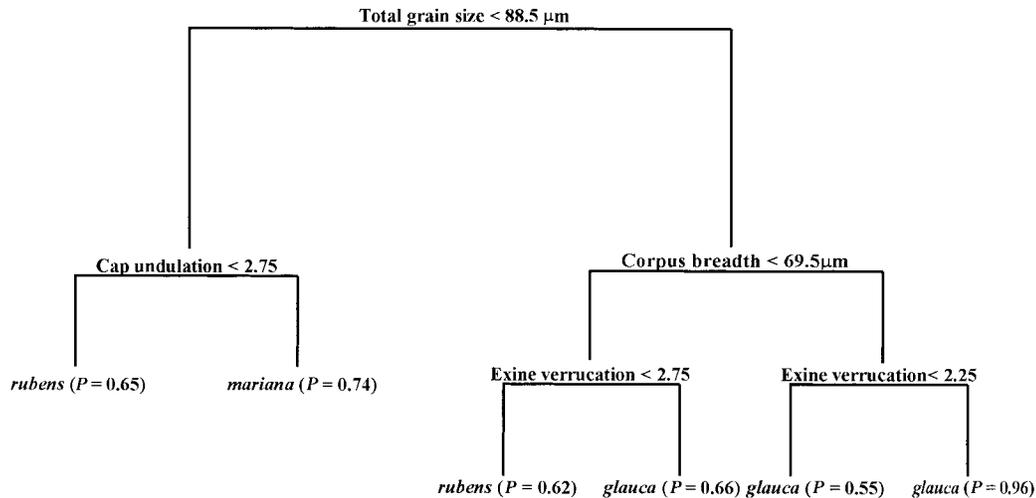


Fig. 3. Structure of classification tree for the simultaneous classification of all three *Picea* species. At each interior node the splitting variable and threshold value are shown and at each end node the commonest species and the probability of being that species with the combination of splitting criteria from end node to root are shown.

P. mariana grains (88 and 78%) than grains of *P. glauca* (54%). Interestingly, these two characters were also identified as critical discriminators in other studies, “cap undulation” by Birks and Peglar (1980) and “exine verrucation” by Richard (1970). However, none of the qualitative variables could be used alone to separate the three spruce species without an unacceptable rate of misclassification. See Appendix at <http://ajbsupp.botany.org/v89> for the entire data from all collections.

Traditional classification trees—Standard CART analysis across the three species reveals that *Picea mariana* grains could be segregated from the others by a combination of a small “total grain size” (<88.5 μm) and a large “cap undulation” score (>2.75) (Fig. 3). *Picea rubens* grains were identifiable either by the combination of a small “total grain size” (<88.5 μm) and a low “cap undulation” score (<2.75) or by a combination of short “corpus breadth” (<69.5 μm) and low “exine verrucation” score (<2.75) if “total grain size” was >88.5 μm. Longer grains were otherwise classified as *P. glauca*.

The probability of grains in a given end node in Fig. 3 being of the species shown for that end node varied across nodes. From left to right across the end nodes, these probabilities were 0.65 (*P. rubens*), 0.74 (*P. mariana*), 0.62 (*P. rubens*), and 0.66, 0.55, and 0.96 (all *P. glauca*). Note that the two rightmost end nodes are both *P. glauca*, so what knowledge the “exine verrucation” score provides is not discrimination

between two species but rather greater certainty of the identification as *P. glauca* when this score is above 2.25. Thus, although the overall misclassification rate of the tree is 30.6%, the rate of misclassification of grains varies across nodes.

Table 3 shows the classification accuracy of the tree for each species in the training set. Species-specific classification accuracies were 77.1, 59.7, and 71.3% for *P. glauca*, *P. mariana*, and *P. rubens*, respectively. However, these figures reflect a classification tree structured to optimize its fate to the training data and therefore are likely to be overoptimistic. Table 3, therefore, also shows the results of applying the classification tree of Fig. 3 to the reserved data set of 103 cases (i.e., grains not used in the training/calibration analysis), yielding figures of 86.2, 59.5, and 73.0%, respectively. Thus, those of the test-set analysis matched closely the accuracy estimates from the training set.

