

both of these core traits exhibit a strong genetic correlation with relevant whole-tree traits (i.e. objective traits; refer to RAYMOND et al., 2001), these findings suggest that only one core needs to be extracted from trees to assess both density and wood chemistry in eucalypt breeding programmes. The adoption of this practice, combined with the use of solid wood rather than ground samples to measure NIR spectra (POKE and RAYMOND, 2006), could considerably reduce the cost of assessing wood properties in eucalypt pulpwood breeding programmes.

Acknowledgements

We thank LINDA BALLARD, CHRIS HARWOOD, LEON SAVAGE, Forestry Tasmania, Norske Skog Ltd., the CRC for Sustainable Production Forestry, the Australian Research Council (Linkage grant LP0453704) and Ensis.

References

AS/NZS 1301.002s (2004): Methods of test for pulp and paper. Method 002s: Preparation of wood samples for chemical analysis. Joint publication of Standards Australia, Homebush, NSW and Standards New Zealand, Wellington, NZ.

DUTKOWSKI, G. W., B. M. POTTS, D. R. WILLIAMS, P. D. KUBE and C. MCARTHUR (2001): Geographic genetic variation in central Victorian *Eucalyptus nitens* in

IUFRO Symposium on Developing the Eucalypt for the Future, CD-Rom communication, Valdivia, Chile. 10–15 September. Instituto Forestal.

GILMOUR, A. R., B. J. GOGEL, B. R. CULLIS, S. J. WELHAM, and R. THOMPSON (2002): ASReml User Guide Release 1.0. VSN International Ltd, Hemel Hempstead, UK.

GRIFFIN, A. R. and P. P. COTTERILL (1988): Genetic variation in growth of outcrossed, selfed and open-pollinated progenies of *Eucalyptus regnans* and some implications for breeding strategy. *Silvae Genetica* **37**: 124–131.

POKE, F. S. and C. A. RAYMOND (2006): Predicting extractives, lignin, and cellulose contents using near infrared spectroscopy on solid wood in *Eucalyptus globulus*. *Journal of Wood Chemistry and Technology* **26**: 187–199.

RAYMOND, C. A., L. R. SCHIMLECK, A. MUNERI and A. J. MICHELL (2001): Nondestructive sampling of *Eucalyptus globulus* and *E. nitens* for wood properties. III. Predicted pulp yield using near infrared reflectance analysis. *Wood Science and Technology* **35**: 203–215.

RAYMOND, C. A. and L. R. SCHIMLECK (2002): Development of near infrared reflectance analysis calibrations for estimating genetic parameters for cellulose content in *Eucalyptus globulus*. *Canadian Journal of Forest Research* **32**: 170–176.

SAS INSTITUTE INC. (2002): The SAS system for Windows, Version 9.1.

TAPPI (1989): Basic density and moisture content of pulpwood. no. T258 om-98, Technical Association of the Pulp and Paper Industries (TAPPI).

Heterogeneity and Spatial Autocorrelation for Chloroplast Haplotypes in Three Old Growth Populations of Northern Red Oak

By Y. FENG¹, W. SUN² and J. ROMERO-SEVERSON^{1,*}

(Received 9th February 2007)

Abstract

In eastern North America, evidence for cryptic northern refugia could contribute to resolving Reid's Paradox, the disparity between the rate of oak recolonization indicated by pollen deposition and the rate indicated by contemporary seed dispersal studies. Severe anthropogenic disturbance of oak-dominated forests throughout eastern North America followed by regeneration from isolated patches and deliberate planting in some regions could obscure the signal of discontinuity expected from small, cryptic refugia. In this study of northern red oak, *Quercus rubra* L., the dominant representative

of *Quercus* section Lobatae in the eastern United States, we address the question of appropriate sample size for accurate detection of the biogeographical distribution of chloroplast haplotype diversity in *Q. rubra*. We examined chloroplast DNA (cpDNA) variation in all *Q. rubra* over 17 cm in diameter (310 trees) in three forest fragments with documented histories of minimal disturbance for the last 100–190 years. cpDNA polymorphisms in three intergenic regions revealed different haplotype frequencies between the two local fragments located within 1 km of each other and complete discontinuity for the predominant haplotype between these two sites and a site 207 km distant. Haplotypes displayed positive spatial autocorrelation over 10–40 meter distances. Sample sizes of 10 or fewer taken at 50 meter intervals along a linear transect yielded poor estimates of haplotype frequencies and did not accurately detect haplotype richness.

Key words: cpDNA, *Quercus rubra*, chloroplast, genetic diversity, patch size, postglacial migration.

¹ University of Notre Dame, Department of Biological Sciences, 327 Galvin Life Sciences, Notre Dame, IN 46556-5645.

² Purdue University, Department of Entomology, West Lafayette, IN 47907.

*) Corresponding author: University of Notre Dame, Department of Biological Sciences, 327 Galvin Life Sciences, Notre Dame, IN 46556-5645, Fax: 574 6317413. E-mail: jromeros@nd.edu

Introduction

Oaks, chestnuts and beeches (Fagaceae) are dominant angiosperm forest trees in the temperate forests of the northern hemisphere. The oaks in *Quercus* section Lobatae, the red oak section of the *Quercus* genus, are native to the Americas and include species in the eastern deciduous forests of North America. Northern red oak (*Quercus rubra* L.), an upland species within the eastern members of the Lobatae, ranges from Ontario to southern Alabama and east from Oklahoma to the Atlantic coast.

