Seed Dispersal, Gene Flow, and Hybridization in Red Oak

by

Emily V. Moran

Department of Biology
Duke University

Date:________________________
Approved:

____
James S. Clark, Supervisor

____
John Willis

____
William Morris

____
Paul Manos

____
John Terborgh

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor
of Philosophy in the Department of
Biology in the Graduate School
of Duke University

2010

ABSTRACT

Seed Dispersal, Gene Flow, and Hybridization in Red Oak

by

Emily V. Moran

Department of Biology
Duke University

Date:_______________________
Approved:

____
James S. Clark, Supervisor

____
John Willis

____
William Morris

____
Paul Manos

____
John Terborgh

An abstract of a dissertation submitted in partial
fulfillment of the requirements for the degree
of Doctor of Philosophy in the Department of
Biology in the Graduate School
of Duke University

2010

Copyright by
Emily V. Moran
2010

Abstract

Understanding the ecological and evolutionary responses of plant species to shifts in climate (and other rapid environmental perturbations) will require an improved knowledge of interactions between ecological and evolutionary processes as mediated by reproduction and gene flow. This dissertation research examines the processes of seed dispersal, intra- and inter-specific gene flow, and reproductive success in two red oak populations in North Carolina; the variation in these processes from site to site; and their influence on genetic structure, population dynamics, and migration potential.

Using genetic and ecological data collected from two large long-term study sites, I develop a hierarchical Bayesian model to identify the parents of sampled seedlings and characterize the scale of effective seed and pollen dispersal. I examine differences in scale of dispersal between the Appalachian and Piedmont sites in light of the spatial genetic structure and ecological differences of the two sites. I then use the pedigree and dispersal estimates derived from these analyses to examine variation in reproductive success and to test hypotheses about the causes and consequences of such variation. Using parentage estimates and measures of genetic differentiation between species, I study the likely extent of hybridization in these mixed-species secondary forests. Finally, using the SLIP stand simulator, I explore the implications of new genetic dispersal estimates for migration potential in oaks.

I find that effective seed dispersal distances are longer than estimated using seed trap data. While at the Piedmont site the large number of seedling found >100 m from their mother trees suggests that animal dispersers play a vital role, at the Appalachian site seedling distributions conform more closely to the original gravity-created pattern of seed density. Individual trees vary widely in their reproductive success. Seedling production was found to be positively associated with annual seed production, but exhibited hump-shaped or reversing relationships with age (suggesting the effect of senescence) and growth rate (suggesting tradeoffs in allocation). Germination fraction was negatively associated with fecundity, suggesting that density-dependent mortality may be acting on the high concentrations of seeds near highly fecund adults. Due to overlapping generations and variation in individual reproductive success, effective population size is estimated to be less than half the size that numbers of “adult” individuals would suggest, with consequences for the relative strength of drift and selection. Hybridization may boost effective population size somewhat; my analyses suggest that inter-specific gene flow is common at both study sites. Finally, simulations show that dispersal has a relatively stronger effect on migration rate and population growth than fecundity or size at maturity, and that genetic estimates of seed dispersal can yield significantly higher rates of migration and/or population persistence than seed-trap based estimates under both competitive and non-competitive conditions.

Dedication

This dissertation is dedicated to my grandmother, Caridad Moran (1914-2010), who was a living example of perseverance and love; and to Emilio Moran and Millicent Fleming-Moran, who are not only wonderful, supportive parents, but also my first role models in science.

Contents

List of Tables

List of Figures

Acknowledgements

I would like to acknowledge the guidance and assistance provided by my advisor, Jim Clark, and my other committee members – with particular thanks to John Willis for letting me use his lab for my genetic analyses. I would also like to thank my friends and family for patiently reading multiple manuscript drafts. Funding for this project was provided by the Duke Biology Department, the Center for Theoretical and Mathematical Sciences, and the Duke Graduate School, as well as by the Association for Women in Science. Funding for long-term research at Duke Forest and Coweeta LTER was provided by NSF.

Chapter 1. Introduction

Why should ecologists care about gene flow?

Genes and ecology

In recent years, there has been increasing interest in the interplay between ecology and evolution, as reflected in the development of new specialties such as community genetics and ecological genomics. The importance of ecological interactions for evolutionary dynamics and vice versa has been recognized since Darwin (1859) and his reflections on a "tangled bank" (an assemblage of interacting plants, insects, birds, etc.), although for practical reasons they have often been treated as separate subjects (Johnson and Stinchcombe 2007). However, new methods are beginning to yield powerful insights into the genetic underpinnings of variation in ecological characteristics, and into the influence of ecological interactions on the distribution of genetic variation, promising to bring ecology and evolution closer than ever.

The term "gene flow" refers to the transport of genes across a landscape due to the movement of individuals or gametes within and between populations. Patterns of gene flow across the landscape have important implications for population dynamics, the maintenance of genetic diversity, and the effectiveness of natural selection (Kawecki 2008). When dispersal and gene flow are limited, genetic diversity in small or isolated populations can quickly be depleted due to strong selection, genetic drift , or a combination of the two; in large populations, these processes are usually balanced by mutation (Gillespie 2004). Reduced genetic diversity can impair a population's viability and ability to respond to environmental change (Newman and Pilson 1997, Robert et al. 2007), especially in self-incompatible species where the loss of S alleles can severely reduce the number of potential mates (Sork et al. 2002). Immigration can boost genetic diversity and adaptive potential (Kimbrell and Holt 2007), but local adaptation at range limits and in marginal habitats can also be inhibited when the influx of maladapted genes from the main part of the species range exceeds the rate at which they are purged by selection (Rehfeldt et al. 1999, Lenormand 2002, Lopez et al. 2007). These processes are of particular interest today, given the selective pressure of anthropogenic global change on ecosystems around the world.

Climate change and biodiversity

Climate change has joined habitat loss, hunting, invasive species, and pollution as one of the major threats to biodiversity (Thomas et al. 2006), and is expected to become an increasingly important force over the coming century. Global mean temperatures are predicted to increase 2.4 to 6.4°C by 2100, accompanied by shifts in the amount and distribution of precipitation (IPCC 2007). Climate change can act as a major selective pressure, directly or indirectly affecting every species on earth (Holt 1990, Skelly et al. 2007). In the past, organisms have responded to rapid shifts in climate, such as the warming at the end of the last ice age, through a combination of range shifts and adaptive evolution (Davis and Shaw 2001, Petit et al. 2002). Species that were unable to respond quickly enough, or that were limited in their response by geographic barriers, went extinct or were reduced to small relict populations (Willis and Niklas 2004). Given the rapid rate of current anthropogenic climate change (Overpeck et al. 1991, IPCC 2007), concerns have been raised about the ability of many species to keep pace (Dyer 1995, Pitelka 1997). Indeed, some have predicted climate-change-related extinctions of 30% or more within several major taxa of plants and animals, primarily due to severe reductions in potential or realized species range (Hannah et al. 2005, Thuiller et al. 2005).

For non-mobile organisms with long generation times such as woody plants, responding to rapid environmental change presents additional challenges. The movement of pollen and seed are the only source of gene flow within and between populations, and thus crucially affect standing genetic diversity and the potential for local adaptation, while seed movement alone allows for range expansion and the colonization of new sites. Both processes can be challenging to measure, particularly in forest trees (Clark et al. 2004).

The extent to which dispersal may constrain plant responses to climate change is a matter of considerable debate. Pollen records suggest rapid rates of spread for tree species at the end of the last ice age (Davis 1969). A “fat-tailed” pattern of seed dispersal, with more long-distance and fewer mid-range events than a Gaussian or exponential distribution, could help account for such rapid spread, because it allows a population to move by leaps, establishing colonists far in advance of the main population front (Clark 1998). Both genetic and ecological data indicate that fat-tailed dispersal kernels provide the best fit to seed and pollen dispersal in many tree species (Clark et al. 1999, Streiff et al. 1999, Hardesty et al. 2006), and invasive tree species frequently exhibit a pattern in which long-range colonists and their offspring provide a base for further population expansion (Petit et al. 2004). On the other hand, genetic data indicating the existence of northern refugia suggest that many species might not have had to move as quickly as was once though (McLachlan et al. 2005). Moreover, the rate of current anthropogenic climate change is substantially higher than the warming that followed the last ice age (Huntley 1991, Overpeck et al. 1991), while habitat fragmentation (Malanson and Cairns 1997, Collingham and Huntley 2000), loss of dispersers (Terborgh et al. 2008), and other environmental modifications may present a further barrier to migration.

