Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 2540–2544, March 1991
Evolution

Interspecific gene flow in sympatric oaks

(hybridization/introgression/chloroplast DNA/*Quercus*)

ALAN T. WHITTEMORE* AND BARBARA A. SCHAAL

Department of Biology, Campus Box 1137, Washington University, St. Louis, MO 63130

Communicated by Peter H. Raven, December 5, 1990 (received for review October 1, 1990)

ABSTRACT **Variation of chloroplast DNA and nuclear ribosomal DNA (DNA encoding ribosomal RNA) was studied for five species of white oak native to the eastern United States. Although these species differ in many morphological characters and have different (though overlapping) geographical ranges and ecological tolerances, they are interfertile and often grow in mixed stands, and hybrids are occasionally found in nature. All individuals studied were morphologically typical members of their respective species—i.e., showed no evidence of recent hybrid ancestry. Restriction site markers in the chloroplast DNA reveal several clear cases of localized gene exchange between species, showing that there is appreciable gene flow between sympatric species in this group. One length variant of the nuclear ribosomal DNA, however, is species specific. The sharp morphological and ecological differences between the species, together with the one ribosomal DNA variant, suggest that nuclear genes may be exchanged less freely between species than are chloroplast genotypes.**

Natural hybrids are found in many groups of plants and animals, and they form a rewarding subject for evolutionary study. Not only is introgression a potentially important source of genetic variation in natural populations, but also the study of hybrids allows the genetic relationships between related species to be assessed. Many recent evolutionary studies have focused on cases of hybridization in narrow zones of contact between pairs of geographically or ecologically isolated species and subspecies (1). Molecular studies of such parapatric species pairs have shown that alleles derived from introgression may be found far outside the zone of morphologically recognizable intermediates (2–8). Natural hybrids are also formed between pairs of sympatric species in many genera, even species that commonly grow mixed together over large parts of their ranges. However, molecular tools have not yet been used to investigate the extent of gene flow in cases of hybridization between sympatric species that frequently grow intermixed, despite its great potential for introducing new alleles into natural populations.

The genus *Quercus* (the oaks) is outstanding for the very poor development of sterility barriers between its species (9, 10). Oak species are interfertile in many combinations, and natural hybrids may be formed between pairs of species that are very different from one another both morphologically and physiologically (11–13). Although some pairs of interfertile species show strong ecological separation (14, 15), many interfertile species pairs show extensive ecological overlap; these may form mixed stands over extensive areas of sympatry and remain distinct despite occasional hybridization.

The white oaks of the eastern United States form a large complex of interfertile species that commonly form mixed stands over large areas of sympatry. The group consists of about 16 species, which differ from one another in many morphological features (16–18). The species have different geographical and ecological ranges, but these ranges overlap broadly, and different species commonly grow in mixed stands. Patterns of hybridization within this complex were reviewed by Hardin (19). All but two of the species are known to form natural hybrids with others in the group; notable are *Quercus alba* and *Quercus stellata*, each known to form natural hybrids with 11 other species in the eastern United States (*Q. stellata* also hybridizes with at least two western species in central Texas). Although many hybrid combinations are known, hybrids are quite rare in natural populations of oaks; where they do occur it is usually only as scattered F_1 individuals, without recognizable later-generation hybrids. Very rarely, however, complex hybrid swarms are found; then species boundaries may break down very locally (19).

The factors controlling hybridization in oaks are not well understood, but it is generally agreed that selection against intermediate phenotypes must be important (19, 20). However, it is difficult to determine whether hybrids are being eliminated before they can pass on their genes, or whether some of them are actually contributing to the variability of the parental species via backcrossing. Oak species are quite variable morphologically (21), and they show a great deal of convergent evolution of morphological characters (22), so that it is difficult to find distinctive genetic markers or to be sure of the homologies of morphological markers. Consequently, it is difficult to be sure from morphological studies whether there is significant interspecific gene flow, masked by strong selection for a limited number of genes controlling striking morphological and physiological features.

Organellar DNAs (i.e., chloroplast DNA and mitochondrial DNA) show great promise for tracing the long-term effects of hybridization in natural populations. Because their inheritance is uniparental and asexual, groups of associated restriction sites are not separated by recombination, so that a great deal of historical information is preserved in these sequences. A cladogram prepared from an organellar DNA therefore reflects only part of the ancestry of the organisms (usually the direct female line), but the ancestry of an organellar genotype will remain recognizable even after many generations of sexual reproduction. If different species are genetically isolated, then conspecific individuals should be grouped together in an organelle-derived cladogram. However, if there has been gene flow between species, then conspecific individuals may be scattered through different parts of the cladogram, but geographically localized genotypes should be shared between different species.

