

# Comparative Genetic Mapping in the *Quercus* Genus: A Prelude

---

Inna Birchenko<sup>1</sup>, Yi Feng<sup>2</sup>, Jeanne Romero-Severson<sup>1\*</sup>

<sup>1</sup>University of Notre Dame  
Department of Biological Sciences  
327 Galvin Life Sciences  
Notre Dame, IN 46556-5645

<sup>2</sup>Lawrence Technological University  
Department of Natural Sciences  
S222 Science Building  
Southfield, MI 48075-1058

Key words: *Quercus rubra*, chloroplast, genetic diversity,  
postglacial migration, genetic map

\*Corresponding author  
University of Notre Dame  
Department of Biological Sciences  
327 Galvin Life Sciences  
Notre Dame, IN 46556-5645  
Phone: 574 631 3938  
Fax: 574 6317413  
E-mail: [jromeros@nd.edu](mailto:jromeros@nd.edu)

## Introduction

Oaks, chestnuts and beeches are dominant angiosperm forest trees in the temperate forests of the northern hemisphere. All of the species in the red oak section of the *Quercus* genus (*Quercus* section *Lobatae*) are native to North and Central America. Among the many of species of red oaks occurring in eastern North America, northern red oak (*Quercus rubra* L.) has the largest range and is now the dominant upland species in oak-dominated ecosystems in the eastern United States. As a living tree, northern red oak provides food and shelter for wildlife and after harvest, a durable, beautiful hardwood for furniture, floors and trim.

Despite wide recognition of the ecological and economic importance of northern red oak, the sheer abundance of stems in forested, suburban and even urban landscapes in the eastern United States produces an impression that this species requires nothing from humanity but a sunny space in which to grow. In fact, northern red oak suffers from range-wide regeneration failure. Exotic pest and diseases (Gypsy moth, oak wilt and sudden oak death among others), fire suppression, short-sighted deer management practices, unsustainable harvesting and sincere but uninformed opposition to good management all contribute to this situation. Even if we were able to prevent or contain the insects and diseases and establish good management practices, our limited understanding of the contemporary genetic structure of naturally regenerated oaks could undermine our best intentions.

## The northern red oak genetics program at Notre Dame

A major research aim of the oak genetics and genomics program at Notre Dame is to develop a set of genetic and genomics tools that will enable us to detect

the biogeographical distribution of genetic diversity in northern red oak and other eastern red oaks. As almost all eukaryotic organisms contain at least two quite different genomes, the organelle genome and the nuclear genome, we are developing two different kinds of tools: one to detect genetic diversity in the chloroplast and mitochondrial genomes and another to detect genetic diversity in the nuclear genome. In oaks the chloroplast and mitochondrial genomes are inherited only from the female parent. Each tree has only one type of mitochondrion and one type of chloroplast and these will be identical to those of the female parent of that tree. In angiosperm plants, the chloroplast genome has high structural stability. Recombination has not been observed and rearrangements are rare. Mutations do occur but these tend to be harmless only in those places where the chloroplast DNA is not coding for a gene. These intergenic regions slowly accumulate neutral mutations, resulting in the appearance of new haplotypes. We use the polymerase chain reaction technique (PCR) and series of specialized enzymes that can detect changes in the DNA sequence in these regions. These enzymes restrict the DNA into precisely sized fragments that will differ in length for different DNA sequences. This combined approach reveals PCR restriction fragment length polymorphism (PCR-RFLP) among the chloroplast genomes of different oaks in the same species, in different species and even in different families.

In oaks, as in humans, male and female parents each contribute equally to the nuclear genome. The progeny will not be identical to either parent but will share the characteristics of both. To detect polymorphism in the nuclear genome we use microsatellite markers. The nuclear genomes of most eukaryotes have extensive noncoding regions that often accumulate long stretches of noncoding nucleotide repeats (e.g. GAGAGAGA). These areas in the genome accumulate mutations, most of which are harmless, at a rate faster than the DNA sequence in functional genes. We use this characteristic to develop DNA markers than can distinguish individuals within a population of the same species or among very closely related species. Most forensic testing of human DNA also employs microsatellites.