Likewise, the extent of gene flow via seed and pollen can have a strong influence on the potential for adaptation to local climatic conditions. Many North American tree species exhibit a high degree of local adaptation despite having undergone large range shifts in the past 15,000 years (Davis and Shaw 2001), indicating that the processes of migration and adaptation may occur simultaneously during periods of rapid climate change. In general, the shorter the distance of gene movement, the finer the scale of local adaptation that is possible (Levin 1981). In wind-dispersed species, high levels of pollen immigration from core populations can inhibit local adaptation in peripheral populations (Kirkpatrick and Barton 1997, Rehfeldt et al. 1999); however, strong latitudinal clines in allele frequencies of climate-related genes can be observed in some tree species even where neutral genetic markers indicate high levels of pollen flow (Mimura and Aitken 2010). As with migration rate, the maintenance of genetic diversity and the potential for local adaptation may be altered by habitat fragmentation or changes in population size and biotic interactions. For instance, even wind-pollinated species such as oaks can exhibit pollen limitation or reductions in genetic diversity when populations are highly fragmented (Fernandez and Sork 2007) or when low reproductive success reduces population density (Sork et al. 2002).

Better estimates of the dispersal ability of forest trees, and the scale of gene flow via seed and pollen, are needed if we are to accurately gauge the sensitivity of these species to rapid environmental change, and to manage them accordingly.

Estimating dispersal – challenges and new approaches

Oaks, hickories, walnuts, and other tree species with heavy rodent-dispersed seeds are generally regarded as being highly dispersal-limited (Sork 1984, Clark et al. 2004, Garcia and Houle 2005). However, because the seed-trap data frequently used to characterize seed dispersal kernels misses secondary dispersal by animals, which for oaks can include birds as well as rodents, their capacity for long-distance dispersal may have been substantially underestimated. Attempts have been made to capture this secondary phase by tagging seeds with metal, magnets, or radioactive tracers (Sork 1984, Roth and Vander Wall 2005), but even when tags do not affect disperser behavior, seeds taken outside the searched area may be missed, while incoming seeds will not be detected. Direct observations of seed-caching behavior can also be used (Johnson et al. 1997, Gomez 2003), but movements longer than a few tens-of-meters can be difficult to follow in all but the most open forests, while deceptive caching behaviors, intended to baffle conspecific thieves (Clayton et al. 2007), could also confound measurements of dispersal distance.

An alternative approach is to use pedigrees to track the movement of seed and pollen. Parentage information is highly informative of dispersal in plants, because movement is restricted to only two stages: seed and pollen. While partial pedigrees for some animals can be obtained through long-term field observations (Pemberton 2008) this approach is impractical for trees, where mating and dispersal are cryptic and individuals may live (and reproduce) for centuries. In such cases, highly variable neutral genetic markers, such as microsatellites, can be used to reconstruct parentage, sibship, or population of origin (Dow and Ashley 1996, Aldrich et al. 1998, Streiff et al. 1999, Godoy and Jordano 2001, Asuka et al. 2005, Bacles et al. 2006, Hardesty et al. 2006, Pairon et al. 2006, Selkoe and Toonen 2006).

The use of molecular markers in parentage and dispersal studies presents challenges of its own. Early analyses used simple exclusion, eliminating as potential parents any adults that, at a given locus, did not share an allele with the juvenile (eg. Dow, Ashley, 1996). This approach can lead to the rejection of true relationships, as the probability of genotyping error for the most useful classes of genetic markers is not negligible

(Dewoody et al. 2006). More sophisticated models of parentage have since been developed (Marshall et al. 1998, Gonzalez-Martinez et al. 2006), but most still do not adequately account for genotyping error or the strongly distant-dependent nature of mating and offspring movement in plants Hadfield et al. (2006) proposed that, because most parentage studies aim to estimate population-level parameters (such as dispersal, self-pollination rates, or extra-pair mating frequencies) rather than simply generate a pedigree, the most sensible approach was to estimate these parameters simultaneously within a Bayesian framework. Such a framework can incorporate not only genetic data, but also ecological information such as social status or territory location. Using data for birds, Hadfield et al.(2006) found that this joint estimation increases the power of pedigree estimates and decreases bias in population parameters when the assumptions of the model are met. In this thesis research I use a similar approach, as described below (and in **Chapter 2**), to estimate effective dispersal distances of seed and pollen for two mixed-species populations of red oak in North Carolina.

Oaks – ecology and conservation

Oaks (*Quercus* spp.) are a major component of many North American deciduous forests. Many species are also important sources of hard mast for wildlife and of high-quality timber (Little 1980, McShea et al. 2006). But despite their dominance in the canopy, recent oak recruitment has been low in many areas, with few new individuals entering the larger size classes (Buckley et al. 1998, McDonald et al. 2002, Aldrich et al. 2005a). This decline in recruitment is most likely due to decreased fire frequency, which increases competition from species such as red maple (*Acer rubrum*), though other factors such as increased deer herbivory may also play a role (Abrams 1992, Elliott et al. 1999, McDonald et al. 2002, Spetich 2004).

Acorns are heavy, dispersed primarily by animals, and, compared to some tree seeds, produced in relatively low numbers (Clark et al. 2004). Seed limitation due either to low fecundity or limited seed movement, might contribute to declines in oak density by reducing the probability of reaching scarce recruitment sites. It could also result in low potential migration rates and low between-population gene flow, which together with low reproductive success would make oak populations more vulnerable to climate change. On the other hand, acorn cacheing birds could provide a mechanism of long distance seed dispersal. The most effective avian dispersers of acorns are the jays, which in North America include the blue jay (*Cyanocitta cristata*) in the eastern deciduous forest, and the scrub jay (*Aphelocoma coerulens*) in Florida and parts of the southwest (DeGange et al. 1989, Johnson and Webb 1989, Vander Wall 2001). Unlike jays, which bury acorns in shallow caches, acorn woodpeckers (*Melanerpes formicivorus*) store acorns by hammering them into tree-trunks (Grivet et al. 2005) and so are probably much less effective as dispersers.

Wind dispersal of pollen could compensate for low seed movement in the maintenance of genetic diversity. Many oaks exhibit low inter-population genetic differentiation, suggestive of high gene flow (Craft and Ashley 2007), although paternity analyses in some oak populations suggest that effective pollination neighborhoods can be quite small (Sork et al. 2002, Fernandez et al. 2006) and, where populations are small or fragmented, oaks can exhibit reductions in fecundity due to pollen limitation (Knapp et al. 2001).

Genetic diversity and adaptive potential in oaks could also be increased by hybridization. Oaks have a relatively high capacity to hybridize (Burger 1975, Cottam et al. 1982, Whittemore and Schaal 1991), and introgression from species with, for example, higher drought or heat tolerance, might provide useful material for adaptive responses to climate change. On the other hand, if hybrids are less fit than parental types, cross-species matings would only lower the reproductive success of the trees involved. The importance of hybridization in the field, and the conditions under which it is more or less likely to occur, is still a matter of lively debate (Rushton 1993, Craft et al. 2002, Coyne and Orr 2004, Dodd and Afzal-Rafii 2004, Muir and Schlotterer 2005).

Dissertation overview

I use red oaks (*Q. rubra*, *Q. velutina*, *Q. falcata*, and *Q. coccinea*) growing in mixed stands at two sites in North Carolina (Duke Forest in the Piedmont and Coweeta LTER in the Southern Apppalachians) as a system in which to explore patterns of seed and pollen dispersal and the consequences of these patterns for migration ability and the distribution of genetic diversity. In **Chapter 2**, I develop a hierarchical Bayesian model which combines genetic and ecological data to estimate effective seed and pollen dispersal and the parentage of seedlings. This model explicitly considers two sources of genotyping error, and is the first such model to treat dispersal from inside and outside the sampled plot in a consistent framework. I illustrate this method using data from the Piedmont site.

I extend this analysis in **Chapter 3**, where I compare the results for Piedmont and Appalachian sites. While both sites exhibit very long-range pollen dispersal, effective seed dispersal distances at the Appalachian site are significantly shorter than for the Piedmont site. I discuss these differences in the context of disperser and herbivore densities. Additionally, I examine patterns of spatial genetic structure at both sites, and ask whether these patterns are consistent with known site history and estimated dispersal distances of seed and pollen.

In **Chapter 4**, I examine variation in reproductive success between individual trees. This is one of the first studies to use pedigree information to estimate female reproductive success (defined as seedling production) in trees. I test several hypotheses about factors that may contribute to variation in reproductive success (such as tree age, fecundity, or proximity to gaps), and explore the consequences of this variation for effective population size and the relative strength of drift versus selection.

In **Chapter 5**, I explore the extent of gene flow between species at these two sites. Two measures of genetic structure, Fst and STRUCTURE 2.2 analysis, show insignificant genetic differentiation between co-occuring species at each site despite the existence of distinct morphologies, suggesting that all red oaks within a site should be treated as a single interbreeding population. I also extend the analyses of Chapters 2 and 3 (for which all oaks are indeed treated as potential parents and offspring) to derive estimates of hybridization for each site, and discuss future research needs.