Chloroplast DNA is the organellar sequence of choice for evolutionary studies in higher plants because of its high copy number, its evolutionary conservatism, and the rarity of large insertions, deletions, and rearrangements (23). To provide a nuclear DNA marker for comparison, variation in the nuclear ribosomal DNA was also studied. Ribosomal DNA is a tandemly repeated chromosomal sequence, present in very high copy number, which codes for three of the four ribo-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

*Present address: Regional Plant Introduction Station, 1109 Experiment Street, Griffin, GA 30223.

somal RNAs. While the coding region of the repeat is extremely conservative, the intergenic spacer varies among closely related individuals and is useful for evolutionary study at lower taxonomic levels (24).

Variation in chloroplast DNA and nuclear ribosomal DNA restriction patterns was studied for several sympatric species of white oak in the eastern United States. The distribution of DNA variants in different species and geographical areas was then used to assess the genetic effects of hybridization in white oaks of the eastern deciduous forest.

MATERIALS AND METHODS

Five species of white oak native to the eastern United States, *Quercus alba* L., *Quercus macrocarpa* Michx., *Quercus michauxii* Nutt., *Quercus stellata* Wang., and *Quercus virginiana* var. *fusiformis* (Small) Sarg., were studied; a total of 128 individuals from 18 populations were examined (Table 1). All individuals studied were typical members of their species; no morphologically recognizable hybrids were encountered at any of the study sites. Samples of 12–19 individuals were studied from each of three representative populations; since the level of within-population polymorphism was low (one or two genotypes per population), the sample size was reduced to six individuals for most of the remaining populations. *Quercus emoryi* Torr., a member of *Quercus* subgenus *Erythrobalanus*, was used as the outgroup. Vouchers for all populations are deposited at the herbarium of the Missouri Botanical Garden (St. Louis).

DNA was extracted from expanding leaves (collected in the spring) or winter buds (late summer and fall). Plant material ground to a powder in liquid nitrogen (≈0.5 g) was extracted in 5 ml of 2× cetyltrimethylammonium bromide (CTAB) extraction buffer (25) with 5% (wt/vol) polyvinylpyrrolidone (PVP-10; Sigma) at 37°C for 15–20 min (leaf material) or 1 hr (bud material); the supernatant was then washed twice with 1 vol of methylene chloride. DNA was precipitated with ⅔ vol of isopropyl alcohol, redissolved in 1 ml of Tris/EDTA buffer (1 mM Tris·HCl, pH 8.0/0.1 mM Na_2 EDTA) plus 100 μl of 2 M NaOAc (pH 5.6), and precipitated with 2.5 ml of 95% (vol/vol) ethanol. Restriction digestions were performed according to the suppliers' specifications; electrophoresis, blotting, and probing followed ref. 26, except that gels were blotted onto nylon (Micron Separations, Westboro, MA) instead of nitrocellulose, and nylon was exposed to long-wavelength UV light for 15 sec to cross-link the DNA prior to baking.

A preliminary assessment of chloroplast DNA restriction-site variation was carried out by surveying one to three individuals per population from 15 of the 18 populations in Table 1 (populations not included were *Q. alba* from Creve Coeur, Missouri, and *Q. macrocarpa* from Kansas and from Houston Co., Texas). This initial survey used 15 restriction enzymes, all five- or six-base cutters (*Bam*HI, *Dra* I, *Eco*RI, *Eco*RV, *Hin*dIII, *Kpn* I, *Pvu* II, *Sac* I, *Sac* II, *Sca* I, *Sma* I, *Sph* I, *Tth*111I, *Xho* I, and *Xmn* I). Blots were probed with a series of chloroplast clones from *Petunia hybrida*, which we have courtesy of J. D. Palmer (Indiana University, Bloomington). *Petunia* probes numbered 1, 3, 4, 6, 8, 10, and 12 (see map in ref. 27) were used; these total about 90 kilobases (kb) (subtracting sequences duplicated in the inverted repeat) and represent about 73% of the chloroplast genome. After completion of the initial survey, the remaining individuals were studied, using only the enzymes and probes that showed restriction-site polymorphisms in the initial survey. Cladistic analysis was carried out by hand. The outgroup (*Q. emoryi*) was investigated with only those probe–enzyme combinations necessary for polarization of ingroup characters.