Almost all of the living tissues of a tree, including the embryos, contain proplastids, colorless organelles that will become green chloroplasts after seed germination. Matrilineal inheritance of the proplastid genome enables investigators to detect successful seed dispersal through genetic analysis of the chloroplast genome in individual 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. Thus the chloroplast genomes of the original migrants will tend to persist in local populations, permitting reconstruction of recolonization patterns (PETT *et al.* 2002). The long-lasting genetic impact of the first migrants is called a founder effect.

On an island, the genetic diversity of both the organelle genomes (the mitochondrial and the chloroplast genome) and the nuclear genome are profoundly affected by the founders of the population, even in outcrossing, wind-pollinated taxa. On a large land mass, only the organelle genomes will show a persistent founder effect and then, only if the descendants of the first migrants suppress establishment of late arrivals and remain at the colonizing front of the population. This limited type of founder effect is expected in terrestrial organisms in which the reproductive adult is sessile and long-lived and the propagules are not dispersed by wind or water. Long-lived forest trees with large seeds (e.g. oaks and chestnuts) are expected to show strong matrilineal founder effects.

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 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 suggested 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 pollen deposition may have been too thin for consistent detection if small populations 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 Rocky Mountains or 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 (Fig. 1).

Large-scale changes in rainfall patterns also induced Paleoclimate range shifts along longitudinal gradients. Pollen records indicate that during the Hypsithermal drying interval 8,700-5,000 yr BP, a tall grass prairie peninsula extended eastward to Indiana and southwest Michigan, partially isolating northern *Q. rubra* populations from southern *Q. rubra* populations (BAKER *et al.* 2002). The contemporary western edge of the *Q. rubra* range still 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.

Regardless of how many refugia there were or whether or not cryptic northern refugia played a major role, founder effects during postglacial migration are expected to result in a latitudinal gradient of chloroplast genetic diversity that declines in the direction of glacial retreat. In the northern hemisphere the northern edge of the contemporary range should have the lowest chloroplast genetic diversity if founder effects persist over time. Oak pollen recovered from lake bottom cores and other suitable sites indicates that during the last glacial maximum (LGM) oak populations were most dense in east Texas and northern Florida (JACKSON *et al.* 2000). If the Texas populations recolonized the western part of the contemporary range and the Florida populations the east, the haplotypes found on the western northern edge may differ from the haplotypes found on the eastern northern edge. The descendants of Texas migrants that colonized Kansas, Missouri, Iowa and Minnesota may have been eliminated or pushed eastward during the Hypsithermal event (Figure 1). To further complicate matters, small populations in cryptic refugia in the Appalachians may have had an advantage over southern populations. Ahead lay some of the richest and flattest land in the world, ideal for forest establishment under mesic conditions. If there were many cryptic refugia distributed across the Appalachians, and many of these contained trees with unique haplotypes, then

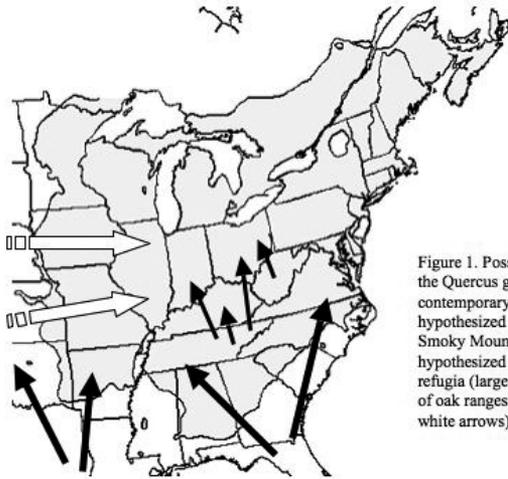


Figure 1. Possible routes of postglacial migration for the *Quercus* genus in eastern North America. The contemporary range of northern red oak (light gray) hypothesized paths out of Blue Ridge and Great Smoky Mountain refugia (small black arrows), hypothesized paths out of the Texas and Florida refugia (large black arrows) and eastward shrinkage of oak ranges during the Hypsithermal event (large white arrows).

the center of genetic diversity for chloroplast haplotypes would be located in this region and not further south.

