Identification of seed origin of slash pine plantations

By A. E. Squillace, H. T. Schreuder and H. T. Bhattacharyya[1])

(Received March / June 1980)

Summary

Likelihood analysis of monoterpene composition in slash pine is shown to be a useful tool for ranking possible geographic origins of slash pine plantations. The procedure could also be used to check suspected errors in provenance study plots and in seed certification. Additional information such as morphological traits and historical records should also be used when available.

Key words: monoterpene composition, likelihood, *Pinus elliottii*

Zusammenfassung

Wahrscheinlichkeitsanalysen der Monoterpen-Zusammensetzung bei Pinus elliottii var. elliottii wurden als ein nützliches Mittel für die Erstellung einer Rangordnung möglicher geographischer Herkünfte von entsprechenden Pflanzungen festgestellt. Das Verfahren konnte ebenfalls benutzt werden, um vermutete Fehler in Herkunftsunterbrechungen und in der Samenzertifizierung zu prüfen. Zusätzliche Informationen, wie morphologische Merkmale und geschichtliche Protokolle sollten, falls vorhanden, ebenfalls zu Rate gezogen werden.

Introduction

Managers frequently wish to determine the geographic origin of seed used in establishing plantations. For example, phenotypically superior plantations are often used in seed procurement, to establish seed production areas, or to make individual tree selections. However, realizing the danger of moving seed out of the natural range of the species, they prefer to know the seed origin. In this paper, we outline a technique for identifying the geographic origin of slash pine