Range shifts resulting from Paleoclimate fluctuations may have had a significant impact on the biogeographical distribution of genetic diversity in northern red oak. During the last glacial maximum in North America (18–20,000 BP), the Wisconsinian extension of the Laurentide ice sheet covered portions of the northern states in the eastern United States, causing forests to retreat southward and later reestablish as the ice sheet retreated (DELCOURT and DELCOURT, 1987; JACKSON *et al.*, 2000; SCHLARBAUM *et al.*, 1982). The palynological record suggests that oaks recolonized at a rate far in excess of that inferred by the behavior of contemporary seed dispersal agents (JACKSON *et al.*, 2000). Several investigators (MCLACHLAN *et al.*, 2004; STEWART and LISTER, 2001) have suggested that this inconsistency, known as Reid's paradox (CLARK *et al.*, 1998; REID, 1899), arises from a misinterpretation of the pollen record. Pollen deposition may have been too thin for consistent detection if small populations of northern red oaks persisted in cryptic refugia north of the main population at the last glacial maximum. The contemporary northern edge of the range for *Q. rubra* (~ 48° N) and the presence of trace amounts of oak pollen in northern lake sediments suggests that this species could have persisted in mesic microhabitats in locations as far north as Tennessee. Although the eastern United States lacks geographical barriers comparable to the Alps, the Appalachian-Blue Ridge forest ecoregion was a mesic and thermal refuge for other species during the Pleistocene glaciations and as such could have provided refuge for *Q. rubra* and other oaks. Oak recolonization from these cryptic refugia could have resulted in biogeographical discontinuities in genetic diversity as cryptic populations recolonized northward, leaving descendants with alleles that may have been lost in unsheltered populations as the range shifted southward.

Large-scale changes in rainfall patterns also induced Paleoclimate range shifts along longitudinal gradients. Pollen records indicate that by 10,000 yr BP oak hardwood forests had advanced to the western edges of Missouri, Iowa and Minnesota and then retreated east as the Hypsithermal drying interval 8,700–5,000 yr BP favored the eastward expansion of prairie grassland (BAKER *et al.*, 2002). As the tall grass prairie peninsula extended eastward to eastern Indiana and southwest Michigan, *Q. rubra* populations in the north were partially isolated from *Q. rubra* populations south of the peninsula. A tall grass prairie and deciduous forest ecotone still dominates southern Iowa, southern Wisconsin, central Illinois and the northwest corner of Indiana, sep-

arating the northern hardwood forest from the southern hardwood forest. Fire, whether natural or deliberately set, also played an important role in the maintenance of the open woodlands present in this region prior to European colonization (FRALISH *et al.*, 1991). In southern Minnesota, Iowa and northern Missouri, the contemporary western edge of the *Q. rubra* range lies 200–500 km east of the western edge of the oak hardwood forest 10,000 yr BP. Thus, the contemporary biogeographical distribution of genetic diversity in northern red oak may be result of complex recolonization patterns in some regions.

The spatial pattern of northern red oak regeneration within stands, the rate of colonization into new habitats and the spatial pattern of this colonization depend primarily on the caching behavior of small mammals and birds. Squirrels (*Sciurus carolinensis* (Gmelin) and other *Sciurus* species) transport and cache northern red oak acorns an average of 14 m from the maternal tree (SMALLWOOD *et al.*, 1998), although some dispersal distances exceed 30 m (STEELE *et al.*, 2001). Blue jays (*Cyanocitta cristata* L.) can transport acorns at least 4 km (JOHNSON *et al.*, 1997; JOHNSON and WEBB, 1989). Field observations (JOHNSON *et al.*, 1997) indicate that blue jays cache preferentially in regenerating woody habitats, forest canopy gaps and forest edges.

Oak embryos contain proplastids, the organelles that will become chloroplasts after seed germination. Matrilineal inheritance of the haploid proplastid genome in the genus *Quercus* enables investigators to detect successful seed dispersal through genetic analysis of the chloroplast genomes present in mature trees. As red oak seedlings are suppressed by shade, seedlings tend to be more successful at edges and within gaps, favoring the descendants of trees already in the stand. Lastly, the caching behavior and foraging ranges of blue jays favor edges and gaps within existing stands. Maternal chloroplast inheritance, seed dispersal mechanisms and seedling suppression as the canopy closes all suggest that only a few chloroplast haplotypes will predominate within local forest fragments and that the matrilineal descendants of the original colonists will persist in local populations, permitting reconstruction of recolonization patterns (PETIT *et al.*, 2002).

The expected chloroplast haplotype richness and distribution within stands merits serious consideration in investigations of recolonization rates and patterns, given the large natural range of many *Quercus* species and limited program resources. Seed dispersal simulations designed to include both local and rare long distance dispersal events produce spatially patchy haplotype patterns (DAVIES *et al.*, 2004). If most of the chloroplasts in a population are descended from a single maternal line, then sampling a few trees at many locations would be more productive than sampling many trees at only a few locations (PONS and PETIT, 1995). Using the “few trees at many locations” strategy to examine chloroplast DNA polymorphisms in 2613 populations of European white oaks (PETIT *et al.*, 2002), European investigators detected three distinct routes of oak recolonization and evidence for haplotype mixing as

populations from different refugia met and then continued the recolonization process. The first chloroplast haplotype study for *Q. rubra* also used the “few trees at many locations” sampling strategy and reported a population differentiation value ($G_{ST} = 0.731$) close to the G_{ST} range (0.781–0.961) reported for European white oaks (ROMERO-SEVERSON *et al.*, 2003). Using sample sizes of 2–10 per population, a second study (MAGNI *et al.*, 2005) detected 12 chloroplast haplotypes for five polymorphic loci in 66 populations of *Q. rubra* but reported a much lower value for population differentiation ($G_{ST} = 0.46$). The difference in these estimates is unexpected, given that the first study was limited to the state of Indiana, while the second was range-wide. The two studies used similar methodology, a very low number of samples per site, which may be the primary cause of this disparity.