### Research results and discussion

To test the hypothesis of a latitudinal gradient (an effect expected regardless of the number or locations of refugia) we compared chloroplast genetic diversity among 29 populations of *Quercus rubra* in old-growth and minimally disturbed forests in eastern North America. Our sampling approach included collecting at least 30 trees per site. Many studies of chloroplast diversity assume beforehand that only one haplotype will be present in a given population and therefore, large sample sizes are unnecessary. Our studies of forest fragments in which we genotyped all of the *Quercus rubra* in the stand (Feng et al. submitted) suggest that haplotype heterogeneity (two or more haplotypes at frequencies greater than 5% of the population) is a natural occurrence that will be missed when sample sizes are very small (<10). Our aims in this study were to answer these three questions: 1) does *Q. rubra* cpDNA diversity decline across a latitudinal gradient 2) are populations at the northern edge of range fixed for a single haplotype and 3) can the most recent postglacial recolonization routes be reconstructed using the contemporary biogeographical distribution of chloroplast haplotypes?

In four years of genotyping, we have found only the five haplotypes we encountered in our preliminary study (Romero-Severson et al., 2003). Haplotypes I and II were the most common. In the eight sites north of 44° N, we detected only haplotypes I and II and six out of eight sites had only one haplotype (Figure 2). The 59 trees sampled from Isle Royale (47.55° N 89.02° W), the largest island in Lake Superior and close to the northern range limit for northern red oak, were monomorphic for haplotype II. Haplotype III, locally abundant in the Pioneer Mothers Memorial Forest (Feng et al. submitted), and detected at a site near the Pioneer Mothers Memorial Forest in a previous study (Romero-Severson et al. 2003), occurred in only one other site, the Indiana Dunes, where it is also locally

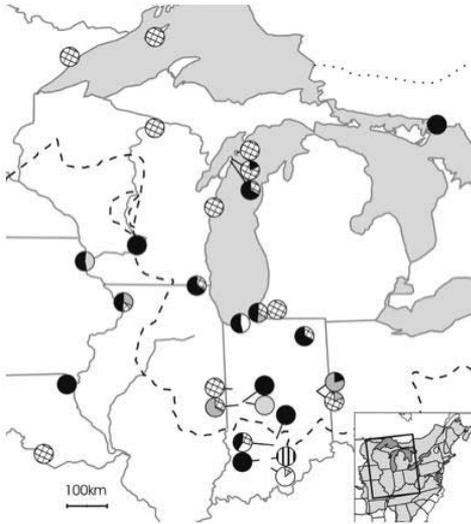


Figure 2. Biogeographical distribution of chloroplast haplotypes detected in *Q. rubra* from 2002-2006. Haplotype I (black), Haplotype II (cross-hatched), Haplotype III (white), Haplotype IV (vertical bars) and Haplotype V (light gray). The LGM is indicated by the dashed line. Inset shows the contemporary native range of *Q. rubra*.

abundant. Previous studies detected haplotype IV at a site south of glaciation line (Romero-Severson *et al.* 2003), but we have not found haplotype IV at any other site. Haplotype V occurred at five Indiana sites north of the glaciation line and two sites west of the glaciation line. Although haplotype V is locally abundant at four of these sites it was not found north of 43° N or south of 40° N.