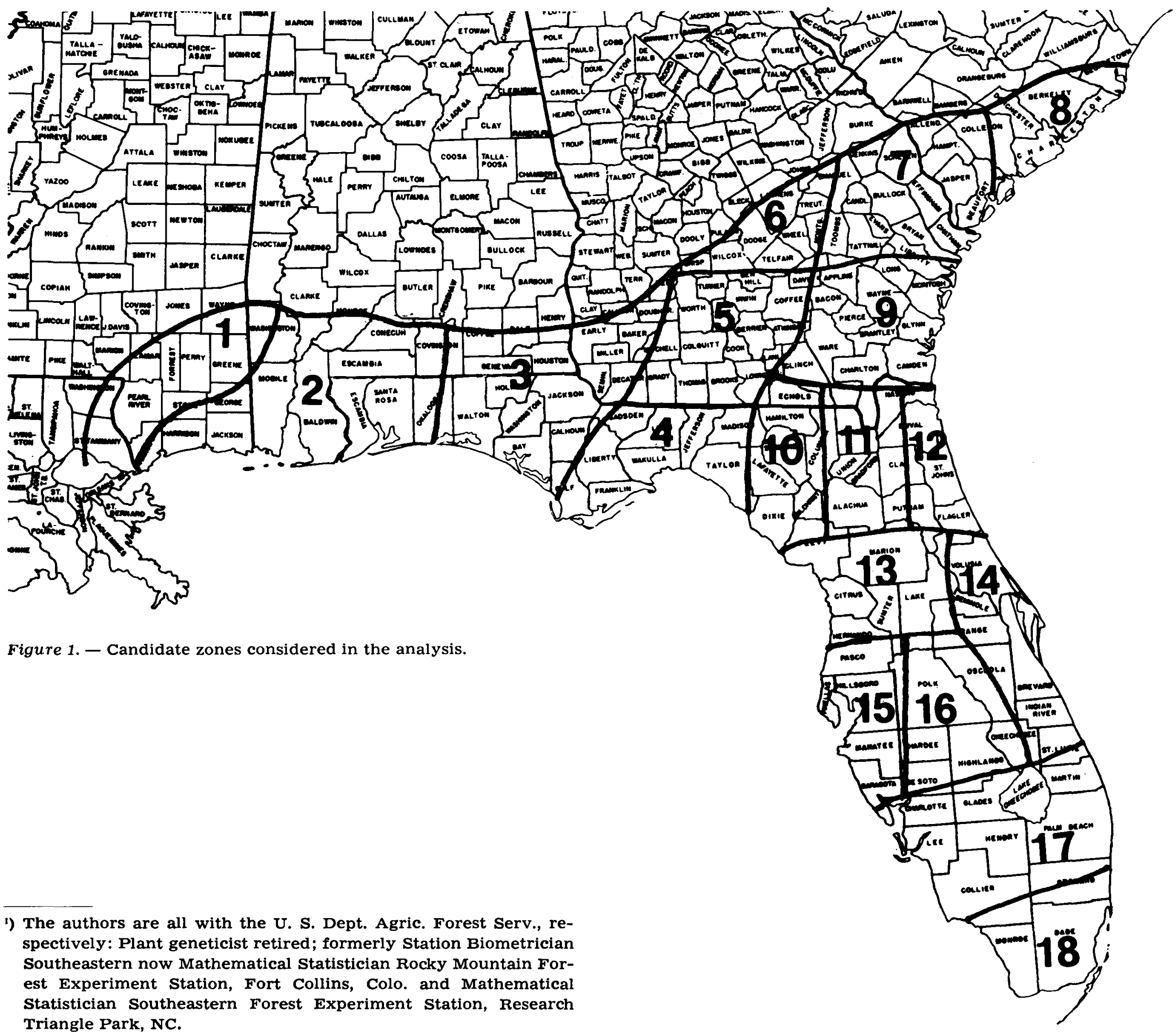

Figure 1. — Candidate zones considered in the analysis.

[1]) The authors are all with the U. S. Dept. Agric. Forest Serv., respectively: Plant geneticist retired; formerly Station Biometrician Southeastern now Mathematical Statistician Rocky Mountain Forest Experiment Station, Fort Collins, Colo. and Mathematical Statistician Southeastern Forest Experiment Station, Research Triangle Park, NC.

(*Pinus elliottii* ENGELM.) plantations after determining the monoterpene composition. Two examples of the application of the technique are given. Although not discussed further here, the techniques employed could also be useful in checking suspected errors in the origin of seed in provenance study plots and in determining origin of purchased seed. Besides monoterpenes, isoenzymes have also been demonstrated to be useful for identifying seed origin (BERGMANN, 1975 and 1977).

In order to judge the origin of trees from a single, unknown source, there must be variation over the species range and the pattern must be known. In slash pine, rather distinct patterns of variation in monoterpene composition were recently reported (GANSEL and SQUILLACE, 1976). These patterns, for example, show that the percentage of trees having high myrcene varies from as high as 100 % in extreme south Florida to less than 10 % at the northern limits of the species range in Georgia. Similarly, the other three simply-inherited monoterpenes β-pinene, limonene, and β-phellandrene also show distinctive patterns. These patterns permit judging the probable seed origin of a sample of trees. SQUILLACE (1977 a and b) recently made such a determination for a plantation in west Florida by comparing composition of the sample trees with samples covering the species range. SQUILLACE eliminated unlikely areas and concluded that the plantation was not of local origin and that the seed most likely came from north central Florida.

Monoterpene Composition as a Tool for Identification

Monoterpene composition of the cortical oleoresin in slash pine is well suited for identifying seed origin. Four of the major monoterpenes in the cortical oleoresin of slash pine are simply inherited (SQUILLACE, 1977 a and b). Trees can readily be classified as having relatively high or low amounts of each of these substances. The monoterpene composition is not greatly affected by environment (e.g., the monoterpene composition of ramets of the same clone planted in different areas remains relatively constant). The proportions of trees having high and low amounts of each monoterpene in a stand can readily be determined by analyzing small quantities of cortical oleoresin obtained from the branches of sample trees. Two men can collect oleoresin samples from between 50 to 100 trees (sufficient for determining seed origin of a stand of trees) within about a half-day. Modern gas chromatographs will automatically determine the percentage composition for 35 trees within 12 hours or less.

Identification of Origin

Likelihood Technique

A more refined, statistical method involves the computation of a likelihood for each candidate geographic area from which the seed may have originated. To use this method, the range of slash pine was first arbitrarily divided into 18 geographic areas, comprising adjacent counties (*Figure 1*). Then, the percentage of trees (p) having high or low amounts of each monoterpene in each zone was computed from the total sample of 1255 trees of known origin.