Binary trees—Traditional discriminant function analysis attempts to create functions that typically describe multivariate axes and a multidimensional space within which the different groups to be classified are mutually segregated. In some cases, however, the variables measured and their distributions may be such that individual groups could be separated from the others by some simple criterion, e.g., group A members may share high values of variable X_1 , while B, C, and D all share low X_1 values, and group B members may in turn segregate from the pool of A, C, and D members in having a particularly low value of variable X_2 . This opens the classification strategy to becoming one of multiple binary classifications (e.g., as A or not A) that can be combined via Boolean logic to describe various outcomes (R. O’Connor and M. Lindbladh, unpublished data). Recursive segregation of successive groups from the residual pool of cases is, in principle, possible as a default outcome of CART analysis, but this would have a different probability of outcome than a strict combination of all binary discriminants. We therefore evaluated the performance of combinations of binary classification trees, each discriminating one species from the pooled data for the other two, and each operating on the complete sample.

Figure 4a–c presents the binary trees for each species. *Picea*

TABLE 3. Classification accuracy by species using a classification tree across all three *Picea* species.

Identified as	Data	Tree species					
		<i>P. glauca</i>		<i>P. mariana</i>		<i>P. rubens</i>	
		N	%	N	%	N	%
<i>P. glauca</i>	Training	111	77.1	21	15.1	22	16.2
	Test	25	86.2	5	13.5	4	10.8
<i>P. mariana</i>	Training	12	8.3	83	59.7	17	12.5
	Test	0	0.0	22	59.5	6	16.2
<i>P. rubens</i>	Training	21	14.6	35	25.2	97	71.3
	Test	4	13.8	10	27.0	27	73.0