If the trees are randomly distributed within sites and there is little or no spatial autocorrelation with respect to chloroplast haplotypes within sites, then sampling strategies can be efficiently designed using existing statistical techniques for estimating the effects of sample size on the allele richness detected and the magnitude of fixation indices. Detection of phylogeographical structure among sites, the estimation of effective population sizes and estimation of the effect uniparental versus diparental inheritance on spatial heterogeneity (CHESSER and BAKER, 1996) all ultimately depend on the degree to which the sampling strategy employed provides accurate information for each site sampled.

Maternal inheritance of the chloroplast genome, the high impact of long distance seed dispersal events on migration rate, seedling suppression in shade and the caching behaviors of blue jays all suggest that nonrandom distribution of oaks in contemporary stands and high spatial autocorrelation with respect to haplotypes may be typical rather than rare. If patch size, the area occupied by a founder tree and its matrilineal descendants, is comparatively small and the source population has a mixture of haplotypes, then recolonization could result in haplotype heterogeneity as islands with different haplotypes coalesce. Haplotype heterogeneity would then be a natural condition and not an artifact of anthropogenic disturbance.

The specific objectives of this study were to ascertain the distribution of chloroplast haplotypes with respect to trees by genotyping all of the mature *Q. rubra* in three old growth forest fragments with documented histories of minimal human disturbance. Two locations are in close proximity, but are historically distinct forest fragments in a glaciated region and the third location is an unglaciated site 207 km further south. In minimizing the need to make assumptions about parameters, we are then able to test four hypothesis in a straightforward manner: 1) haplotype richness will be low but not zero within a single site; 2) haplotype mixtures are common; 3) haplotypes within populations will show spatial autocorrelation; and 4) small sample sizes (≤ 10) taken along a linear transect do not yield accurate estimates of haplotype frequency and haplotype richness when haplotype mixtures are present.

Materials and Methods

Study sites

The Purdue-Davis Research Forest (PDRF) and Pioneer Mothers Memorial Forest (PMMF) are old growth forests (forests containing trees at least 150 years old). PDRF is in east-central Indiana (N 40.11° W 85.11°). Deeded to Purdue University in 1917, the PDRF contains five distinct forest fragments surrounded by open fields and managed with minimal human disturbance since 1917. The largest and smallest fragments, PD_1 (20.64 hectares) and PD_4 (2.83 hectares) lie ~ 1 km from each other. In 1926, Professor Burr Prentice of the Purdue Department of Forestry identified tree species and placed a permanent tag on each tree larger than 10.2 cm in diameter at breast height (DBH). Professor Prentice tagged all of the trees in PD_1 and 60 of the 87 trees in PD_4 that were selected for genotyping. Diameter distribution studies support the verbal history of minimal disturbance in these fragments during the last 100 years (ALDRICH *et al.*, 2005; BAKER *et al.*, 2002; PARKER *et al.*, 1985). The PMMF is located in the Hoosier National Forest in southern Indiana (N 38.33° W 86.27°). Historical records and size class distributions indicate that the central old growth (20.2 hectares) section, formerly called Cox Woods, has experienced little human disturbance and the peripheral old growth (12.2 hectares) minimal disturbance over the last 200 years (SCHMELZ and LINDSEY, 1965). The PDRF is located ~75 km north of the southernmost advance of Wisconsinian ice sheet, whereas the PMMF is located ~132 km south of glacial maximum. At PD_4, all of the *Q. rubra* > 17 cm DBH and not tagged by Professor Prentice (27 trees) were included in the study. Tissue samples were taken from all *Q. rubra* trees at the PMMF site, including the 13 trees in the smallest size class (17–30 cm DBH).

Virtual Transects

Grids of 50 by 50 meters were drawn over maps of the three study sites, with the genotyped trees in the spatial arrangement actually observed for each site. Maps included all genotyped individuals except 10 trees in PD_1, for which position data was not available. The strategy was to sample the first tree within 5 m of a line starting at the southwest corner of the fragment and then proceed north along the along this line, sample the next tree at least 50 m distant from the first and so on until either five or ten trees were sampled. If the transect reached an edge before the required number of trees was sampled, the adjacent 50 meter transect was used and the sampling continued in the direction opposite to that used before. The first north south transect started at southwest corner and the first east west transect started at the northeast corner. PD_4, the smallest fragment, had eight total transects, PD_1 15 and PMMF 25.

Plant Materials and DNA extraction

DNA from leaves or cambium samples was extracted using QIAGEN DNeasy™ Plant Mini Kits (QIAGEN Inc. Valencia, CA) as described previously (ROMERO-SEVERSON *et al.*, 2003) and stored at –20 °C until amplification.

PCR-RFLP

Amplicons from 25 trees from each of the three study sites were made using the universal primers for the CD, FV and TC regions and digested with a mix of seven restriction enzymes having compatible buffers: BamHI, EcoRI, AluI, HhaI, MspI, HaeIII and RsaI. This procedure queries 10239 bp of intergenic sequence for seven different sets of restriction sites. The restriction fragments were size separated and scored as described previously. Detailed descriptions of all three regions in northern red oak are presented in the 'Results' section.

Cleaved Amplified Polymorphic Sequence (CAPS) markers

Starting primers were synthesized by MWG-Biotech (High Point, NC). PCR conditions were: an initial denaturing step at 95°C for 5 min, followed by 36 cycles of 94°C for 30 sec, 56°C (CD_seq), 52°C (FV_seq) and 60°C (TC_seq) for 1 min and 70°C for 4 min. Reagent mixture was 1X Ex Taq™ buffer, 2.0 mM MgCl₂, 0.2 mM dNTP each, 10 pmol for each primer, 10 ng template DNA and 2.5 units of Takara Ex Taq™ DNA polymerase per 25 l reaction.