In **Chapter 6**, I examine the implications of new genetic dispersal estimates for migration potential in forest trees. Using an individual-based forest stand simulator, I test the sensitivity of migration rate to changes in dispersal distance relative to changes in other demographic parameters, and compare the rate of spread based on genetic dispersal estimates (with and without heterospecific competitors) to projected rates of climate change.

Finally, in **Chapter 7**, I conclude with a summary of my findings and suggest directions for future research. While ‘dispersal’ and ‘gene flow’ are often used interchangeably in the ecological literature, it is important to recognize that not all dispersal results in gene flow while, conversely, effective gene flow distances may in some case be considerably longer than observed dispersal distances due to inbreeding depression or lack of detection of long distance dispersal events (Levin 1981). Furthermore, prediction of population responses to environmental change is often complicated by demographic stochasticity and the interactive effects of competitors, mutualists, and natural enemies. Predicting and managing the ecological and evolutionary responses of plant species to current anthropogenic climate change will require a better understanding of the scales of dispersal and gene flow in natural populations, and of the interactions between dispersal and the various biotic and abiotic factors which may limit the success of immigrant seed and pollen.

Materials and methods

The study populations

Because of the time and cost involved, most gene flow studies are based on a single population (Burczyk et al. 2004). However, gene flow needs to be considered in a broader context, especially in widespread species that experience a range of different environments. Comparison of two or more populations in different regions can begin to give some idea of the potential variation in the scale of seed and pollen dispersal.

This study makes use of two long-term research plots, one in the Blackwood Division of the Duke Forest in the Piedmont (35°58'N, 79°5'W, elevation 155m) and one at the Coweeta LTER site in the Southern Appalachians (35º03' N, 83 º27' W, elevation 1030 m). In order to regularize plot shape and expand the number of potential parents, all oaks within a 30-60 m border area were mapped and genotyped, bringing the total plot area to 12 ha and 7.5 ha, respectively. Both plots are located within >80 year-old secondary forest. Seed-trap, seedling-plot, adult demographic, and environmental data have been collected since 1998. The tree community in the Duke Forest plot includes mature loblolly pines (*Pinus taeda*) inter-mixed with hardwoods, including *Quercus*, *Carya*, *Acer*, and *Liquidambar*. It was largely cleared for farming, and abandoned in the early 20th century. The Coweeta site is dominated by mixed hardwoods including *Quercus*, *Acer*, and *Liriodendron*; *Tsuga* (currently succumbing to woolly adelgid); and *Rhododendron,* and is known to have been selective logged in the early 1900's. The Appalachian site is more topographically complex than the Piedmont site, and exhibits a much higher density of seedlings of all species (1.27 vs. 0.41 red oak seedlings/m^2), most likely due to lower deer herbivory. Deer densities at Duke Forest (unpublished data) are extremely high: 55 deer/mi^2 in 2005, and 41/mi^2 in 2010 following two years of hunting, compared to a recommended density of 15-20/mi^2 (Marissa Hartzler, pers. comm.). Estimates of deer densities for Coweeta are not available, but according to the NC Wildlife Resources Commission deer densities in surrounding Macon county are approximately 8.5/mi^2 (Mike Carraway, pers. comm.). Site history and disperser density are further discussed in **Chapter 3**, and descriptions of the study sites are available at http://www.env.duke.edu/forest and http://coweeta.ecology.uga.edu.

The focal species

Most North American oaks can be divided into the red oak (section *Lobatae*) and white oak (section *Quercus*) clades. Within these larger taxonomic groups, oak species are believed to have a high ability to hybridize (Burger 1975, Aldrich et al. 2003b, Dodd and Afzal-Rafii 2004). For this reason, I include in this study not only northern red oak (*Q. rubra*), which has published microsatellite primers (Aldrich et al. 2002, Aldrich et al.

2003a) and is particularly abundant at both study sites, but also species present at each study site likely to hybridize with it: black oak (*Q. velutina*) at both sites, southern red oak (*Q. falcata*) at Duke Forest, and scarlet oak (*Q. coccinea*) at Coweeta. Microsatellite primers were found to transfer effectively to all three species. Trees were identified to species based on morphology (Little 1980); the extent of genetic differentiation and hybridization is examined in **Chapter 5**.

All four oak species are broadly distributed in eastern North America (Little 1980). *Quercus rubra* has the widest range, extending from Ontario to Alabama. Southern red oak (*Q. falcata*), as the name suggests, has a more southerly distribution, from southern Indiana and New Jersey to northern Florida. *Quercus velutina* does not grow as far north as *Q. rubra*, extending only to mid-Michigan, but is otherwise similarly distributed. *Quercus coccinea* has perhaps the most restricted distribution, being concentrated around the Appalachian Mountains from SW Maine to Georgia. *Quercus falcata*, *Q. velutina*, and *Q. coccinea* all tolerate drier soils than *Q. rubra*. Nevertheless, species present at each site are frequently found growing side-by-side.

Acorn production in red oaks is variable and episodic, with large crops occurring every 3-5 years, often followed by high recruitment (Gribko et al. 2002). They are wind-pollinated and self-incompatible (Schwarzmann and Gerhold 1991, Ducousso et al. 1993). Their most important dispersers in the Southeastern US are the gray squirrel (*Sciurus carolinensis*) and the blue jay (*Cyanocitta cristata*) (Vander Wall 2001), which both scatter-hoard acorns in 1-2 nut caches. Jays are the only dispersal vector of oaks known to regularly transport acorns hundreds of meters from the parent tree (Johnson and Webb 1989). Long-distance dispersal of acorns by the European jay is thought to have an important influence on the distribution of Spanish oaks (Gomez 2003, Purves et al. 2007). Whether blue jays or scrub jays play a similarly important role in North America relative to short-range rodent dispersers has not been established; apart from a few studies in the 1980’s and ‘90’s, there has been little research on the subject (DeGange et al. 1989, Johnson and Webb 1989, Johnson et al. 1997).

Data collection

Sites originally contained 1x2 m permanent seedling census plots (124 at Duke forest, 70 at Coweeta) arrayed in cross-shaped transects centered on canopy gaps and extending into the understory. In the Duke Forest, where the seedling layer is sparser, I added 79 additional $1m^2$ and 70 $7m^2$ census plots to increase sample size and to provide better representation of short- and long-range dispersal events. These plots were censused each spring to identify newly emerged or dead individuals. Seedlings from the permanent plots were sampled non-destructively, with < 1 g of leaf tissue required for genotyping. Adult canopy leaves were obtained using a slingshot. Leaf tissue was stored at -80° C prior to DNA extraction. Total genomic DNA was extracted from leaf tissue using a modified CTAB protocol (**Appendix 1.1**). Six nuclear microsatellites isolated by Aldrich et al.(2002, 2003a) were analyzed using GeneMarker (Softgenetics LLC, 2004). All loci were highly polymorphic and all individuals had unique genotypes.

Table 1: Number of individuals genotyped.

"Trees" defined as individuals > 2 m tall, regardless of diameter. Trees > 10 cm dbh considered as potential parents

Site	Oak trees	Potential parents (trees >10 cm dbh)	Oak seedlings
Duke Forest	173	118	219
Coweeta	201	199	179

Only trees > 10 cm diameter-at-breast-height were considered potential parents (table 1). As red oaks < 25 cm dbh are seldom reproductively mature (Clark et al. In press), this is a conservative cutoff intended to avoid excluding rare early-maturing individuals.

Chapter 2. Estimating seed and pollen movement in a monoecious plant: a hierarchical Bayesian approach integrating genetic and ecological data

Introduction

Because seed arrival is a prerequisite to population establishment, dispersal ability has a strong influence on migration and invasion potential in plants. The spatial scale of gene flow also has important implications for population dynamics, the maintenance of genetic diversity, and the effectiveness of natural selection (Kawecki 2008). Where dispersal and gene flow are limited, genetic diversity can be quickly depleted due to drift, strong selection, or a combination of the two (Gillespie 2004), especially in self-incompatible species (Sork et al. 2002). Immigration can improve adaptive potential by increasing genetic variation (Kimbrell and Holt 2007), but local adaptation at range limits and in marginal habitats can also be inhibited when the influx of maladapted genes from the main part of the species range exceeds the rate at which they are purged by selection (Kirkpatrick and Barton 1997, Rehfeldt et al. 1999, Lenormand 2002, Lopez et al. 2007). As species have historically responded to the strong selective pressure of climate change via both migration and local adaptation (Davis and Shaw 2001), the influence of seed and pollen dispersal on these processes is of particular interest in the context of current anthropogenic climate change (Holt 1990, Skelly et al. 2007), but unobserved dispersal processes and unknowns associated with genetic data have presented challenges. Here I introduce a new approach that can be used to estimate seed and pollen movement, taking into account the many types of uncertainty associated with sampling genotypes and with the dispersal process itself.