Haplotype III has a clear latitudinal gradient, from southern to northern Indiana, but haplotype V occupies the center of the sampled area. The maternal ancestors of these trees could have originated from Texas and successfully colonized the western part of range while the maternal ancestors of haplotype III trees originated from Appalachian refugia and colonized north, across the open plain of central Indiana. The pattern of insertions and deletions in the genotyped part of the chloroplast genome suggests that haplotype V predates the others, while haplotype III is most recent. Tests of these hypotheses require more sampling in regions west and south of the LGM.

Repeated founder effects and adverse interactions with pests, diseases and aggressive, shade-tolerant species as recolonization proceeded northward could also have resulted in loss of genetic diversity in the oak nuclear genome. Loss of genetic diversity in the nuclear genome depends on the distribution of reproductively competent female parents and the effective size of the pollination neighborhood. As the pollen of all *Quercus* species is small (~x) and smooth, long distance transport by wind could result in pollination neighborhoods many kilometers in diameter, mitigating both the founder effect and severe anthropogenic disturbance arising from land clearing, timber harvest and urban sprawl. However, previous studies (SMOUSE *et al.* 2001; SORK *et al.* 2002) have shown that oak pollination neighborhoods are remarkably small (tens of meters), although long-distance pollinations can occur (DODD and AFZAL-RAFI 2004).

Another complication arising in genetic studies of the nuclear genome is the infamous tendency of oaks to hybridize with one another. Based on whole tree silvicultural characters, northern red oak can hybridize with nearly a dozen other red oaks (JENSEN 1995). The oaks within *Quercus* section *Quercus* also hybridize with one another. This classic characteristic of the *Quercus* genus provides us with an excellent opportunity to examine the genetic basis of “species typical characters” and elucidate how taxon distinctness is maintained when barriers to gene flow among taxa appear minimal or absent. Perhaps the Lobatae and *Quercus* sections of the *Quercus* genes are phylogenetically young and not enough time has passed for barriers to arise among the members of both sections. Alternatively, the sections are phylogenetically old but multiple glaciation events during the Pleistocene may have resulted in species assemblages not normally found together. In these circumstances, where some species are duress, hybrids that would ordinarily be noncompetitive may be superior. Under the strictest definition of the biological species concept (no viable and fertile hybrids permitted), the Lobatae in the eastern United States would lump into a handful of species at most. Fortunately, these hypotheses are now testable.

We have now developed 34 microsatellite markers for *Quercus rubra* (ALDRICH *et al.* 2002; ALDRICH *et al.* 2003a). These microsatellite markers are highly polymorphic not only in *Q. rubra*, but also in *Q. shumardii* and *Q. palustris*. In one of our studies we used 15 microsatellites to compare the genetic relatedness of these three species with taxonomic designations (ALDRICH *et al.* 2003b). In the 20.6 hectare old growth forest fragment we chose for this study, mature trees numbered (>10 cm DBH) numbered 620 (*Q. rubra*), 108 (*Q. shumardii*) and 101 (*Q. palustris*). Although the trees did group into distinct genetic clusters, these clusters did not correspond to the species identification, indicating that whole tree silvicultural characters may not reveal the true extent of hybridization (Figure 3).

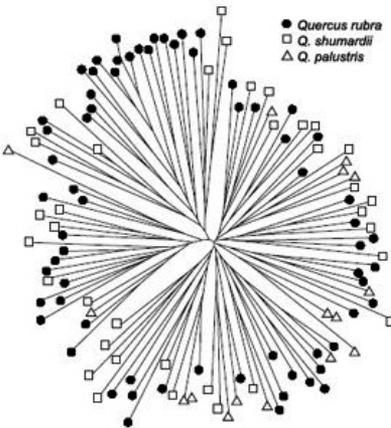


Figure 3. Genetic similarity clusters. Trees on a common stem are more closely related than trees on other stems. Groups of related clusters trace back to a single stem. There are three major groups and many subgroups. Note that the taxonomic classifications do not correspond to the genetic relatedness clusters.