Amplicons were cloned using TA Cloning® kit for ligation and transformation (Invitrogen, Carlsbad, CA). Positive clones for each of the FV, CD and TC amplicons were identified by blue/white selection. For each amplicon, two clones were sequenced with M13 forward and reverse primers. Nested primers were then designed and used to sequence forward. Four sets of nested primers were designed for each set of amplicons then one consensus sequence assembled using Vector NTI Suite V.6 (Informax, Bethesda, MD). Predicted restriction fragment patterns for each finished sequence were

also detected with Vector NTI™. The restriction fragments responsible for the RFLP were selected, and primers designed to amplify across the polymorphic regions. The amplified DNA bands were removed from the gel under UV light and purified with QIAquick (QIAGEN Inc. Valencia, CA) gel extraction kit. Sequences for each allele were aligned to identify the polymorphism. Based on the sequencing results, primers CD234, FV 240 and TC 915 (Table 1) were designed to amplify the polymorphic regions. The PCR program included 3 min denaturation at 94°C; 39 cycles of 30 sec at 94°C; 30 sec annealing at 50.5°C for CD234, 51°C for FV240 and 57.8°C for TC 915, respectively, 30 sec extension at 70°C for CD234 and FV240 and 1 min extension for TC915 at the same temperature. For CD and FV, PCR products were electrophoresed in 2.5% TreviGel (Trevigen) and stained with ethidium bromide. Alleles were detected visually by differences in fragment length patterns. Haplotypes for 109 trees at PD_1, 62 at PD_4 and 64 at PMMF were detected with the CAPS markers. The haplotype designations (Table 2) are the same as reported previously.

Probability expression

The identification probability of the predominant haplotype is a function of sample size, if the sample of haplotypes is unbiased. A haplotype sample is unbiased if the probability of detecting any given haplotype is equal to the frequency of that haplotype in the population. The predominant haplotype *H* is defined as a haplotype shared by more than half of a given population. The probability of correctly detecting haplotype *H* for a given sample size depends on the true frequency *p* of haplotype *H*. Given *p* and a sample of size *n*, we define one success *r* as obtaining exactly

Table 1. – CAPS Primers for the polymorphic regions of CD, FV and TC.

Name	Orientation	Sequence	Notes
CD234	Forward	CCAGTTCAAATCTGGGTGTC	Consensus CD forward primer
	Reverse	CGATCCCTTTTCAATATTCA	
FV240	Forward	CACTTCACTTCGGATTTGAA	Consensus FV reverse primer
	Reverse	CCGAGAAGGTCTACGGTTCG	
TC915	Forward	GGTAGAGTAACGCCATGGTA	First 4 bp (GGTA) overlaps with consensus TC forward primer
	Reverse	GTTGTCGGAACAAGAATTC	

Table 2. – Allele combinations for CD, FV and TC for five haplotypes.

Locus	Haplotypes				
	I	II	III	IV	V
CD	2	2	2	2	1
FV	1	2	3	1	1
TC	1	1	1	2	1

$$\left[\frac{(n+1)}{2} \right]$$

individuals of haplotype *H*. The probability is given by the standard expression

$$\binom{n}{r} p^r (1-p)^{n-r}$$

For our analysis, all samples of size *n* that contain at least

$$\left[\frac{(n+1)}{2} \right]$$

individuals of the predominant haplotype *H* are successful. Then the probability of obtaining at least

$$\left[\frac{(n+1)}{2} \right]$$

individuals is the sum of the probabilities for all successful outcomes, i.e., the probability of getting

$$\left[\frac{(n+1)}{2} \right] \text{ and } \left[\frac{(n+2)}{2} \right] \text{ and } \left[\frac{(n+3)}{2} \right] \text{ and so on up to}$$

$$\left[\frac{(n+n)}{2} \right].$$

The arithmetic sum of all these probabilities correctly identifies the predominant haplotype given a true predominant haplotype frequency *p* and a sample size *n*.

Spatial autocorrelation

Equally spaced lag distance intervals of 10 m were used for PD_1 and PD_4 and 20 m for the PMMF. Moran's I was calculated using GS⁺ v. 7 (Gamma Design Software LLC, Plainwell Michigan).

Results

DNA sequence polymorphism within the CD, FV and TC regions

The restriction fragments predicted from the complete sequence for the CD, FV and TC regions matched those

actually detected with individual digests and with the multiple digest of seven restriction enzymes. CAPS amplicons produced from primers designed to amplify across the polymorphic site showed sequence identity except for the expected polymorphic region for each locus. The sizes for the detected CAPS amplicons were 234 bp and 222 bp for CD1 and CD2, respectively, and 240 bp, 231 bp and 223 bp for FV1, FV2 and FV3, respectively. The TC1 CAPS amplicon (915 bp) has no HaeIII site while the TC2 amplicon yields two fragments (755 bp and 160 bp) after HaeIII digestion.

The consensus primers for CD are anchored on the ends of the coding regions for the highly conserved genes *trnC* and *trnD*. The CD amplicon for allele 1 has 3483 bp, includes two small genes (*ycf6* and *psbM*) and has three intergenic regions. Sequence comparison of the two CD alleles detected in this study showed that the polymorphism is a 12 bp indel in the intergenic region between *trnC* and *ycf6* (Table 3). Consensus primers for FV are anchored in the coding regions of genes *trnF* and *trnV*. The amplicon for allele 1 has 3110 bp, includes three genes (*ndhJ*, *ndhK* and *ndhC*) and has four intergenic regions. Sequence comparison of the three FV alleles detected in this study showed that the polymorphism occurs in the intergenic region between *ndhC* and *trnV*. Relative to FV_1, FV_2 has a nine-nucleotide deletion, whereas FV_3 has a 17-nucleotide deletion. The deletion site of FV_2 is inside that of FV_3 (Table 3). Consensus primers for TC are anchored in the coding regions of genes *trnT* and *psbC*. The amplicon for allele 1 has 3646 bp, includes one gene (*psbD*) and has two intergenic regions. Sequence comparison of the two TC alleles revealed that the polymorphism is a SNP in the intergenic region between *trnT* and *psbD*. The transition at a single nucleotide (A-G) in TC_1 creates a restriction site for *HaeIII* in TC_2 (Table 3).