Nuclear ribosomal DNA variation was assessed by using the same blots as the chloroplast DNA studies and probing with pGmr-1, a clone of the whole ribosomal DNA repeat from *Glycine max* (28). Preliminary mapping experiments indicate that this probe binds to *ca.* 80% of the *Quercus* ribosomal DNA repeat. Blots from the initial chloroplast survey described above were used to assess restriction site variation and length variation in the sequence. Restriction site variation was rare, and was not investigated further. Length variants for all individuals were characterized by using digests of *EcoRV*, one of several restriction enzymes that cuts the ribosomal DNA repeat only once.

RESULTS

The chloroplast DNA survey using 15 restriction enzymes yielded only eight polymorphisms (Table 2). All of these are consistent with single restriction site gains or losses except for polymorphisms 3A–C; these all involve the same enzyme and the same DNA restriction fragment. The low level of

Table 1. Locations of the populations studied, with sample sizes and the distribution of genotypes A–G

Quercus species	Location	Genotypes
Q. alba	1. Green Co., WI	13 F + 2 A
	2. Creve Coeur, St. Louis Co., MO	4 A + 2 E
	3. Tyson Reserve, St. Louis Co., MO	15 A + 4 D
	4. Jefferson Co., AR	5 A
	5. San Jacinto Co., TX	6 B
Q. emoryi	6. Arco Iris, Chihuahua, Mexico	1 H
Q. macrocarpa	7. Fairbault Co., MN	6 A
	1. Green Co., WI	6 F
	2. Creve Coeur, St. Louis Co., MO	6 A
	8. Osborne Co., KS	6 C
	9. Madison Co., TX	5 A + 1 B
	10. Houston Co., TX	6 B
Q. michauxii	11. Cleveland Co., AR	6 A
	5. San Jacinto Co., TX	6 B
Q. stellata	3. Tyson Reserve, St. Louis Co., MO	12 A
	4. Jefferson Co., AR	5 A
	12. Bastrop Co., TX	1 A
	13. Travis Co., TX	6 G
Q. virginiana	13. Travis Co., TX	6 G

Localities are numbered from 1 through 13; populations of different species from the same locality represent collections from mixed stands of two species.

Table 2. Chloroplast DNA polymorphisms

Character	Enzyme	Probe	Fragment size, kb	
			Primitive	Derived
1	*Bam*HI	3	3.3	2.0 + 1.3
2	*Bam*HI	7	7.7	5.0 + 2.7
3A	*Dra* I	1	2.0	2.2
3B	*Dra* I	1	2.0	2.4
3C	*Dra* I	1	2.4	1.5
4	*Dra* I	3	6.1 + 3.3	9.4
5	*Eco*RI	1 + 4	4.6	2.5 + 2.1
6	*Eco*RV	7	6.1	4.4 + 1.7
7	*Xmn* I	4	5.2	3.1 + 2.1
8	*Xmn* I	8	2.0 + 1.5	3.5

Numbering of cloned probes follows Sytsma and Schaal (27). Characters 3 and 7 are in the small single-copy region, character 5 is in the inverted repeat, and all others are in the large single-copy region.

restriction-site variation in the chloroplast DNA of these species is apparently characteristic of the family Fagaceae as a whole (29). However, the resulting cladogram (Fig. 1) is free of homoplasy. The majority of individuals examined share a single chloroplast DNA restriction pattern, but there are several distinct genotypes that may be used as genetic markers.

Except for the common genotype (genotype A in Fig. 1), all chloroplast genotypes are geographically localized. However, only the two rarest genotypes (D and E) are confined to single species. There are three clear cases of local sharing of distinctive chloroplast genotypes among species. Genotype F is confined to a single location, a mixed stand of *Q. alba* and *Q. macrocarpa* from southern Wisconsin, where it is the most common genotype in both species (Fig. 2). Genotype B is found only in a small area in eastern Texas, but it predominates in all three species (*Q. macrocarpa*, *Q. alba*, and *Q. michauxii*) sampled from this area. Genotype G is found in the single population of *Q. virginiana* studied and also in *Q. stellata* from the same site, although it is absent from all other populations examined.