The development of a suite of nuclear DNA markers is the first step in detecting the order of genes along the chromosomes. As genes tend to occur in the same order within the same species, gene order can be used as a taxonomic character. The relative degree of polymorphism across genes can reveal regions where genes that confer fitness lie. When natural selection for a given trait is strong (e.g. resistance to an endemic disease), the genes that confer this trait tend to become less and less polymorphic as selection weeds out the weaker alleles. The genes that happened to be linked to the gene carrying the best alleles will also tend to become monomorphic if selection is severe. This “hitchhiker” effect results in a significant drop in polymorphism in the region where pest resistance genes are located. A genetic map is also required to properly interpret the meaning of associations among alleles of different genes. If two genes are located very close together (tightly physically linked), then, because genetic recombination will occur between them only rarely, the alleles of both genes tend to be inherited together. If, on the other hand, the alleles of two genes show strong association and we know that they are not physically linked (linkage disequilibrium without physical linkage), then we have a good circumstantial case for active selection of these specific allele combinations.

The classic approach to making a genetic map requires that we have two known parents and progeny from only those parents. Other approaches are possible but the classic design has the most power for a given progeny population size. Most of the Lobatae, including northern red oak, take two years to produce a given acorn. This makes structured cross problematic. However, a mature oak can produce thousands of acorns every year and the data on other oaks suggest that the pollination neighborhood is small. Therefore, there might be hundreds of acorns on a tree that were pollinated by a single male. We have encountered disbelief from nearly everyone (including funding agencies) by single-mindedly pursuing this idea, and have found that our assumptions were correct. We now have in hand a full sib population of nearly three hundred seedlings and are actively screening and mapping our own microsatellites and other types of genetic markers as they become available.

This work is merely a prelude to the effort that is to come. Our goal is to establish an integrated genetic and genomics program for our native oaks that has the enthusiastic support of applied foresters, silviculturists, forest tree geneticists, taxonomists, population geneticists, private foundations, public funding agencies and dedicated ordinary people who love our native oaks and do not want our children and grandchildren to witness their loss and wonder why those who could did not do more to save them.

### References

- Aldrich, P., C. Michler, W. Sun and J. Romero-Severson, 2002 Microsatellite markers for northern red oak (Fagaceae: *Quercus rubra*). *Molecular ecology notes* **2**: 472-474.
- Aldrich, P. R., M. Jagtap, C. Michler and J. Romero-Severson, 2003a Amplification of North American red oak microsatellite markers in European white oaks and Chinese chestnut. *Silvae genetica* **52**: 176-179.
- Aldrich, P. R., G. R. Parker, C. Michler and J. Romero-Severson, 2003b Whole-tree silvic identifications and the microsatellite genetic structure of a red oak species complex in an Indiana old-growth forest. *Canadian journal of forest research* **33**: 2228-2237.

- Baker, R. G., E. A. B. 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.
- Delcourt, P. A., and H. R. Delcourt, 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.
- DODD, R. S., and Z. AFZAL-RAFI, 2004 Selection and dispersal in a multispecies oak hybrid zone. **58**: 261-269.
- Jackson, S. T., R. S. Webb, K. H. Anderson, J. T. Overpeck, T. Webb *et al.*, 2000 Vegetation and environment in Eastern North America during the Last Glacial Maximum. *Quaternary Science Reviews* **19**: 489-508.
- Jensen, R. J., 1995 Identifying Oaks: The Hybrid Problem. *Journal of the International Oak Society*: 47-54.
- 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.
- 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.
- Smouse, P. E., R. J. Dyer, R. D. Westfall and V. L. Sork, 2001 Two-generation analysis of pollen flow across a landscape. I. Male gamete heterogeneity among females. **55**: 260-271.
- Sork, V. L., F. W. Davis, P. E. Smouse, V. J. Apsit, R. J. DYER *et al.*, 2002 Pollen movement in declining populations of California Valley oak, *Quercus lobata*: where have all the fathers gone? **11**: 1657-1668.