For plants, the movement of pollen and seed are the only source of gene flow within and between populations, while seed movement alone allows for range expansion and the colonization of new sites. The probability distributions of seed and pollen movement, also known as dispersal kernels, can be challenging to estimate (Clark et al. 2004). Because movement in plants is restricted to two life stages, parentage information is highly informative of dispersal at these stages. For most wild populations, pedigrees are unknown. In some animal populations, partial pedigrees have been obtained through long-term field observations (Pemberton 2008), but this approach is not feasible for many species, including trees, where mating and dispersal are cryptic and individuals may live (and reproduce) for centuries. For this reason, molecular markers, particularly microsatellites, are increasingly used to infer parentage, sibship, or population of origin (Dow and Ashley 1996, Streiff et al. 1999, Godoy and Jordano 2001, Asuka et al. 2005, Bacles et al. 2006, Hardesty et al. 2006, Pairon et al. 2006, Selkoe and Toonen 2006).

The use of molecular markers in parentage and dispersal studies presents its own challenges. Early parentage analyses were based on excluding adults that, at a given locus, did not share an allele with the juvenile under consideration (eg. Dow, Ashley, 1996). But the probability of genotyping error is not trivial for microsatellites (Dewoody et al. 2006). Simple exclusion may lead to the rejection of true relationships when a mutation or genotyping error causes the observed genotype of parent and offspring to differ. The popular parentage analysis software CERVUS (Marshall et al. 1998) is an extension of the rejection approach, calculating the log likelihood ratio (termed the LOD score) for each proposed parent based on the observed genotypes and a user-specified number of allowed mis-matches. This approach is most effective in species with separate sexes, as parentage assignments are based solely on genotype and do not include other adult characteristics that might narrow the pools of potential mothers and fathers. CERVUS does not explicitly allow for reproductive contributions by individuals outside the sampled area; all offspring are matched to genotyped adults, even if the LOD score is very low. The first model developed to investigate bi-parental gene flow in monoecious plants was the ‘seedling neighborhood model’ (Gonzalez-Martinez et al. 2006). However, as the authors acknowledge, because it relies on arbitrary ‘neighborhoods’ around seedlings and mother trees, this model does not adequately characterize the dispersal kernel – which is instead calculated using an LOD approach. Moreover, genotyped adults outside the neighborhood are not explicitly considered as parents, and genotypes are assumed to be observed without error. Because most parentage studies aim to estimate other population-level parameters (such as dispersal, selfing rates, or extra-pair mating frequencies) rather than simply generate a pedigree, Hadfield et al. (2006) proposed a Bayesian approach that estimates these parameters simultaneously and incorporates other types of data such as social status or territory location. They found that this joint estimation improves pedigree estimates and decreases bias in population parameters (when the assumptions of the specific model are met). For plants, the problem is more complex, because any given individual of a monoecious species can be a mother and a father, and pollen and seed dispersal kernels can differ substantially.

The model presented here is the first to simultaneously estimate parentage, seed dispersal, and pollen dispersal in a hermaphroditic plant species. Building on the approach of Hadfield et al.(2006), I take into account both genetic and ecological data as well as genotyping error. A unique feature of the model is that it treats dispersal coherently, the same process governing seed and pollen movement both inside and outside the defined area where all potential parents are genotyped.

In the present study, I develop a Bayesian hierarchical model to estimate parentage and effective dispersal distances in the general case of a self-incompatible monoecious tree. I use this model to measure the scale of gene flow via seed and pollen in a mixed-species population of red oaks (*Quercus rubra, Q. velutina, Q. falcata*) located in central North Carolina. Red oaks, important both as timber trees and as providers of hard mast for wildlife (Little 1980, McShea et al. 2006), are currently experiencing recruitment failures in many parts of their range (Abrams 1992, Aldrich et al. 2005b). This is usually attributed to changes in disturbance frequencies and herbivory by deer (Elliott et al. 1999, Spetich 2004). In addition, because oaks have heavy seeds dispersed primarily by rodents, they may be less likely to reach scarce recruitment sites than species dispersed by wind or frugivorous birds (Sork 1984, Clark et al. 2004, Garcia and Houle 2005). They may also be at a disadvantage in shifting their geographic ranges in response to climate change (**Chapter 6**). Limited dispersal by seed or pollen or both could lead to the erosion of genetic diversity and adaptive potential under strong selective pressures such as climate change. Because the focal life stage in this study is seedlings, not seeds, dispersal parameters reflect "effective dispersal" after germination and initial seedling mortality (see **Chapter 3**).

Methods

In this chapter, I make use of data collected from the Duke Forest population, as described in **Chapter 1**. Oaks are self-incompatible (Schwarzmann and Gerhold 1991, Ducousso et al. 1993), so I hereafter assume that the same individual cannot be both mother and father to a given seedling. Genetic structure analyses for oak species at Duke Forest show little between-species differentiation at neutral loci, supporting the hypothesis that the three morphologically-defined species *Quercus rubra*, *Q. velutina*, and *Q. falcata* interbreed at this site (**Chapter 5**). Consequently, all three species are considered as potential parents and offspring in the analysis that follows.

Model development:

Genotypes and dispersal:

Consider a population in which mature individuals I produce both pollen and seeds. These adult trees exist in a mapped area that is exhaustively sampled. Any individual $i \subseteq I$ can serve as a mother or a father. Pollen released from individual i' may disperse to and fertilize a flower from individual i. Due to self-incompatibility, $i' \neq i$, but this assumption of exogamous pollen could be relaxed in species known to self-pollinate. Potential parents are characterized by genotype and location $\{(G_{i,l}, s_i), i = 1, \ldots, n; l = 1, \ldots L\}$, where $s_i = (x_i, y_i)$ are map coordinates, l are loci, and $G_{i,l} = (a_{1i}, a_{2i})_l$ is the length-two vector of alleles at locus l, where $(a_{1il}, a_{2il}) \in A_l$, and A_l is the set of all n_l alleles in the population at that locus. The frequency of alleles in the population at locus l is the length n_l vector $freq(a_l) = \lfloor a_{l1}, \ldots, a_{ln_l} \rfloor$, each element being equivalent to the probability of drawing allele $1 \ldots n_l$ for locus l at random from the population. Assuming alleles are independent (as one would expect in a largely outbreeding population), the probability of a given genotype (a_1,a_2), drawn at random from the population, will be $p(G_l) = freq(a_{1l})freq(a_{2l})$.

The probability of fertilization of individual i by i' depends on pollen production $c_{i'}$ and dispersal distance $d_{i,i'} = \|s_{i'} - s_i\|$. Seeds dispersed from mother i to the location of offspring k, depend on distance $d_{ik} = \|s_i - s_k\|$. Seed production, $f_{i,t,}$ by all trees for years t = 2000, …, 2008 in the plot has been estimated in a separate study (Clark et al. 2004, Clark et al. In press). Because many of the sampled seedlings recruited before the beginning of the present study and their exact age is not known, seedling parentage effectively integrates over multiple years of seed production. I therefore incorporate variation and uncertainty in fecundity by defining a mean and standard deviation for f_i over the 2000-2008 period and, at each iteration of the Gibbs sampler, drawing a new value for f_i from this distribution (See "Implementation", **Appendix 1.2**). Trees that are large and fecund tend to produce more pollen than trees that are small or immature. However, detailed studies on male and female allocation within individual oaks (as opposed to at the stand level (Saito et al. 1987)) are lacking in the literature. In the absence of more detailed information, pollen grains produced per father per year $c_{i'}$ is assumed to be proportional to estimated seed production, $f_{i'}$, for the same individual. Genotype data is the ultimate arbiter of whether a tree is a potential father.

In addition to the adult trees, there is a sample of seedlings $k = 1,\ldots,K$, each characterized not only by genotype G_k and location s_k, but also by pedigree, where $P_k = (i', i)$ indicates that k has mother i and father i'. The pedigree is not known, but rather will be estimated based on genotype and dispersal. The genotype of k at a given locus consists of one allele contributed by the mother and one by the father.