We assume a multinominal distribution (16 categories) for each candidate zone allowing for high (H) or low (L) content of each of the four monoterpenes. We then estimate the probabilities (p_i) of each of the 16 categories (HHHH, HHHL, . . . , LLLL) for each zone using the sample of trees known to be from that zone. For example, for zone 17 we had a sample of m = 34 (out of the 1255) known trees. The distribution of these trees and the corresponding estimated probabilities p_i for the 16 categories are given in columns 3 and 4 of *Table 1*. To classify the sample of N trees of unknown origin, we compute its likelihood L of being from each of the 18 zones and select that zone which maximizes the likelihood L. For each candidate zone,

$$L = (N!/[n_1!n_2!,..n_{16}!])\ p_1^{n_1} p_2^{n_2} \ldots p_{16}^{n_{16}} \quad (1)$$

where the n_i (i = 1, . . . , 16) are the number of trees in the unknown sample falling into category i. For the p_i (i = 1, . . . , 16) for each zone we used

$$p_i = (M_i + \frac{1}{16})/(M + 1) \quad (2)$$

instead of the standard estimator $p_i = M_i/M$. Here M and M_i represent the sample size and the number of trees falling in category i in the original sample of known trees. Our estimator is the multinominal analogue of the Bayesian estimator suggested by CHEW (1971) for the binomial case when a beta prior distribution is assumed. This avoids the possible problem of encountering zero probabilities caused by limited sample sizes of known trees.

We computed likelihoods for two plantations of known (Liberty County and Charlotte County, Florida) and two of unknown seed origin (St. Johns County and Alachua County, Florida) to check the effectiveness of the procedure. The plantations were sampled for other purposes but the data had not been used in estimating the multinomial distributions for the 18 zones. Results are given in *Table 2* for all zones. *Table 1* illustrates the procedures involved in computing, for example, the likelihood L corresponding to candidate zone 17 for an unknown sample (actually Charlotte County) of size N = 32. Since likelihoods are usually very small, we shall work with the logarithm of the likelihood (1n L). The results in *Table 2* show that the Liberty County sample (N = 90) most likely comes from zone 9 with zone 11 second best. The actual seed source is located in zone 11. The Charlotte County sample most likely comes from zone 18 which is in fact the actual location of the seed source .

Table 1. — An example illustrating the distribution of zone 17 and a procedure for computing the likelihood L for unknown sample (Charlotte County) corresponding to zone 17.

Phenotypic[1] class	(i)	Sample Charlotte Co. (N_i)	Candidate zone 17 (M_i)	(p_i)
HHHH	1	0	5	0.1446
HHHL	2	0	0	0.0020
HHLH	3	1	3	0.0875
HHLL	4	0	0	0.0020
HLHH	5	0	0	0.0020
HLHL	6	0	0	0.0020
HLLH	7	0	2	0.0589
HLLL	8	0	0	0.0020
LHHH	9	16	13	0.3732
LHHL	10	0	0	0.0020
LHLH	11	11	7	0.2018
LHLL	12	0	0	0.0020
LLHH	13	4	4	0.1161
LLHL	14	0	0	0.0020
LLLH	15	0	0	0.0020
LLLL	16	0	0	0.0020
Total		N = 32	M = 34	1.0021

$$L = \frac{32!}{1!\ 16!\ 11!\ 4!} (0.0875)^1(0.3732)^{16}(0.2018)^{11}(0.1161)^4 = 668 \times 10^{-9}$$

$\ln L = -14.22$

[1] Trees belonging to class HHLH, for example, are high in monoterpenes 1, 2, 4, and low in monoterpene 3.

Table 2. — Log likelihoods of four Florida plantations for 18 zones covering the range of slash pine in the Southeast.