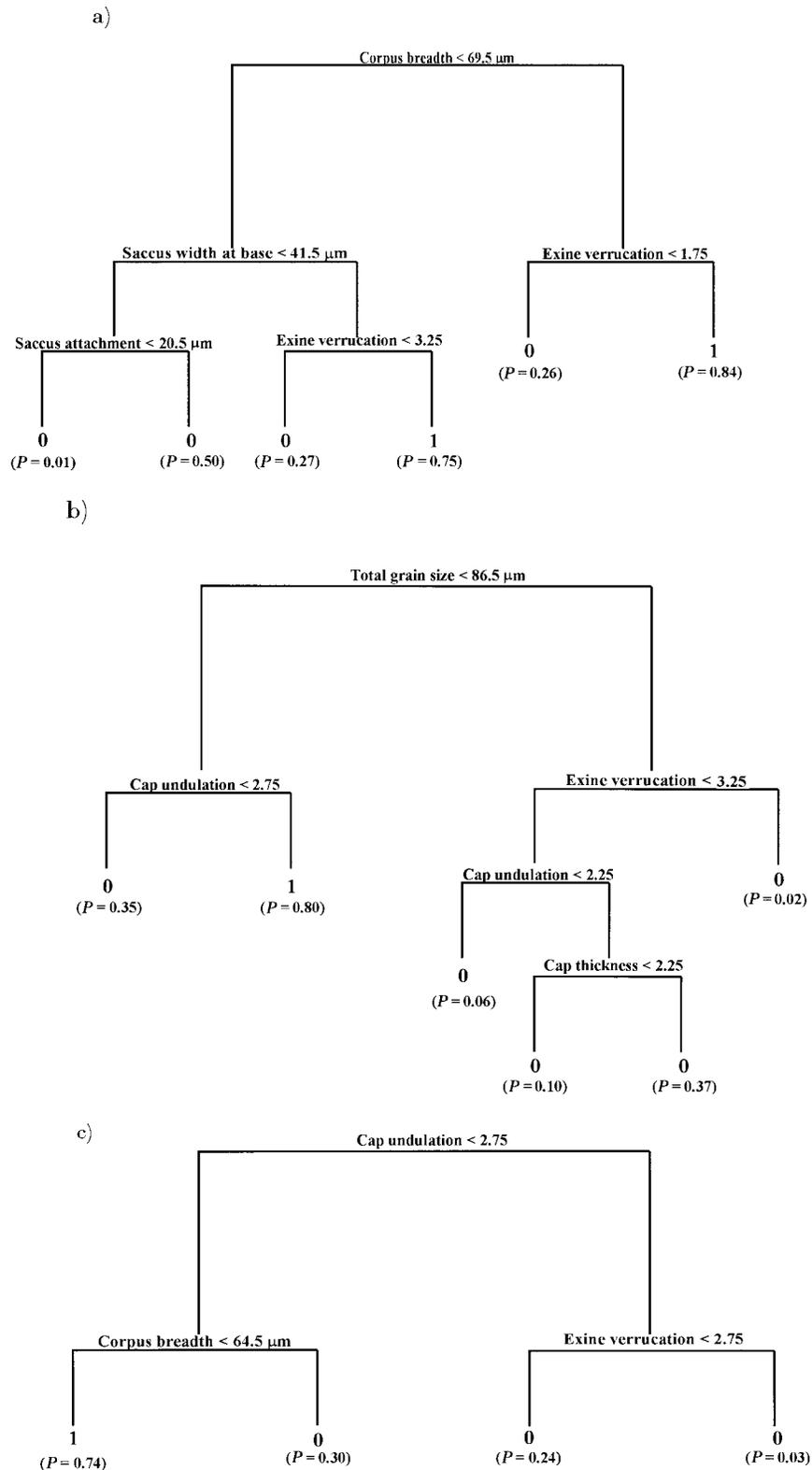


Fig. 4. Structure of binary classification trees for (a) *Picea glauca*, (b) *Picea mariana*, and (c) *Picea rubens*. For each end node a "1" indicates that the focal species was the commonest in that node and a "0" indicates that individuals of the two nonfocal species for that tree were commonest there. The *P* value gives the probability in a given end node for being the focal species. For example, in Fig. (a): 84% of the grains in the training set that had the combination "Corpus breadth > 69.5 μm " and "Exine verrucation > 1.75" were grains of *Picea glauca*.

TABLE 4. All possible outcomes of eight ordered triplets. The order is *Picea glauca*, *P. mariana*, and *P. rubens*.

Code	Meaning
000	Not classified to any species: unidentifiable
001	Unequivocally classified as <i>P. rubens</i>
010	Unequivocally classified as <i>P. mariana</i>
011	Matches both <i>P. mariana</i> and <i>P. rubens</i> : equivocal
100	Unequivocally classified as <i>P. glauca</i>
101	Matches <i>P. glauca</i> and <i>P. rubens</i> : equivocal
110	Matches <i>P. glauca</i> and <i>P. mariana</i> : equivocal
111	Matches all three species: equivocal

glauca (Fig. 4a) was discriminated from the other species by the combination of narrow “corpus breadth,” large “saccus width at base” and a high score of “exine verrucation,” or by a broad “corpus breadth” and medium to high score of “exine verrucation” (>1.75). Classification accuracy for the training set was 84.0%. For *P. mariana* (Fig. 4b), classification as the species was based on a short “total grain size” (<86.5 μm) and a high score of “undulating exine” (>2.75). All other cases (i.e., those not classed as *P. mariana* (see previous sentence) were classed as non-*P. mariana*. Overall *P. mariana* classification accuracy was 79.7%. Finally, *P. rubens* (Fig. 4c) was discriminated from the other species by virtue of a low score of “undulating exine” (<2.75) and narrow “corpus breadth” (<64.5 μm), with overall accuracies of classification equal to 78.5%. These classification accuracies are again from the training set and lower estimates are assessed from the test set, namely 79.3, 56.8, and 59.5% for *P. glauca*, *P. mariana*, and *P. rubens*, respectively.