The nuclear ribosomal DNA showed little restriction site variation, but appreciable length variation was observed. All plants examined contained repeat types between 9 and 10.5 kb in length, each individual having from one to three repeat types in this length range. Variation within this range is high

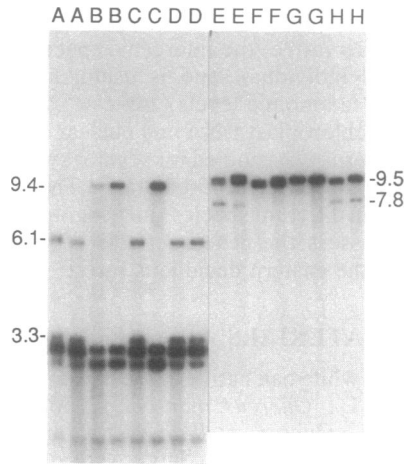

FIG. 2. Selected restriction fragment polymorphisms. Each lane represents an individual tree. A–D, distribution of character 4 (chloroplast probe 3, Table 2). A, *Q. macrocarpa* (Minnesota); B, *Q. macrocarpa* (Wisconsin); C, *Q. alba* (same site as B); D, *Q. alba* (Creve Coeur, Missouri). E–H, distribution of the 7.8-kb nuclear ribosomal DNA repeat. E, *Q. alba* (Texas); F and G, *Q. stellata* (F, Travis Co., Texas, and G, Arkansas); H, *Q. macrocarpa* (Texas).

within populations, and these length variants are not useful for comparing different species or localities. Variants in the 9- to 10.5-kb range were the only repeat lengths present in the 30 individuals of *Q. stellata* and *Q. virginiana* examined. However, all 98 individuals of *Q. alba*, *Q. macrocarpa*, and *Q. michauxii* showed a distinctive short repeat only 7.8 kb long in addition to the longer repeats (Fig. 2). The distribution of this short repeat type is thus species specific in the group.

DISCUSSION

The distribution of chloroplast genotypes in this group indicates that sympatric species of oak in the eastern United States do not represent fully isolated gene pools, but are actively exchanging genes. The cladogram derived from the chloroplast data does not reflect the species boundaries, but it is concordant with the geographical location of the populations. Indeed, in all six cases where mixed stands of two species were examined the commonest chloroplast type in both species was the same. The case of *Q. stellata* and *Q.*

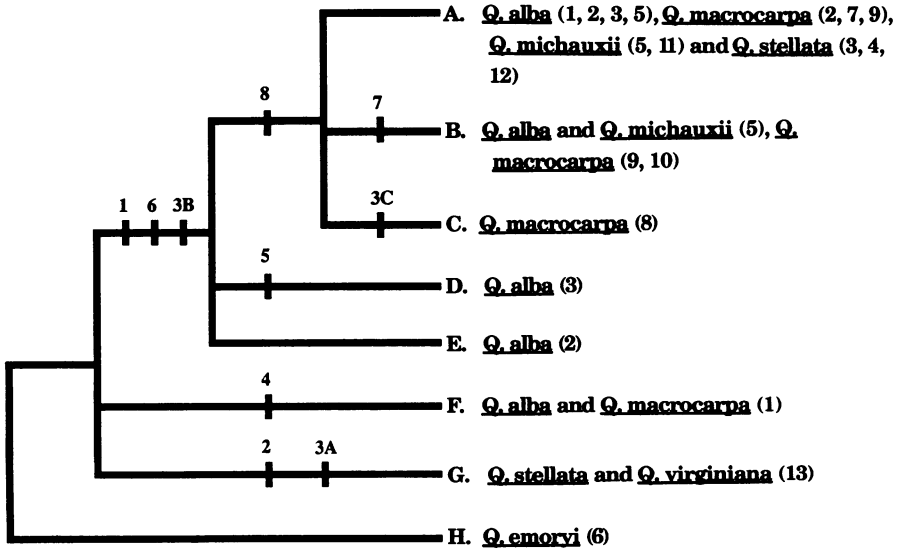

FIG. 1. Cladogram showing the relationships of genotypes A–G. Localities (given in parentheses after the species) are numbered as in Table 1, and characters (given above the short vertical bars) as in Table 2.

virginiana in Travis Co., Texas, is especially interesting, as *Q. virginiana*, an evergreen species, is very different morphologically from the deciduous oak species and the two groups never form extensive hybrid swarms (30). Thus, the chloroplast DNA data suggest that genes are exchanged even between pairs of species that are distantly related and show limited ability to hybridize.