Zone	Sample Size	Liberty Co. (N=90)	Charlotte Co. (N=32)	St. Johns Co. (N=50)	Alachua Co. (N=102)
1	56	-24.63	-182.70	-38.10	-21.96
2	98	-34.97	-200.80	-32.48	-24.64
3	81	-20.69	-194.40	-33.83	-12.58
4	52	-20.21	-180.60	-29.17	-11.44
5	79	-35.54	-193.80	-36.78	-25.56
6	73	-32.28	-192.40	-22.20	-30.46
7	96	-21.03	-199.40	-37.59	-13.27
8	29	-41.12	-164.50	-20.78	-39.62
9	48	-14.41	-178.50	-25.53	-11.07
10	70	-38.19	-189.30	-39.22	-36.90
11	162	-18.34	-215.50	-31.06	-10.53
12	50	-41.07	-179.40	-39.95	-54.45
13	109	-33.55	-152.80	-34.75	-43.41
14	76	-407.30	-31.41	-211.40	-468.70
15	29	-128.40	-27.13	-78.72	-145.20
16	47	-117.90	-35.09	-70.17	-144.70
17	34	-207.50	-14.22	-136.60	-246.10
18	20	-425.10	-13.52	-214.10	-492.10

________ denotes the largest likelihood value.

--------- denotes next largest likelihood value.

For the unknown seed sources, the St. Johns County (N = 50) sample most likely comes from zones 8 or 6, historical records suggesting zone 11 as the most likely source. This is in agreement with historical records.

These results show that the computed likelihoods can be a useful tool in ascertaining the origin of slash pine plantations. Clearly, the results would have been better if the original sample sizes for each zone had been larger.

Use of Other Information

The zones delineated within Florida *(Figure 1)* and south Georgia tend to differ from each other rather appreciably mainly because of the prevalence of north to south clinal patterns in monoterpene composition. However, within northern portions of the species range some zones are similar to each other even though geographically distant. In such cases, if the likely origin is in these zones, additional traits should be used to help judge seed origin. For example,, the patterns of variation given for cone dimensions, seed weight, and needle characteristics reported by Squillace (1966) could be of help in judging seed origin along with monoterpene composition. This is in line with a recommendation by Kleinschmit and Spethman (1977) that morphological and physiological traits be utilized along with biochemical ones.

Likewise, historical records may be of help. Mr. R. Aaron Jordan, who was in charge of seed procurement for the Florida Division of Forestry for many years, provided the following information for slash pine. Plantations established between 1930—1934 frequently were from seed or wild seedlings in the central portion of zone 11. Those established in 1934—1947 were frequently from seed collected in the northwest part of zone 11 in Florida. Those established in 1947—1966 were frequently from seed collected in a 12-county area centering at the southeast portion of zone 6 in Georgia.

Discussion and Conclusions

We conclude that good judments of the origin of slash pine plantations can be made using monoterpene composition. The likelihood procedure is a useful quantitative technique to rank alternative zones of suspected origin. Ideally, at least 50 sample trees should be used for seed source identification, but smaller sample sizes ca be used. Information such as morphological traits and historical records should also be utilized to the extent possible .

Literature Cited

Bergmann, F.: Identification of forest seed origin on the basis of isoenzyme gene frequencies. Allgemeine Forst und Jagdzeitung **146**: 191—195 (1975). — Bergmann, F.: Identification of forest seed origin by means of isoenzyme gene frequencies. EEC Symposium on Forest Tree Biochemistry, Brussels. p. 151—162 (1977). — Chew, V.: Point estimation of the parameter of the binomial distribution. The Amer. Stat. **25**: 47—50 (1971). — Gansel, C. R. and A. E. Squillace: Geographic variation of monoterpenes in cortical oleoresin of slash pine. Silvae Genetica **25**: 150—154 (1976). — Kleinschmit, J. and W. Spethman: Possibilities and problems in the application of biochemical methods to the identification of forest trees. EEC Symposium on Forest Tree Biochemistry, Brussels. p. 263—283 (1977). — Squillace, A. E.: Geographic variation in slash pine. For. Sci. Mono. **10**, 56 p. (1966). — Squillace, A. E.: Monoterpene composition of cortical oleoresin in *Pinus elliottii* and its utility in genetics research. EEC Symposium on Forest Tree Biochemistry, Brussels. p. 39—53 (1977 a). — Squillace, A. E.: Use of monoterpene composition in forest genetics research with slash pine. In: 14th So. For. Tree Imp. Conf., Proc. 227—238 (1977 b).