With three binary classifications (one for each species) conducted independently here, maximum agreement between the three arises when a given grain is successfully classified to only one species and is simultaneously classified as being neither of the other two species. Conversely, total failure to identify a grain would arise if no possible positive classification emerges for that individual. Similarly, maximum confusion about the identity of the grain would arise were it simultaneously classified positively by all three species-specific decision trees. Species confusion would also arise when there is a positive classification for the wrong species tree and a negative for the other two. We can list all possible outcomes in a list of eight ordered triplets or Hamming (1950) codes (Table 4). Three of the eight codes (001, 010, and 100) thus designate unambiguous classification; (000) designates a complete lack of success in classification; and the remaining four denote some (011, 101, and 110) or complete (111) equivocation.

Table 5 summarizes the Hamming codes results obtained on applying the trees of Fig. 4a–c to each grain, both for the training and the test set of data. Inevitably, the raw classification accuracies are lower than for the binary trees alone (as above), at 67.4, 47.5, and 50.7% for *P. glauca*, *P. mariana*, and *P. rubens*, respectively, in the training set and 79.3, 54.0, and 59.5% in the test data set. However, the Hamming codes identify both those cases where the grain was unidentifiable and those cases where the classification was equivocal, so the analyst knows which cases are unequivocally identified. When only these cases are considered, classification rates rise to 94.2 and 100.0% (training and test data, respectively) for *P. glauca*, 72.5 and 76.9% for *P. mariana*, and 77.5 and 75.9% for *P. rubens*. Thus, the Hamming codes yield substantially higher classification certainty than did the binary trees alone. Much of this increase in certainty comes from the identification of particular cases as being unidentifiable, with 20–32% of the grains being in this category. Ambiguities in classification were relatively few (0.0–6.3%) (Table 5).

Table 6 presents a breakdown of grain classification failures by Hamming code. Misclassifications of *P. glauca* were relatively fewer than for *P. mariana* and *P. rubens*; *P. mariana* and *P. rubens* were as likely to be mistaken for each other as for *P. glauca*.

Given that about 20–35% of all cases were unclassifiable or equivocal, it was worth considering whether the failure to classify these grains was associated with particular morphometric features or whether it reflected a random distribution of measurement errors affecting the classification. Considering total grain lengths, for example, unclassifiable and equivocal grains were in general intermediate in morphology (Fig. 5). The unidentifiable grains that were truly *P. glauca* were slightly smaller than was typical for *P. glauca*; on the other hand, the unidentifiable *P. mariana* and *P. rubens* grains were slightly longer than was typical of the species. Consequently these grains were intermediate in size, fitting neither the smaller *P. rubens* and *P. mariana* stereotype nor the larger *P. glauca* stereotype. As a result, the classification algorithm failed to classify them. The same phenomenon was evident in the other morphological metrics considered.

DISCUSSION

When pollen in fossil samples are distinctive morphologically, their discrimination by direct taxonomic keys can be rapid and efficient. Where the difference between pollen in closely related taxa are less pronounced, however, taxonomic keys suffer in their dependence on expert judgment, in which

TABLE 5. Extent of accurate classification, equivocal classifications, and unclassifiable cases when using Hamming-coded (1950) binary classification trees.