Haplotype polymorphism and richness

We detected only four of the five haplotypes (I, II, III and V) reported previously and no new haplotypes. We did not detect the SNP in the TC region in these three populations. Haplotype V is the only haplotype having allele 1 of both CD and FV (Table 2), suggesting that this haplotype may predate the others. The simplest

Table 3. – Polymorphic regions detected for the CD, FV and TC alleles.

Locus	Alleles	Sequences
CD	1	CCTTGC GGAA AGATTATATT AAAGATTATA TTAGTGAATA TTGAAAAGGG
	2	CCTTGC GGAA AGATTATATT A----- ---GTGAATA TTGAAAAGGG
FV	1	TTTATATAAA TTGATTAGAA GGA ACTAAAG GATATAAAGA GCTAAAGAAA
	2	TTTATATAAA TTGATTAGAA GGA ACT---- -----AAAGA GCTAAAGAAA
	3	TTTATATAAA TTGATTAGAA GGA----- ----- GCTAAAGAAA
TC	1	CTTATTCCAA GTCACAACAT ACGAGCCGTT TTAAATTAAA TTAAATATA
	2	CTTATTCCAA GTCACAACAT ACGGGCCGTT TTAAATTAAA TTAAATATA

Table 4. – Haplotype counts for PD_1, PD_4 and PMMF.

Population	Sample Size	Haplotypes				
		I	II	III	IV	V
PD_1	134	4	32	0	0	98
PD_4	87	58	4	0	0	25
PMMF	89	17	0	72	0	0

ordered relationship among the haplotypes given the pattern of indel polymorphism is V, I, II and III, with III being the most recent.

Haplotype frequency, distribution and diversity

At each site, a majority of the trees genotyped had the same haplotype, but all three sites had different majority or predominant haplotypes (Table 4). The observed frequency of the predominant haplotype was 0.666, 0.731 and 0.809 for PD_4, PD_1 and PMMF, respectively. Different predominant haplotypes were detected in PD_1 and PD_4 despite the small distance between them (~1 km). In PD_1, haplotype V is predominant. Only four trees have haplotype I, and the remaining 32 trees have haplotype II. In PD_4, haplotype I is predominant. Only four of the trees have haplotype II and the remaining 25 trees have haplotype V. In PMMF, the largest fragment of the three, haplotype III, is predominant.

Comparison of the 27 trees in PD_4 genotyped in this study that were not tagged by Professor Prentice with the 60 that were tagged and, therefore, likely to be older, revealed no difference in haplotype richness. The proportion of haplotype I in the subset of 60 was higher (0.75) than in the set of 27 (0.407). Three of the latter group had haplotype II. These were clustered together

on the northwestern edge of PD_4 (Figure 1). The only other tree having haplotype II lies next to this cluster, was tagged in 1926 and now has a DBH of 90 cm. Five of the 17 trees having the minority haplotype (haplotype I) in the PMMF have DBH measurements > 80 cm and one has a DBH of 105 cm.

Correct identification probability under the random distribution

The identification probability of the predominant haplotype falls sharply as the true frequency of the predominant haplotype falls (Figure 2). The observed frequencies of the predominant haplotypes in PD_4, PD_1 and PMMF sites were 0.666, 0.731 and 0.809, respectively. If the assumption of random haplotype distribution were correct, then for the PD_4, PD_1 and PMMF sites the minimum sample sizes for identifying the predominant haplotype at $P > 0.95$ are 25, 11 and 5. Even under this best-case assumption, it is clear that small differences in the true frequency of the predominant haplotype have great impact on the probability of correct identification.

Spatial autocorrelation

At all three sites, trees with the same haplotypes tended to be closer to one another than trees with different haplotypes (positive spatial autocorrelation). Within

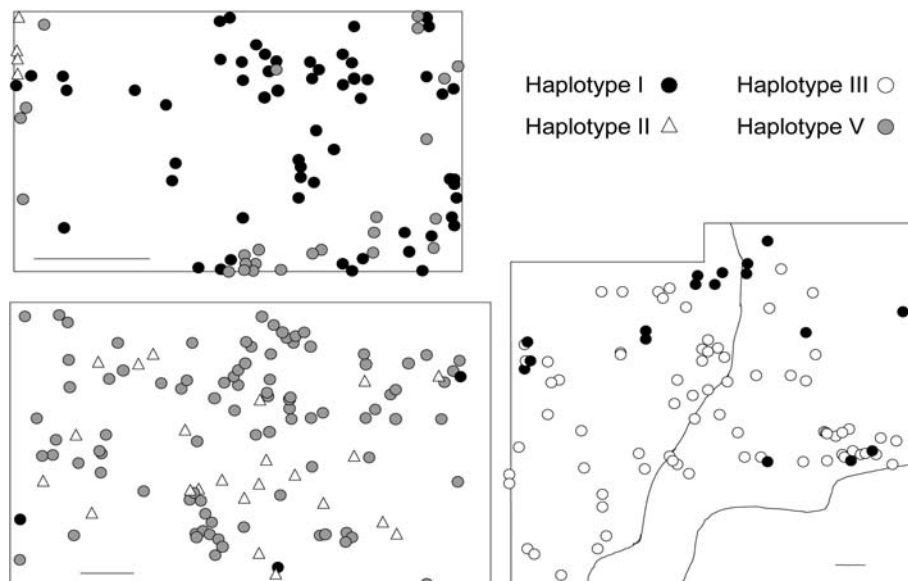


Figure 1. – Distribution of *Q. rubra* cpDNA haplotypes in PD_1, PD_4 and PMMF. Internal horizontal bar = 50 meters. The irregular line in PMMF distinguishes the central undisturbed old growth region from the peripheral old growth region.