The samples taken from sites where two species grow mixed together can be used to generate a numerical comparison of gene flow rates between localities and between species, using the private allele method of Slatkin (31). The sample includes six mixed stands (Green Co., Wisconsin; Creve Coeur and Tyson, Missouri; Jefferson Co., Arkansas; and San Jacinto Co. and Travis Co., Texas), with a total of six genotypes present at these six sites. Five of these six genotypes are confined to single localities, and are thus private alleles with respect to geographical subdivisions of the complex, while only two are confined to single species, and are private with respect to taxonomic subdivisions. Slatkin's correction for a sample size of 5 per population may be used since it is unlikely that randomly chosen subsamples of 5 from any of the populations studied would miss any of the private alleles. The estimated number of migrant chloroplasts per generation (Nm) across these barriers is $Nm = 0.075$ between localities and $Nm = 0.475$ between species at the same locality; the actual number of migrants per generation is higher (corrections for haploid organellar sequences are discussed in ref. 32). These calculations suggest that a typical population of oaks experiences a level of gene flow from other species in the same geographical area very much greater than the level of gene flow from distant conspecific populations. The strong geographical isolation is not surprising, since the seeds are probably less mobile than the pollen in *Quercus*, and the inheritance of chloroplasts is typically maternal in angiosperms. However, the level of interspecific gene flow suggested by these data is very high, considering the scarcity of morphologically recognizable hybrids in the field (as noted above, no recognizable intermediates were seen at any of the sites sampled) and the fact that oaks were absent from much of eastern North America during the Pleistocene and reinvaded the central and northern part of the study area only within the past 10,000–15,000 years [i.e., ≈100 generations (33)]. The figure of $Nm = 0.475$, corresponding to an Nm of 0.95 from a diploid nuclear locus (32), is higher than rates of gene flow between populations within many plant species (34).

Sympatric oak species are able to remain distinct despite considerable introgression, so that species concepts that rely on total genetic isolation between species to explain their distinctness clearly are not applicable in *Quercus* (19, 35, 36). A model more appropriate to oaks considers species as adaptive peaks, in which the tendency of the species to merge due to introgressive gene flow is balanced by selection for groups of coadapted alleles (19). Unfortunately, there are no data to suggest how many loci such coadapted complexes might consist of or how strong the selection on such groups of alleles is likely to be in oaks. Linkage is a strong force constraining recombination within the nuclear genome, so that strong concerted selection acting on loci scattered through the nuclear genome could greatly reduce interspecific exchange of other nuclear alleles (37). Therefore, it is quite possible that the high level of gene flow shown by the chloroplast DNA may not represent the situation in the oak genome as a whole.

Several studies of parapatric hybridization in other taxa have confirmed that extensive exchange of organelles may be coupled with sharp differentiation at nuclear loci (2, 3, 5, 6). The five species of white oak studied here are well differentiated with respect to many morphological characters, isozyme loci (38), and probably, judging from their different ecological and geographical ranges, many physiological traits. As noted above, the distribution of one distinctive nuclear DNA marker, the short (7.8-kb) ribosomal DNA repeat type, also follows taxonomic boundaries in these oak species. These data suggest that nuclear genes may not be so freely exchanged among these species. Further investigation of relative interspecific gene flow rates shown by nuclear and organellar DNAs could provide insight into the nature of the adaptive barrier that isolates such interfertile species.

The demonstration that chloroplast DNA in natural populations may be acquired from distantly related organisms, so that relationships shown by this molecule may not be representative of the genome as a whole, suggests that chloroplast DNA data must be used with caution in cladistic studies of plant relationships where hybridization may occur. It is difficult to determine how common introgressive plastid exchange may be in plants, since many published studies of chloroplast DNA variation are based on very small samples, often only one individual per species. Such small samples would not have been adequate for understanding the genetic relationships among the oak species studied here. The data presented here, together with other recent studies showing the coexistence of different chloroplast DNA lineages within single species (27, 39, 40), suggests that hypotheses of relationships need to be based on adequate surveys of within-species variation, preferably coupled with analyses of data derived (directly or indirectly) from other parts of the genome.

We thank Mark Eberle, Sherry Fergusson, Carl Keener, Lloyd Stark, and Michael Warnock for help collecting the plant material; Michael Cain, Eric Routman, Victoria Sork, Richard Spellenberg, and Billie Turner for helpful discussions; and the Alfred P. Sloan Foundation for funding this project.