We now consider the probability of pedigree P_k, i.e., the probability that i is the mother and i' is the father of k, which depends on the genotypes of all three individuals and on any other factors that affect the probability that individual k could have parents (i',i): $p(P_k) = p((G_i, G_{i'}) \rightarrow G_k) w_{i',i,k}$

The parameter $w_{i',i,k}$ is a weight that can be assigned to $P_k(i',i)$ relative to other potential pedigrees for k. In this example, the weights depend on fecundity and dispersal of pollen from $s_{i'}$ to s_i, and of seed from s_i to s_k,

$$w_{i',i,k} = c_{i'}\, p(d_{i'i} | u_p) f_i\, p(d_{ik} | u_s)$$

but other criteria could be used. This probability is relative to all other potential parents, expressed as the ratio

$$p(P_k = (i,i') | G_k, G_i, G_{i'}, w_k) = \frac{p(G_k | P_k = (i,i'), G_{i'}, G_i) w_{i',i,k}}{\sum_{i' \in I} \sum_{i \in I} p(G_k | P_k = (i,i'), G_{i'}, G_i) w_{i',i,k}} \qquad [1]$$

This equation can be broken down further into its component parts. We begin with genotype. Given that *i* and *i′* are parents of *k*, the probability of the offspring genotype is:

$$p\left(G_k \middle| P_k = (i',i), G_{i'}, G_i\right) \propto \prod_{l=1}^{L} p\left(G_{kl} \middle| P_k = (i',i), G_{i'l}, G_{il}\right) \qquad [2]$$

The two sides of this equation are expressed as a proportionality, because it will be normalized after accounting for pedigree weights (equation 1). The factors on the right-hand side of eqn 2 are the standard Mendelian probabilities for diploid organisms. For example, if the parental genotypes are *Aa* and *AA*, the probability that they produce an offspring with genotype *Aa* is 0.5, while the probability of an offspring with genotype *aa* is zero. In other words, the probability equals zero if $G_{kl} = (a_{1kl}, a_{2kl}) \not\subset (G_{il} \cup G_{i'l})$. Note that we could exchange subscripts *i* and *i′*, representing the equivalent case for the mother being the father and vice versa, and the probability of producing a given offspring genotype would be the same. Probabilities are not equivalent once dispersal is taken into account, because dispersal probabilities of seed and pollen differ.

The dispersal kernel is a density function representing the probability of seed or pollen traveling a given distance from the parent tree. Previous studies show that for animal dispersed seed and wind dispersed pollen the dispersal kernel is usually fat-tailed (Clark et al. 1999, Streiff et al. 1999, Hardesty et al. 2006), with both more short-distance and more long-distance events than in a Gaussian distribution. To facilitate comparison with previous work by Clark et al. (1999, 2001, 2005), a 2Dt kernel was chosen to represent both seed and pollen dispersal probabilities, although the general model structure can accommodate other types of distributions. The probability of pollen or seed traveling a given distance d is given as:

$$p(d) = \frac{1}{\pi u\left(1+\frac{d^2}{u}\right)^2} \qquad [3]$$

where the shape of the kernel is determined by the parameter u (u_p for pollen, u_s for seed). The mean dispersal distance is given by:

$$E(d) = \pi/_2\sqrt{u} \qquad [4]$$

The expected density of seed or pollen reaching a given point k from a particular source tree i is equal to the probability of the seed or pollen grain travelling the distance d_{ki}, given by the dispersal kernel, times the fecundity or pollen production of the source tree. Because the expected amount of seed is in units of seeds/m^2, this quantity is multiplied by the size of the plot to approximate the number of seeds expected to reach that plot.

Because focal populations in population-genetic studies are seldom completely isolated, it is important to allow for the possibility that parents of a sampled offspring reside outside the sampled area. In this model, I assume that the sampled area is part of a continuous population and that the density of adult trees outside the plot is equal to the density of adults inside, allowing us to approximate expected seed and pollen received from out-of-plot sources via numerical integration (**Appendix 1.3**). This assumption is appropriate when dealing with continuous forest, as in the present example, but may not be justified in all situations. If information exists about the distribution of out-of-plot seed or pollen sources, this can and should be included.

Combining equations 1 and 2 we have:

$$p\left(P_k = (i,i')\middle|\{G\},\{d\}\right) = \frac{\prod_{l=1}^{L} p\left(G_{k,l}\middle|P_k = (i,i')\right)c_{i'}\,p\left(d_{i'i}\right)f_i\,p\left(d_{ik}\right)}{\sum_{i'\in I}\sum_{i\in I}\prod_{l=1}^{L} p\left(G_{k,l}\middle|P_k = (i,i')\right)c_{i'}\,p\left(d_{i'i}\right)f_i\,p\left(d_{ik}\right)} \qquad [5]$$

Genotype error:

Genotype errors (figure 1) are predominantly of two varieties: Mistyping causes one allele to be mistaken for another (usually of similar length), while allelic dropout causes a heterozygote to look like a homozygote (Dewoody et al. 2006). Both can be estimated by repeated genotyping of individuals and loci (Bonin et al. 2004). This was done for all six loci, using data from two study populations. Across loci, mistyping occurred at an average of 5.7% (range 2-18%) of re-genotyped alleles and dropout at 5% (range 2-8%) (table 15). These rates are high, but microsatellites often exhibit high error rates (Bonin et al. 2004, Burczyk et al. 2004, Dewoody et al. 2006). In this case, the high concentrations of tannins and other secondary compounds in oak leaves made it challenging to obtain clean DNA samples of consistent concentration; Amplification success for a single individual could vary considerably from one extraction to another (see **Appendix 1.1**). I develop models for the two main error types.

Mistyping occurs when an allele is amplified using PCR and some copies are longer or shorter than the true length. This "stutter" can cause the length of an allele to be misread by one repeat-length (Garant et al. 2001) - in this case, two base-pairs. Previous models (Marshall et al. 1998, Hadfield et al. 2006) have generally assumed that if an allele is mistyped, the probability of observing any "false" allele is proportional to the frequency of the "false" allele in the population. However, because it is unlikely that the observed allele will differ greatly in length from the true allele except in the rare case of contamination or sample mis-labeling (Garant et al. 2001, Bonin et al. 2004) I assume that only alleles adjacent in length (differing by 1-2 base pairs, and expressed in table 2 as $a^o = a \pm 1$) can be mistaken for one another. Differences of one repeat length between parent and offspring or between two samples from the same tree may also occur due to mutation. Microsatellite markers have high mutation rates, which generate the high intra-population variation that makes them useful for parentage analysis. However, by allowing for a relatively high rate of "mistyping", we prevent the inadvertent exclusion of potential parents due to either mistyping or mutation.

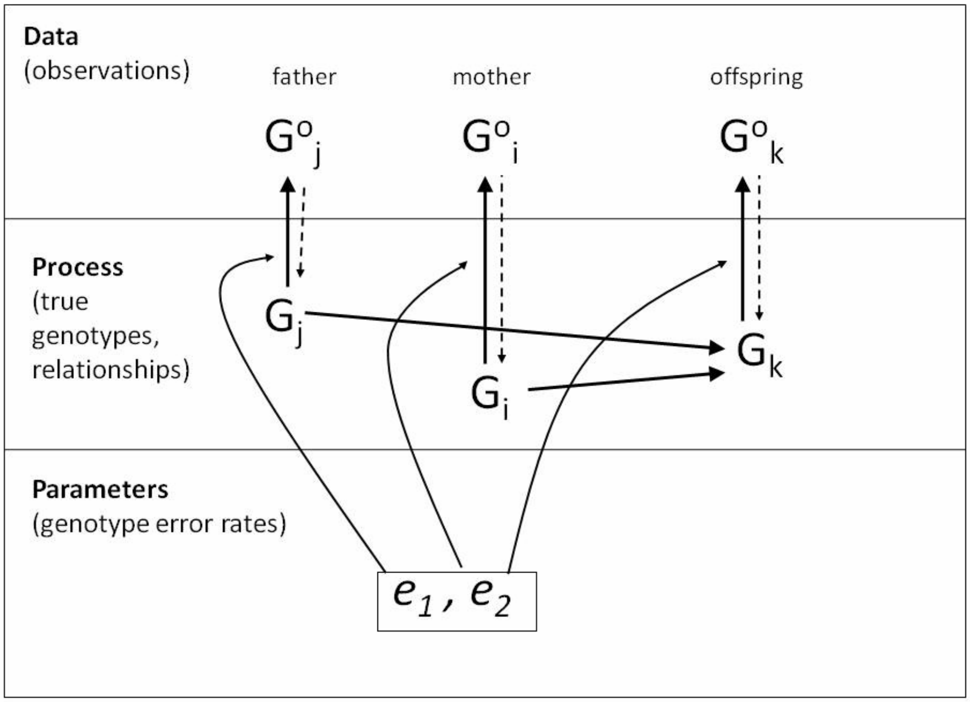

Figure 1: Relationship between true and observed genotype.
Dashed arrows indicate that we can calculate the probability of a given true genotype given the observed genotype, as well as the probability of the observed given the truth (solid line).