	<i>P. glauca</i>		<i>P. mariana</i>		<i>P. rubens</i>	
	Training	Test	Training	Test	Training	Test
Grains of this species	144	29	139	37	136	37
Identification made of this species	97	23	66	20	69	22
Unequivocal identification	103	23	91	26	89	29
Equivocal classifications	9	0	7	1	3	0
Unidentifiable	32	6	41	10	44	8
Percentage correctly identified	67.4	79.3	47.5	54.0	50.7	59.5
Percentage unequivocal identification correct	94.2	100.0	72.5	76.9	77.5	75.9
Percentage equivocal classifications	6.3	0.0	5.0	2.7	2.2	0.0
Percentage unidentifiable	22.2	20.7	29.5	27.0	32.4	21.6

TABLE 6. Distribution of species-specific codes for grains not correctly classified to species by the combined binary classification trees.

Species	Code	Training set		Test set		Interpretation
		Frequency	Percentage	Frequency	Percentage	
<i>P. glauca</i>	000	32	22.2	6	20.7	Unidentifiable
	001	4	2.8	0	0.0	Misclassified as <i>P. rubens</i>
	010	2	1.4	0	0.0	Misclassified as <i>P. mariana</i>
	101	4	2.8	0	0.0	Equivocal
	110	5	3.5	0	0.0	Equivocal
<i>P. mariana</i>	000	41	29.5	10	27.3	Unidentifiable
	001	16	11.5	3	8.1	Misclassified as <i>P. rubens</i>
	100	9	6.5	3	8.1	Misclassified as <i>P. glauca</i>
	101	1	0.7	0	0.0	Equivocal
	110	6	4.3	1	2.7	Equivocal
<i>P. rubens</i>	000	44	32.4	8	21.6	Unidentifiable
	010	10	7.4	4	10.8	Misclassified as <i>P. mariana</i>
	100	10	7.4	3	8.1	Misclassified as <i>P. glauca</i>
	101	2	1.5	0	0.0	Equivocal
	110	1	0.7	0	0.0	Equivocal

not all experts may concur, and, more importantly, in their risk of suboptimal performance. A less subjective analysis might well reveal fewer or more effective attributes than those in the expert's key, allowing discrimination among species with a lower investment in morphological measurement. Moreover, even with expert knowledge and a reliable key several taxa cannot be discriminated because of overlapping attributes or combinations of characters that are difficult to judge. These considerations have led a number of workers to turn to discriminant function analysis as a complement to taxonomic keys (Birks and Peglar, 1980; Hansen and Engstrom, 1985; Brubaker, Graumlich, and Anderson, 1987). Discriminant function analysis constructs linear functions of the measured variables, determining weights for each variable, such that the different species receive, as far as possible, different values of the function. Two or more functions may be needed to separate species. This approach is often highly effective, the classic example being Fisher's (1936) separation of three *Iris* species on the basis of petal and sepal morphology. Although discrim-

inant function analysis is widely available in modern statistical packages, and although the species assignments can be made with an individual probability estimate, the procedure lacks the transparency of the traditional taxonomic key. In this respect, the application of CART (Breiman et al., 1984) is advantageous, combining transparency with objective and optimized analysis of morphometric data. It is, of course, important that the classification tree method be at least as powerful for classification accuracy as discriminant function analysis. This appears to be the case: when we subjected our training set to discriminant function analyses we obtained classification rates of 76% for *P. glauca*, 63% for *P. mariana*, and 70% for *P. rubens* (R. O'Connor and M. Lindbladh, unpublished data), rates comparable to those derived from the standard classification tree model here (Table 3).