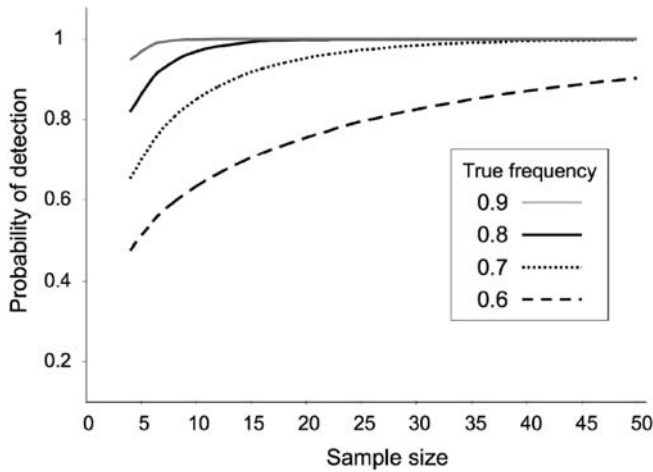


Figure 2. – Probability of correct identification of the predominant haplotype for sample sizes from 5 to 50 given the true predominant haplotype frequency.

a 10 meter distance, spatial autocorrelation at PD_4 was high (average Moran's $I = 0.59$). Not only do similar haplotypes group together but also 25 of the 29 trees having minority haplotypes grow within 20 m of the forest edge (Figure 1). At PD_1, spatial autocorrelation was lower but still positive (average Moran's I for the 10 meter distance = 0.23). PD_1 has only three forest edges, as the southern edge of the study site is an interior boundary. Only one of the trees having a minority haplotype lies within 20 m of the three edges. Although haplotypes V and II occur throughout the site, haplotype V occurs in distinct clusters and haplotype II is more evenly distributed. At PMMF, spatial autocorrelation is positive for both the 20 and 40 meter distances (average Moran's $I = 0.31$ for both). Of the 17 trees having haplotype I, 14 occur along the edges of the *Q. rubra* distribution at that site.

Transect performance

Transect performance was measured in two ways. The accuracy measure for haplotype richness is the proportion of all transects per site that detected only the true predominant haplotype, the true predominant haplotype and one other haplotype or all three haplotypes if three were actually present. For haplotype frequencies, frequencies for each transect were calculated, then the

range of frequencies across all transects reported for each haplotype. Measures were constructed in this way to evaluate the standard procedure of sampling along a single linear transect until the desired number of trees are sampled. All transects for both the five and 10 tree sample sizes correctly identified the predominant haplotype for all three sites (Table 5). Three of the eight transects for PD_4 failed to detect the second or third haplotype at $N = 5$. Two of the eight transects that found two haplotypes found the most and least frequent haplotypes, missing the second most frequent haplotype despite the relatively high actual frequency (0.24) of this haplotype. Only two transects detected all three haplotypes. At PD_1, three of the 15 transects failed to detect a second or third haplotype at $N = 5$ and one of 15 transects failed to detect a second or third haplotype at $N = 10$. At the PMMF, the largest fragment, 10 of the 25 transects failed to detect the second haplotype at $N = 5$ and 4 of the 25 transects failed to detect the second haplotype at $N = 10$. Haplotype frequencies based on a single transect were poor estimates of the true frequencies, ranging between 0.6 and one for the predominant haplotype and 0 to 0.4 for the second most frequent haplotype, when it was detected at all.

Discussion

DNA polymorphism within the CD, FV and TC regions

The significance of DNA sequence that lacks polymorphism is easy to overlook. The CD, FV and TC regions were chosen because these were the most polymorphic regions found in our initial study of old growth and older secondary growth stands in Indiana. Our PCR-RFLP approach tested 10239 bp of chloroplast sequence for seven different sets of specific nucleotide arrangements. In most of the previously published studies on chloroplast genetic diversity in the *Quercus* genus, more intergenic regions are tested but amplicons are restricted with a single enzyme. The SNP in the TC region was not detected in these three populations. A related investigation using the same markers employed in this study (unpublished data) indicates that the A to G transition is rare, occurring in only 5 of 703 trees genotyped to date. The lack of SNP and indel polymorphism within the CD, FV and TC intergenic regions for both the PCR-RFLP amplicons and the shorter CAPS markers sug-

Table 5. – Haplotype richness and frequencies inferred from linear transects.

Site	Transects	N ¹	Haplotype detection ²			Haplotype frequencies ³					
			1	1,2	1,2,3	Actual	Detected (range)			3rd	
						1st	2nd	3rd	1st	2nd	3rd
PD_4	8	5	3	3	2	0.66	0.29	0.05	0.6-1.0	0-0.4	0-0.2
		10	0	4	4				0.5-0.9	0.1-0.4	0-0.1
PD_1	15	5	3	8 ⁴	2	0.73	0.24	0.03	0.6-1.0	0-0.4	0-0.2
		10	1	9	5				0.5-0.8	0.2-0.4	0-0.1
PMMF	25	5	10	15	-	0.81	0.19	-	0.6-1.0	0-0.4	-
		10	4	21	-				0.7-1.0	0-0.3	-

¹ The number of trees sampled/transect.

² The number of times transects detected only the predominant haplotype (1), only the predominant and the secondmost frequent haplotype (1,2) or all three haplotypes (1,2,3) in the correct ranking.