1. Barton, N. H. & Hewitt, G. M. (1985) *Annu. Rev. Ecol. Syst.* **16**, 113–148.
2. Powell, J. R. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 492–495.
3. Ferris, S. D., Sage, R. D., Huang, C.-M., Nielsen, J. T., Ritte, U. & Wilson, A. C. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 2290–2294.
4. Gyllensten, U. & Wilson, A. C. (1987) *Genet. Res.* **49**, 25–29.
5. Tegelstrom, H. (1987) *J. Mol. Evol.* **24**, 218–222.
6. Marchant, A. D., Arnold, M. L. & Wilkinson, P. (1988) *Heredity* **61**, 321–328.
7. Keim, P., Paige, K. N., Whitham, T. G. & Lark, K. G. (1989) *Genetics* **123**, 557–565.
8. Rieseberg, L. H., Beckstrom-Sternberg, S. & Doan, K. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 593–597.
9. Stebbins, G. L., Jr. (1950) *Variation and Evolution in Plants* (Columbia Univ. Press, New York).
10. Grant, V. (1981) *Plant Speciation* (Columbia Univ. Press, New York), 2nd Ed.
11. Cottam, W. P., Tucker, J. M. & Drobnick, R. (1959) *Ecology* **40**, 361–377.
12. Tucker, J. M. (1952) *Evolution* **6**, 162–180.
13. Tucker, J. M. (1968) *Madroño* **19**, 256–266.
14. Muller, C. H. (1952) *Evolution* **6**, 147–161.
15. Salisbury, E. J. (1940) in *The New Systematics*, ed. Huxley, J. S. (Clarendon, Oxford), pp. 329–340.
16. Harlow, W. M. & Harrar, E. S. (1950) *Textbook of Dendrology* (McGraw-Hill, New York), 3rd Ed.
17. Hardin, J. W. (1979) *Bull. Torrey Bot. Club* **106**, 313–325.
18. Solomon, A. M. (1983) *Am. J. Bot.* **70**, 481–494.
19. Hardin, J. W. (1975) *J. Arnold Arbor. Harv. Univ.* **56**, 336–363.
20. Stebbins, G. L., Jr., Matzke, E. B. & Epling, C. (1947) *Evolution* **1**, 79–88.
21. Baranski, M. J. (1975) *Bull. Agric. Exp. Stn. (N.C.)* **236** (i–iv), 1–176.
22. Tucker, J. M. (1974) *Taxon* **23**, 129–154.
23. Palmer, J. D., Jansen, R. K., Michaels, H. J., Chase, M. W. & Manhart, J. R. (1988) *Ann. Mo. Bot. Gard.* **75**, 1180–1206.

24. Schaal, B. A. & Learn, G. H., Jr. (1988) *Ann. Mo. Bot. Gard.* **75,** 1207–1216.
25. Doyle, J. J. & Doyle, J. L. (1987) *Phytochem. Bull.* **19,** 11–15.
26. Learn, G. H., Jr., & Schaal, B. A. (1987) *Evolution* **41,** 433–438.
27. Sytsma, K. J. & Schaal, B. A. (1985) *Evolution* **39,** 594–608.
28. Doyle, J. J. & Beachy, R. N. (1985) *Theor. Appl. Genet.* **70,** 369–376.
29. Manos, P. S., Doyle, J. J. & Nixon, K. C. (1989) *Am. J. Bot.* **76,** Suppl., 258 (abstr.).
30. Muller, C. H. (1961) *Am. Midl. Nat.* **65,** 17–39.
31. Slatkin, M. (1985) *Evolution* **39,** 53–65.
32. Birky, C. W., Jr., Fuerst, P. & Maruyama, T. (1989) *Genetics* **121,** 613–627.
33. Davis, M. B. (1983) *Ann. Mo. Bot. Gard.* **70,** 550–563.
34. Govindaraju, D. R. (1988) *Oikos* **52,** 31–35.
35. Burger, W. C. (1975) *Taxon* **24,** 45–50.
36. Van Valen, L. (1976) *Taxon* **25,** 233–239.
37. Barton, N. & Bengtsson, B. O. (1986) *Heredity* **57,** 357–376.
38. Guttman, S. I. & Weigt, L. A. (1989) *Can. J. Bot.* **67,** 339–351.
39. Doyle, J. J., Doyle, J. & Brown, A. D. H. (1990) *Am. J. Bot.* **77,** 772–782.
40. Doebley, J. (1989) *Evolution* **43,** 1555–1559.