Allelic dropout occurs when one of the two alleles at a locus fails to amplify (expressed in table 2 as $a^o = 0$). Like mistyping, this error rate can be estimated by regenotyping multiple individuals and loci, because frequently the allele that was missed on the first genotyping will be detected in the second, and vice versa. The probability that a heterozygote will be observed to be a homozygote in my model is based on this regenotyping data. Null alleles can also cause a heterozygote to be typed as a homozygote, and are more difficult to identify because they *never* amplify. Thus, if an *A0* individual mates with a *BB* individual, their offspring can have the genotype *B0*, which looks like *BB*. The presence of null alleles is suggested by an excess of homozygotes in a population but, because this can also result from inbreeding, null allele frequencies are difficult to estimate with confidence. Many oak populations exhibit mild heterozygote deficiency, generally attributed to spatial genetic structure leading to non-self inbreeding, or to assortitive mating (Ducousso et al. 1993). As with mistyping and mutation, my method of treating allelic dropout will ensure that individuals that are homozygous due to null alleles are not eliminated as potential parents or offspring, but it should be noted that the total probability of a heterozygote being identified as a homozygote may be an underestimate.

Let G_i^o be the observed genotype, which can differ from the true genotype of individual *i* by mistyping or dropout error. A mistyping error is the event E_1 and occurs with probability $p(E_1) = e_1$ and a dropout is event E_2 with probability $p(E_2) = e_2$. These probabilities are taken as fixed for each locus, and are determined by re-genotyping many individuals and loci. For two alleles at each locus, define the matrix $\mathbf{E} = \begin{pmatrix} E_{1,a1}, E_{2,a1} \\ E_{1,a2}, E_{2,a2} \end{pmatrix}$ of binary indicators, where the first row represents one allele and the second row the other allele. Both types of errors can occur for either allele. Both errors *can* occur simultaneously at a locus, but if two alleles are observed we know that event E_2 has not occurred because dropout always results in the appearance of a homozygote, even if the other allele has been mistyped. If a mistyping and a dropout event were to occur at the same allele, only the dropout event will be observed. In table 2, I take this event into account. When neither allele at a locus is observed this could be due to the same processes that cause only one allele to amplify, but it may also be due to other causes – a badly degraded sample, for instance. I therefore assume that we do not observe the case where both alleles drop out, which occurs with probability $\mathrm{p}\begin{pmatrix}0,1\\0,1\end{pmatrix}+p\begin{pmatrix}1,1\\0,1\end{pmatrix}+p\begin{pmatrix}0,1\\1,1\end{pmatrix}+p\begin{pmatrix}1,1\\1,1\end{pmatrix}=e_2^2$; the probabilities in table 2 are normalized by $(1-e_2^2)$. Notice that the consequences of each possible event are different for homozygotes and heterozygotes.

Table 2: Genotyping error probabilities

	No error	1 allele mistyped	1 allele dropped	1 allele dropped, the other mistyped	Both alleles mistyped
Observed genotypes	$(a_1^o = a_1, a_2^o = a_2)$	$(a_1^o = a_1 \pm 1, a_2^o = a_2)$ or $(a_1^o = a_1, a_2^o = a_2 \pm 1)$	$(a_1^o = a_1, a_2^o = 0)$ or $(a_1^o = 0, a_2^o = a_2)$	$(a_1^o = a_1 \pm 1, a_2^o = 0)$ or $(a_1^o = 0, a_2^o = a_2 \pm 1)$	$(a_1^o = a_1 \pm 1, a_2^o = a_2 \pm 1)$
Event $\mathbf{E} = \begin{pmatrix} E_{1,a1}, E_{2,a1} \\ E_{1,a2}, E_{2,a2} \end{pmatrix}$	$\begin{pmatrix} 0,0 \\ 0,0 \end{pmatrix}$	$\begin{pmatrix} 1,0 \\ 0,0 \end{pmatrix}$ or $\begin{pmatrix} 0,0 \\ 1,0 \end{pmatrix}$	$\begin{pmatrix} 0,0 \\ 0,1 \end{pmatrix} \begin{pmatrix} 0,0 \\ 1,1 \end{pmatrix}$ or $\begin{pmatrix} 0,1 \\ 0,0 \end{pmatrix} \begin{pmatrix} 1,1 \\ 0,0 \end{pmatrix}$	$\begin{pmatrix} 1,0 \\ 0,1 \end{pmatrix} \begin{pmatrix} 1,0 \\ 1,1 \end{pmatrix}$ or $\begin{pmatrix} 0,1 \\ 1,0 \end{pmatrix} \begin{pmatrix} 1,1 \\ 1,0 \end{pmatrix}$	$\begin{pmatrix} 1,0 \\ 1,0 \end{pmatrix}$
# of observable combinations	1	4	2	4	4
Probability of each combination	$\frac{(1-e_1)^2(1-e_2)^2}{1-e_2^2}$	$\frac{e_1(1-e_1)(1-e_2)^2}{2(1-e_2^2)}$	$\frac{((1-e_1)^2 e_2(1-e_2))}{1-e_2^2}$ + $\frac{(e_1(1-e_1)e_2(1-e_2))}{1-e_2^2}$	$\frac{e_1(1-e_1)e_2(1-e_2)}{2(1-e_2^2)} + \frac{e_1^2 e_2(1-e_2)}{2(1-e_2^2)}$	$\frac{e_1^2(1-e_2)^2}{4(1-e_2^2)}$
Total probability $p(\mathbf{E})$	$\frac{(1-e_1)^2(1-e_2)^2}{1-e_2^2}$	$\frac{2e_1(1-e_1)(1-e_2)^2}{1-e_2^2}$	$\frac{2(1-e_1)e_2(1-e_2)}{1-e_2^2}$	$\frac{2e_1e_2(1-e_2)}{1-e_2^2}$	$\frac{e_1^2(1-e_2)^2}{1-e_2^2}$

Building on equation 5, the full model for all individuals is:

$$p(P,u_s,u_p \left| \{G^o\},\{d\}) \propto p(\{d\}\right| u_s,u_p,P)p(\{G^o\}|P)p(u_s)p(u_p) \qquad [6]$$

where $\{d\}$ is the set of distances between pairs of individuals, u_s and u_p are seed and pollen dispersal parameters, $\{G^o\}$ is the set of observed genotypes of all individuals, and p(u_s) and p(u_p) are Gaussian priors on the dispersal parameters. Based on data from the literature, I assigned u_s a prior mean of 253, corresponding to a mean dispersal distance of 25 m, and a prior standard deviation of 2000. I assigned u_p a prior mean of 1000, corresponding to a mean dispersal distance of 70.2 m, and a prior standard deviation of 1500 (see **Appendix 1.4**). In expanded format, the model may be written as:

$$p(P,u_s,u_p | \{G^o\},\{d\},e_1,e_2,\{f\},\{c\}) \propto$$

$$\prod_k \frac{c_{i'} p\left(d_{i'i} \middle| u_p\right) f_i p\left(d_{ik} \middle| u_s\right)}{\sum_{i,i'} c_{i'} p\left(d_{i'i} \middle| u_p\right) f_i p\left(d_{ik} \middle| u_s\right)} \prod_l \frac{p\left(G^o_{k,l} \middle| G^o_{i',l}, G^o_{i,l}, e_1, e_2\right)}{\sum_{i,i'} p\left(G^o_{k,l} \middle| G^o_{i',l}, G^o_{i,l}, e_1, e_2\right)} p(u_s)p(u_p) \qquad [7]$$

Thus far, the model appears computationally demanding, because conditional probabilities are expressed in terms of latent genotypes (equations 1 and 2), which are observed with error and therefore must be estimated. Implementation with MCMC would require substantial overhead to sample latent variables due to the large number of potential pedigree combinations. These true states do not appear in equation 7 because we can marginalize them away, expressing observed offspring genotype conditioned directly on observed parent genotypes. Here we demonstrate that this is the case.

Consider the factor of equation 7 relating to the probability of the observed offspring genotype given the observed genotypes of the proposed parents. The existence of genotyping error means that the observed genotype for an individual is not necessarily identical to its true genotype. However, because the only quantities of interest are *P* and the dispersal parameters, we will not estimate these true genotypes directly. Using the observed genotypes, and the genotyping error distributions, we can calculate the probability that a given pair of parents could give rise to an observed offspring genotype:

$$p\left(G_k^o \middle| G_i^o, G_j^o\right) = \prod_l \sum_{G_{k,l}} p\left(G_{k,l}^o \middle| G_{k,l}\right) \sum_{G_{i,l}} \sum_{G_{j,l}} p\left(G_{k,l} \middle| G_{i,l}, G_{j,l}\right) p\left(G_{j,l} \middle| G_{j,l}^o\right) p\left(G_{i,l} \middle| G_{i,l}^o\right)$$

$$= \prod_l p\left(G_{k,l}^o \middle\| G_{i,l}^o, G_{j,l}^o\right) \qquad [8]$$

The probabilities $p(G_l^o | G_l)$ are contained in table 2. To obtain $p(G_l | G_l^o)$, we use Bayes theorem:

$$p(G_l | G_l^o) = \frac{p(G_l^o | G_l) P(G_l)}{\sum_G p(G_l^o | G_l) P(G_l)} \qquad [9]$$

Marginalizing away the true states allows us to build efficient algorithms for posterior simulation.