³ 1st, 2nd and 3rd: The most frequent, secondmost frequent and least frequent haplotype.

⁴ Two transects, not included here, detected only the most and least frequent haplotype.

gests that sequence variation in these regions is limited both within and among these three populations. Assuming maximum parsimony, haplotype V, found at both PDRF sites, but not found in the PMMF, is ancestral to haplotype III, which is not found in either PD_1 or PD_4.

Biogeographical distribution of genetic diversity

The PDRF is located ~75 km north of the southernmost advance of Wisconsinian glaciation, whereas the PMMF is located ~132 km south of the glaciated region. However, absolute differences in haplotypes and the patterns of indel polymorphism indicate that the ancestors of the northern red oak population at PMMF were not the progenitors of the populations now present in the PDRF. Even if some of the haplotypes observed were the result of in-migration from acorns of trees in of nonlocal origin, stand histories suggest that the observed haplotype heterogeneities at all three sites predates European settlement.

Sample size and patch size

If the spatial distribution of trees with respect to chloroplast haplotypes is random, the probability of correctly identifying the predominant haplotype depends only on the sample size and the true frequency of the predominant haplotype. For sample sizes typically taken in previous studies of chloroplast haplotypes in oaks (5–10), the probability of correct identification exceeds 0.95 only if the frequency of the predominant haplotype exceeds 0.90. This study has shown that haplotype frequencies in excess of 0.90 may not be at all typical. Furthermore, if haplotypes are not randomly distributed, as shown in this study, small sample sizes (<10) taken along a linear transect at 50 m intervals will result in poor estimates of haplotype frequencies and can fail to detect haplotypes whose actual frequencies exceed 20% of the trees in the stand.

Patch size is directly affected by the maximum dispersal distance (d_{max}) and the frequency with which dispersals of this distance occur. The actual patch size is of critical importance. Simulations have shown that the patch size shrinks as d_{max} increases and as the frequency of d_{max} increases (DAVIES *et al.*, 2004). Patch size is also more sensitive to long distance seed dispersal than the lag between populations formed by a given long-distance dispersal event and the more slowly advancing front of the main population. Small patch sizes (< 1 km) result in higher values of G_{st} on a regional level. Patch sizes less than 1 km across are consistent with the haplotype distributions found in this study. The difference in haplotype frequencies between PD_4 and PD_1 could have resulted from the separation of local patches when the intervening forest was cut. Blue jays caching acorns along edges could have brought haplotypes II and V to PD_4. At PD_1, haplotype II is not limited to the edges of the fragment and is relatively frequent. The PD_1 site may span a very old patch boundary whose edges are now obscured through death and in-migration. In the simulation study, persistent small patch sizes were the result of astonishingly long d_{max} values (80–100 km). Refinement of simulation parameters (e.g., differential

reproductive success, shade suppression and seed predation) may allow for persistent small patch sizes with smaller values of d_{max} .

Effect of sampling strategy on population differentiation measures

G_{st} , the standard measures of population differentiation for uniparentally inherited haploid genomes, is acutely sensitive to sampling strategies when sample size is low. G_{st} depends on the number and frequency of haplotypes detected within and among populations. For the three forest fragments examined in this study, using the simple transect method of sampling only five trees every 50 m could have easily resulted in a data set that consisted of three haplotypes: Haplotype II for PD_4, haplotype V for PD_1 and haplotypes III for the PMMF, in which case $G_{st} = 1$. Complete population differentiation and biogeographical differentiation in the absence of geographical barriers (between the PDRF and PMMF sites lies the Tifton till plain, without mountains, major rivers or lakes) would be a surprising result in a population study involving a single species. Three sites are clearly too few for the application of such measures, but we would argue that five samples per site is also too low. Even as the number of populations rises, keeping the number of samples per site very low will miss the rarest haplotypes and will often miss haplotypes in the 20–30% range. Even under the random distribution assumption, a sample size of five has a correct identification probability of 95% or higher only if the true predominant haplotype has a frequency of 80% or higher. At a true frequency of 80%, the probabilities associated with all successes are 0.21 (3 of 5), 0.42 (4 of 5) and 0.32 (5 of 5). Even if the predominant haplotype was correctly identified, the true frequency will be poorly estimated in over half of the trials. Yet, in the calculation of G_{st} a great difference among haplotype frequencies between two sites has about as much weight as an absolute difference in haplotype richness and kind between two sites. If most contemporary, unmanaged stands of *Q. rubra* in the United States have two or more haplotypes present in the frequencies detected in this study, then range-wide estimates of G_{st} based very low sample sizes could misrepresent the degree of biogeographical differentiation present, especially if collections are made without regard to stand age or degree of anthropogenic disturbance and samples are taken along the edges of a stand.

We have detected chloroplast haplotype heterogeneity at each of three old growth sites, ranging in size from 2.83 to 20.64 hectares in size and located in two geographically distinct regions 207 km apart. All three sites have a predominant haplotype but the proportion of the secondmost frequent was ranged from 0.19 to 0.29, demonstrating that minority haplotypes are not necessarily rare. The location of the trees having the secondmost frequent haplotype, the stand ages and the stand histories all suggest that haplotype heterogeneity is a natural condition resulting from relatively small patch sizes. Even under the assumption of random distribution of haplotypes, if a site has a mixture of haplotypes, the probability of accurate identification of the predomi-

nant haplotype falls sharply as sample size decreases. Positive spatial autocorrelation weakens the value of any analytical approach based on the assumption of random distribution of haplotypes. Positive spatial autocorrelation results in poor estimates of both haplotype richness and haplotype frequency for the sample sizes typically taken (≤ 10) along a linear transect. If these three sites are representative for *Q. rubra*, measures of population differentiation in studies based on this sampling method must be interpreted with caution. While only three sites cannot provide strong support for biogeographical discontinuities in haplotype distribution, the data show that such discontinuities are possible.