The notion of using Hamming codes to combine the results of binary classification trees appears to be an innovation in classification work. The resulting raw classification rates were necessarily lower than those for the binary trees (Table 5), but reliance on the binary trees alone may be risky: the training set binary classification rates were markedly higher than those from the standard multispecies classification tree—84.0 vs. 77.1% for *P. glauca*, 79.7 vs. 59.7% for *P. mariana*, and 78.5 vs. 71.3% for *P. rubens*—but the values obtained from the test set with the binary trees were lower (79.3, 56.8, and 59.5%, respectively). It is likely, therefore, that the binary results for individual species involve significant overfitting to the training data. When combined as Hamming codes, however, the species-specific binary trees have the potential to reinforce each other in clear-cut cases and to reveal conflicting classifications with intermediate grains. Thus, a major advantage of the Hamming codes is to reveal those cases for which classification is uncertain. That it is the difficulty of classification rather than a deficiency in the classifying algorithm is indicated by the finding here that the problematic cases in our training set were of intermediate morphology, often too large in a given dimension to be *P. mariana* or *P. rubens* but at the same time smaller than typical *P. glauca* grains (Fig. 5). In operational terms, this ability of the Hamming codes to identify problematic grains that would otherwise reduce the overall classification accuracy is of special value in palynological studies. (It is, of course, possible in a discriminant function analysis to plot the location of individual grains within the *n*-

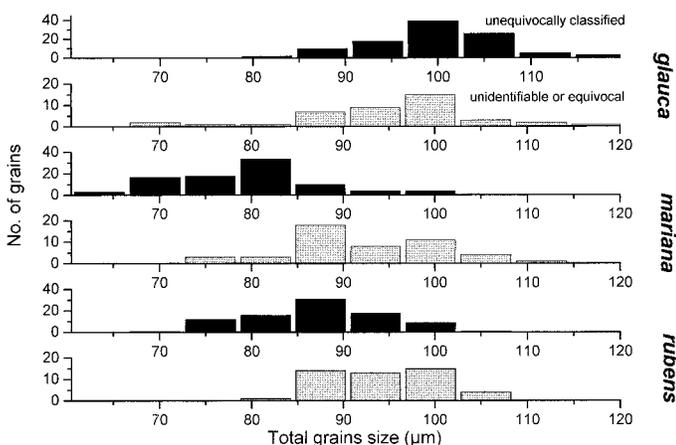


Fig. 5. Distributions of "total grain size" for grains of each of the three *Picea* species according to whether the binary triplet procedure classified the grains correctly to species or whether the grain proved problematic in all cases for the training data. From top to bottom the distributions are: *Picea glauca* (unequivocally classified [black]; unidentifiable or equivocal [shaded]), *Picea mariana* (unequivocally classified [black]; unidentifiable or equivocal [shaded]), and *Picea rubens* (unequivocally classified [black]; unidentifiable or equivocal [shaded]).

dimensional space of the n discriminant functions and to determine which grains are peripheral to the species-specific clusters identified. This approach is less direct, and more complex, than the generation of an unequivocal Hamming code). Intuitively, one would expect that conclusions based on relatively reliable estimates of species composition from a subset of available grains are scientifically superior to less reliable estimates made from the complete sample. Furthermore, any spatial/temporal variation in the proportion of unclassifiable grains in the samples for different locations or times may itself be informative. It is worth noting that the Hamming procedure may be of relatively greater value as the number of species involved in the study increases. In principle, Hamming codes can be created for samples with four, five, six, etc. species, yielding quadruplet, quintuplet, and sextuplets instead of triplets (with correspondingly more values, e.g., $2^6 = 64$ values for a six-species mix). One would intuitively expect discriminant function analysis to be relatively more difficult as the number of species in the mixture increases, so that Hamming codes identifying problematic cases would be increasingly valuable. On the other hand, the possibilities for equivocal codes increases in a binomial distribution, and one could envisage a relatively greater frequency of equivocal rather than unclassified grains as species richness increases. However, this is not a problem for analysis of grains in less diverse samples.

Conclusions—The large number of pollen (522) from many collections (38) used in this study was the prerequisite for the separation of the three closely related *Picea* species. The large between- and within-species similarities discovered in the study emphasize the need for a large collection of reference grains. Furthermore, the apparently large differences in the quantitative characters among the *Picea* pollen from different regions (i.e., eastern North America and Midwest) suggest that applying the comparative data from one region to another might not work. It may be useful for the analysis to be repeated for each region of interest. Classification tree analysis appears to provide a robust and transparent procedure, and it is at least as powerful as traditional discriminant function analysis, while retaining the conceptual simplicity of traditional taxonomic keys. Use of Hamming codes to link multiple binary trees allows identification and exclusion of problematic specimens and correspondingly greater classification certainty among the remaining grains. This study shows that there is a large proportion of problematic specimens in a modern reference collection of *Picea* pollen, and thus it is likely that a number of fossil grains will also fall into this category. Hence, the new method allows more confidence than is possible with

previous methods about the identity of unequivocally classified grains.