Implementation:

Computation was implemented in R. The $P(G_k^o | G_i^o, G_{i'}^o)$ are computationally expensive and independent of the dispersal parameters, and so were evaluated before MCMC simulation, as described below. For each offspring, we create an $(n_a+1) \times (n_a+1)$

matrix $Amat_k$, where n_a is the number of genotyped adults. $Amat_k[i,i']$ represents the probability of obtaining the observed genotype of offspring k given $P_k = (i,i')$ relative to all possible parent combinations, and where row (n_a+1) represents a hypothetical out-of-plot mother, and column (n_a+1) a hypothetical out-of-plot father.

First, $p\left(G_{i,l} \middle| G_{i,l}^o\right)$ is calculated at a given locus l for all potential true genotypes G_l for each adult i using equation 11 and the probabilities given in table 2. If an adult is ungenotyped at locus l, or if i represents a hypothetical ungenotyped out-of-plot parent, then $p\left(G_{i,l} \middle| G_{i,l}^o\right) = p(G_l) = freq(G_{1,l})freq(G_{2,l})$. Then for the parent pair (i,i') we calculate $p\left(G_{k,l} \middle| G_{i,l}, G_{j,l}\right)$ using Mendelian inheritance probabilities and $p\left(G_{i,l} \middle| G_{i,l}^o\right)$, and store these probabilities in an n_l x n_l matrix, *Nmat*. We then calculate $p\left(G_{k,l}^o \middle| G_{k,l}\right)$ for each offspring using the probabilities in table 2, and store these probabilities in an n_l x n_l matrix, *Omat*. Finally, $p\left(G_{k,l}^o \middle| G_{i,l}^o, G_{i',l}^o\right) = \sum_G \mathrm{Nmat[a1,a2]Omat[a1,a2]}$, where G indicates all potential true offspring genotypes (a_1,a_2), and $Amat_k[i,i'] = \prod_L p\left(G_{k,l}^o \middle| G_{i,l}^o, G_{i',l}^o\right) = p\left(G_k^o \middle| G_i^o, G_{i'}^o\right)$

The MCMC was then implemented in the following sequence:

1) Initialize chain.

An initial pedigree $P_k = (m_k,f_k)$ is generated for each seedling using $Amat_k$, with a random draw $(m_k,f_k) \sim \text{multinom}(Amat_k)$. Initial values for u_s and u_p are chosen such that

$u_p > u_s$, because in a wind-pollinated, largely rodent-dispersed plant, the average distance of pollen movement is expected to be longer than average seed movement.

Then for each step in the Gibbs sampler:

2) Draw values for f_i, c_i

Distributions of fecundity values reflecting both year-to-year variation and uncertainty in annual fecundity estimates are developed for each tree as described in **Appendix 1.2**. A new value for f_i is drawn at the beginning of each Gibbs step to mix over this variation and uncertainty; c_i is assumed to be proportional to fecundity.

3) Sampling of u_s,u_p conditioned on P_k

Dispersal parameters are sampled with a Metropolis step from the conditional distribution:

$$p(u_s,u_p) = \frac{\prod_k c_j p\left(d_{ji}\middle|u_p\right) f_i p\left(d_{ik}\middle|u_s\right)}{\prod_k \sum_i \sum_j c_j p\left(d_{ji}\middle|u_p\right) f_i p\left(d_{ik}\middle|u_s\right)} p(u_p|m_p,s_p)p(u_s|m_s,s_p)$$

where i and j are the currently imputed parents, m_p and m_s are the prior means, and s_p and s_s are the prior standard deviations. If $p_{new} > p_{now}$, where p_{new} is the conditional probability of the proposed vaules and p_{now} is the conditional probability of the current values, the proposed values are accepted. If $p_{new} < p_{now}$, the proposed parameter values are accepted with probability $a = p_{new}/p_{now}$.

4) Sampling of P_k conditional on u_s,u_p

Each seedling has a currently imputed pedigree – a mother/father pair (i,j). For the purposes of proposing new pedigree values, an $(n_a+1) \times (n_a+1)$ matrix, $ppmat_k$, is created for each seedling such that $ppmat_k$ [x,y] = 1 if $Amat_k$ [x,y] > 0; otherwise, $ppmat_k\,[x,y]$ = 0. A new pedigree is proposed from $(i^*,j^*) \sim$ multinom($ppmat_k$). This step speeds convergence by avoiding proposing parent pairs deemed impossible based on genotype, while allowing all combinations of parents *not* ruled out by genotype to be explored

We then evaluate the conditional probability of the proposed pedigree relative to the current pedigree, given the currently imputed dispersal parameters using:

$$p(i,j) = \frac{c_j p\left(d_{ji} \middle| u_p\right) f_i p\left(d_{ik} \middle| u_s\right) p(G_k^o | G_i^o, G_j^o)}{\sum_i \sum_j c_j p\left(d_{ji} \middle| u_p\right) f_i p\left(d_{ik} \middle| u_s\right) p(G_k^o | G_i^o, G_j^o)}$$

for father *j* and mother *i*. The proposed values are accepted or rejected for each seedling as described in the previous step.

5) Steps 2-4 are repeated until the chain converges.

Simulation:

Multiple simulations were conducted from different initial conditions to assure that chains converged to the posterior distribution (**Appendix 1.5**). Estimates of u_s and u_p converged quickly – generally within 100 to 1500 steps, depending on initial conditions. Testing with simulated data sets showed that the approach assigned the highest probabilities to the correct parent pair 97% of the time on average. Incorrect parentage assignment was usually caused by a large (>3) number of genotyping errors or ungenotyped loci in the parent-offspring pair. For an average of 86% of seedlings in a given simulation, the most frequently identified mother and father were the true mother and father, whereas for 11% the parents were "inverted" – the true mother identified as the father and vice versa. Inversions occur because the only information we have that can help to identify mother versus father is their location; however the occurrence of inversions did not have large effects on the dispersal parameter estimates. For simulations in which the stand dimensions and plot number were those of the actual Duke Forest stand, true parameter values fell within the 95% CI of the dispersal estimate in all simulations, and 83.3% of the simulations for both parameters. Estimates deteriorated with small sample sizes (few trees or seedlings in the simulation).

Application to field data:

The mapped plot contained 118 potential parent trees, while seedling plots contained 219 red oak seedlings. Multiple independent runs were performed with different initial values for parentage and dispersal parameters. The chains were run for a total of 50,000 steps, with a burn-in of 30,000 steps.

Results

Parentage:

Independent runs show that both parentage and dispersal estimates converged to the posterior distributions. Of 219 genotyped seedlings, for 16% the estimated parents were both genotyped, in-plot adults. For 19.6% of seedlings, the father was estimated to be an in-plot individual and the mother an out-of-plot (unsampled) individual, while for

27.4% the father was identified as an out-of-plot individual and the mother as an in-plot individual. For 37% of seedlings, neither parent was estimated to be among the genotyped trees within the 12 ha plot. In-plot mother-offspring pairs are shown in figure 2.