In summary, we have shown that in three forest fragments with histories of minimal human disturbance during the European settlement period, haplotype richness is low but not zero within a single site, haplotype mixtures occur at all three sites and that haplotypes within sites exhibit spatial autocorrelation. Thus, small sample sizes (≤ 10) taken along a linear transect did not yield accurate estimates of haplotype frequency and haplotype richness. The degree of chloroplast haplotype richness and the spatial autocorrelation per site ultimately depends on species-specific fitness components, seed dispersal mechanisms, seed predation and other factors that merit serious investigation.

Comparative DNA sequencing of chloroplast regions within *Q. rubra* populations and sampling in areas that may have served as cryptic refugia will provide a more complete picture of the existing biogeography and genetic diversity of chloroplast haplotypes in northern red oak and ultimately will enable us to detect patterns of postglacial migration for this ecologically and economically important species.

References

- ALDRICH, P. R., G. R. PARKER, J. ROMERO-SEVERSON and C. H. MICHLER (2005): Confirmation of recruitment failure of red oaks in Indiana old-growth forest. *Forest Science* **51**: 406–416.
- BAKER, R. G., E. A. BETTIS III, R. F. DENNISTON, L. A. GONZALEZ, L. E. STRICKLAND *et al.* (2002): Holocene Paleoenvironments in Southeastern Minnesota – Chasing the Prairie-Forest Ecotone. *Palaeogeography, Palaeoclimatology, Palaeoecology* **177**: 103–122.
- CHESSER, R. K. and R. J. BAKER (1996): Effective Sizes and Dynamics of Uniparentally and Diparentally Inherited Genes. *Genetics* **144**: 1225–1235.
- DAVIES, S., A. WHITE and A. LOWE (2004): An investigation into effects of long-distance seed dispersal on organelle population genetic structure and colonization rate: a model analysis. *Heredity* **93**: 566–576.
- DEL COURT, P. A. and H. R. DEL COURT (1987): Long Term Forest Dynamics of the Temperate Zone: A Case Study of Late-Quaternary Forests in Eastern North America, pp. 246–252 in *Ecological Studies, Analysis and Synthesis*, edited by W. D. BILLINGS, F. GOLLEY, O. L. LANGE and J. S. OLSEN. Springer-Verlag, New York.
- FRALISH, J. S., F. B. CROOKS, J. L. CHAMBERS and F. M. HARTY (1991): Comparison of Presettlement, Second-growth and Old-growth Forest on Six Site Types in the Illinois Shawnee Hills. *American Midland Naturalist* **125**: 294.
- JACKSON, S. T., R. S. WEBB, K. H. ANDERSON, J. T. OVERPECK and T. WEBB *et al.* (2000): Vegetation and environment in Eastern North America during the Last Glacial Maximum. *Quaternary Science Reviews* **19**: 489–508.
- JOHNSON, W. C., C. S. ADKISSON, T. R. CROW and M. D. DIXON (1997): Nut caching by blue jays (*Cyanocitta cristata* L.): Implications for tree demography. *American Midland Naturalist* **138**: 357–370.
- JOHNSON, W. C. and T. WEBB (1989): The Role of Blue Jays (*Cyanocitta-Cristata* L.) in the Postglacial Dispersal of Fagaceous Trees in Eastern North-America. *Journal of Biogeography* **16**: 561–571.
- MAGNI, C. R., A. DUCOUSSO, H. CARON, R. J. PETIT and A. KREMER (2005): Chloroplast DNA variation of *Quercus rubra* L. in North America and comparison with other Fagaceae. *Molecular ecology* **14**: 513–524.
- PARKER, G. R., D. J. LEOPOLD and J. K. EICHENBERGER (1985): Tree Dynamics in an Old-Growth, Deciduous Forest. *Forest Ecology and Management* **11**: 31–57.
- PETIT, R. J., S. BREWER, S. BORDACS, K. BURG, R. CHEDDADI *et al.* (2002): Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence. *Forest ecology and management* **156**: 49–74.
- PONS, O., and R. J. PETIT (1995): Estimation, Variance and Optimal Sampling of Gene Diversity. 1. Haploid Locus. *Theoretical and Applied Genetics* **90**: 462–470.
- ROMERO-SEVERSON, J., P. ALDRICH, Y. FENG, W. L. SUN and C. MICHLER (2003): Chloroplast DNA variation of northern red oak (*Quercus rubra* L.) in Indiana. **26**: 43–49.
- SCHLARBAUM, S. E., R. P. ADAMS, W. T. BAGLEY and W. J. WAYNE (1982): Post-glacial migration pathways of *Quercus rubra* L., Northern Red Oaks, as indicated by regional genetic-variation patterns. *Silvae genetica* **31**: 150–158.
- SCHMELZ, D. V. and A. A. LINDSEY (1965): Size class structure of old growth forests in Indiana. *Forest Science* **11**: 258–264.
- SMALLWOOD, P. D., M. STEEL, A., E. RIBBENS and W. J. MCSHEA (1998): Detecting the effect of seed hoarders on the distribution of seedlings of tree species: Gray squirrels (*Sciurus carolinensis*) and oaks (*Quercus*) as a model system: In: pp. 211–222 in *Ecology and evolutionary biology of tree squirrels*. Virginia Museum of Natural History, Martinsville, Virginia.
- STEELE, M. A., G. TURNER, P. D. SMALLWOOD, J. O. WOLFF and J. RADILLO (2001): Cache management by small mammals: Experimental evidence for the significance of acorn-embryo excision. *Journal of mammalogy* **82**: 35–42.