The method has successfully been applied to fossil *Picea* grains from eastern North America (Lindbladh, Jacobson, and Schaffler, in press). It is not restricted in applicability to *Picea* but can readily be applied to any other taxon for which a large reference collection is available. Extensive experience with CART indicates that an initial sample of several hundred cases is needed before adequate discrimination is achieved. With smaller samples, the algorithm fails gracefully rather than proceeding to misidentify cases.

LITERATURE CITED

- ANDERSEN, T. 1960. Silicone oil as a mounting medium for pollen grains. *Danmarks Geologisk Undersøgelse* 4, 4(1): 1–24.
- BIRKS, H. J. B., AND S. M. PEGLAR. 1980. Identification of *Picea* pollen of Late Quaternary age in eastern North America: a numerical approach. *Canadian Journal of Botany* 58: 2042–2058.
- BREIMAN, L., J. H. FRIEDMAN, R. A. OLSHEN, AND C. J. STONE. 1984. Classification and regression trees. Wadsworth, Belmont, California, USA.
- BRUBAKER, L. B., L. J. GRAUMLICH, AND P. M. ANDERSON. 1987. An evaluation of statistical techniques for discriminating *Picea glauca* from *Picea mariana* pollen in Northern Alaska, USA. *Canadian Journal of Botany* 65: 899–906.
- CAIN, S. A. 1948. Palynological studies at Sodon lake. I. Size frequency study of fossil spruce pollen. *Science* 108: 115–117.
- CLARK, L. A., AND D. PREGIBON. 1992. Tree-based models. In J. M. Chambers and T. J. Hastie [eds.], *Statistical models in S*, 377–419. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, California, USA.
- CUSHING, E. J. 1961. Size increase in pollen grains mounted in thin slides. *Pollen et Spores* 3: 265–274.
- FAEGRI, K., AND J. IVERSEN. 1964. *Textbook of pollen analysis*, 2nd ed. Blackwell Scientific, Oxford, UK.
- FISHER, R. G. 1936. The use of multiple measurements in taxonomic problems. *Annals of Eugenics* 7: 179–188.
- HAMMING, R. W. 1950. Error detecting and error correcting codes. *Bell System Technical Journal* 29: 147–160.
- HANSEN, B. S., AND E. J. CUSHING. 1973. Identification of pine pollen of Late Quaternary age from the Chuska Mountains, New Mexico. *Geological Society of America Bulletin* 84: 1181–1200.
- HANSEN, B. C. S., AND D. R. ENGSTROM. 1985. A comparison of numerical and qualitative methods of separating pollen of black and white spruce. *Canadian Journal of Botany* 63: 2159–2163.
- JACKSON, S. T., AND C. WENG. 1999. Late Quaternary extinction of a tree species in eastern North America. *Proceedings of the National Academy of Science* 96: 13 847–13 852.
- LINDBLADH, M., G. L. JACOBSON, JR., AND M. SCHAFFLER. In press. The postglacial history of three *Picea* species in New England, USA. *Quaternary Research*.
- MOORE, P. D., J. A. WEBB, AND M. E. COLLINS. 1991. *Pollen analysis*, 2nd ed. Blackwell Scientific, Oxford, UK.
- RICHARD, P. 1970. Atlas pollinique des arbres et de quelques arbustes indigènes du Québec. I Introduction generale. II Gymnospermes. *Naturalist Canadian* 97: 1–34.