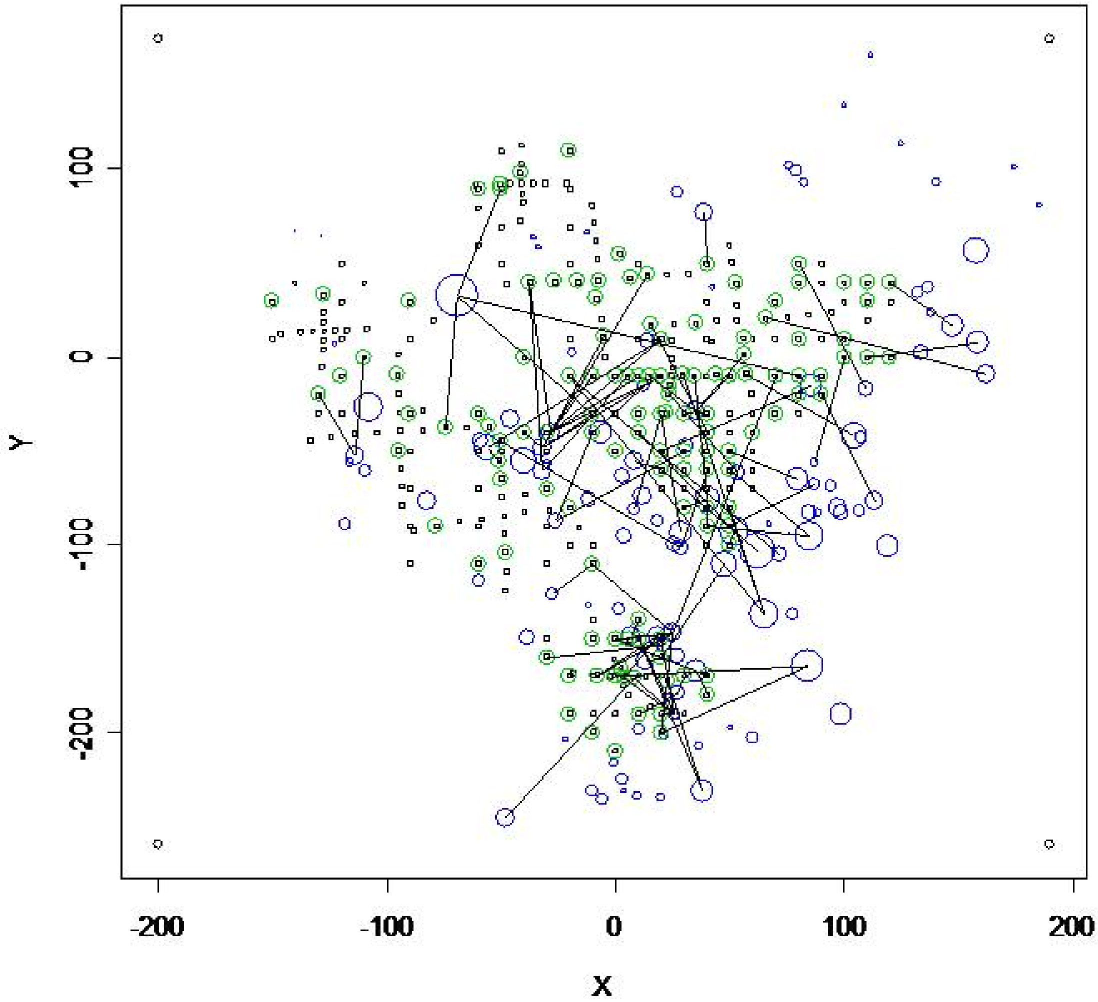

Figure 2: In-plot mother-offspring pairs (black lines)

Blue circles- adult trees, size proportional to diameter. Green circles – seedlings. Black squares – seedling sampling plots.

I compared the performance of my model to CERVUS (Marshall et al. 1998), which uses maximum likelihood to assign parentage based on genotype alone, using both simulated data and Duke Forest data. Simulated data was created using the methods described in **Appendix 1.5** for the true stand size and seedling plot configuration. Dispersal parameters were set to $u_s = 6300$, $u_p = 12{,}900$. For my simulated individuals, the overall genotyping error rate was 5%, and the 3.5% of loci were ungenotyped. Of the 209 simulated seedlings, 20.6% had out-of-plot mothers, 41.1% out-of-plot fathers, so the percentage of parents sampled was set to 70%. For 40.7% of seedlings, CERVUS identified the correct parent pair, while in a further 21.1% it correctly identified the one true in-plot parent (assigning another individual in place of the out-of-plot parent). In 13.4% of seedlings it only identified one parent correctly even though both parents were present in the plot, while in the remaining 24.8% of seedlings neither parent was correctly identified.

For the actual Duke Forest data, I limited potential parents to individuals >10 cm DBH, for greater consistency with the Bayesian model. For 10% of Duke Forest seedlings the Bayesian model identified the same parent pair as CERVUS, while for a further 37.9% it identified one of the parents with the highest LOD score. Where results of the Bayesian model and CERVUS differed (52.1% of seedlings), this was generally because a) the parent tree identified by CERVUS had extremely low estimated fecundity, b) the models weight the likelihood of certain allelic mismatches differently, or c) there were several untyped loci in the parent or offspring or both, which increases uncertainty. Because CERVUS must choose parent pairs from among genotyped trees (ie. no out of plot parents), LOD scores tended to be low, and the number of mismatched alleles high. Only 6% of the best matches between parent pairs and seedlings for the Duke Forest CERVUS analysis exceeded the 95% critical LOD level. Offspring matches that require no genotyping errors are consistently identified by both models, but such matches were comparatively rare for this dataset (28.7% of parent-offspring matches).

Dispersal:

The seed dispersal parameter, u_s, converged at 6300 (se 460), corresponding to a mean dispersal distance of 127.7 meters. The pollen dispersal parameter converged at 12,900 (se 510), corresponding to a mean dispersal distance of 178.2 m. The mean distance between mothers and offspring within the plot was 72.4 m (range 3.1 to 248 m) and the mean father-mother distance was 101.6 m (range 8.7 to 229 m). These values are both shorter than the means for the dispersal kernel, as the overall estimate takes into account dispersal from outside the plot. For comparison, the average distance from a seedling to the nearest adult tree was 14.4 m with a maximum observable mother-offspring distance of 374 m, and the average nearest-neighbor distance between adults 14.9 m with a maximum observable father-mother distance of 413 m.

Discussion

This model is the first developed to simultaneously estimate parentage, seed dispersal, and pollen dispersal in a hermaphroditic plant species. One notable feature of the model is that it treats dispersal coherently, the same process governing seed and pollen movement inside and outside the mapped area. It also explicitly considers two common types of genotyping error. This model is also substantially better at identifying parents of juveniles than CERVUS when a) individuals are hermaphroditic; b) the probability of mating or parentage depends on distance between individuals; and c) a large number of potential parents are located outside the exhaustively sampled area.

This model makes several simplifying assumptions, which must be considered when applying these methods. Where data on pollen production can be obtained, this information can and should be substituted for the simplistic assumption of proportional seed and pollen production. Likewise, if data contradicts the assumption of similar adult density on all sides outside the plot, this data should be incorporated. For instance, if the plot is located at a forest edge, such that the only nearby adults would be to the south and east, it would be appropriate to consider only those directions as potential seed and pollen sources. The multinomial genotyping error probabilities can also be modified to allow for mistyping errors between alleles more than one repeat-length apart, although since this will increase the number of possible parent pairs this change greatly increases computation times.

The estimated seed dispersal parameter in this study (u_s = 6300, mean distance = 124.7 m) was considerably higher than that estimated using seed trap data (u_s = 34.9, mean distance = 9.28m)(Clark et al. In press). Some increase was expected, given that seed-trap data reflects the initial pattern of seedfall before animal dispersal, and undispersed seeds are unlikely to fall much beyond the maternal canopy, but the difference between the gravity-created seed shadow and effective dispersal distance at this site is striking. Some of the increase in average parent-offspring distance could be due to density-dependent mortality acting between germination and the time of sampling; nevertheless, the large number of seedlings found far beyond the maternal crown (>20 m) indicates the importance of animal-mediated dispersal for patterns of gene flow and seedling recruitment (**Chapter 3**). Distance-dependent mortality (due to adults harboring pests or pathogens) is likely to be of minor importance to the distribution of seedlings in this population: Seedlings exhibit high annual survival (**Chapter 4**) even though most are located within 14 m of an adult, and none are more than 60 m from an adult. My results suggest that oaks may be less dispersal limited than is commonly thought, with a greater potential for migration and gene flow via seed movement. They also suggest that in areas with reduced oak recruitment, red oaks are more likely to be limited by lack of suitable recruitment sites than by their ability to reach those sites (**Chapter 6**).

The pollen dispersal parameter converged at a value of 12900, corresponding to a mean effective dispersal distance of 178.2 m. Previous studies in *Quercus* species have also found high out-of-plot paternity – 50 to 70% (Dow and Ashley 1996, Streiff et al. 1999, Nakanishi et al. 2004). Because my plot was much larger than the area censused in previous studies, I expected that a greater proportion of seedlings would have in-plot fathers. Wind-blown pollen can travel extremely long distances, but because oak pollen degrades relatively quickly in UV light (Schueler et al. 2005) and can be washed out of the air by precipitation, and because nearby trees may produce large amounts of pollen, effective pollen dispersal can be much shorter than physical pollen transport distances (Ducousso et al. 1993). In some closed-canopy oak forests, short-distance matings appear to predominate (Fernandez et al. 2006), while some fragmented oak forests show evidence of pollen limitation (Knapp et al. 2001). However, even though my plot was 12 ha in size, 27.4% of seedlings were estimated to have out-of-plot fathers, while an additional 37% had neither parent among genotyped adults.

Several factors may account for the high amount of out-of-plot parentage for the population examined here. Because oaks are self-incompatible, nearby individuals, if they are closely related to the pollen recipient, could be at a disadvantage in siring seeds(Levin 1981). As there is no significant clustering of related adult trees at this site, shared incompatibility alleles are unlikely to exert a strong influence on pollination success (**Chapter 3**). Another possibility is that some true in-plot parents have died, leading to an over-estimate of the number of out-of-plot parents. Oak reproduction is episodic; no mast year occurred during the three years of this study, so acorn production was low and few new seedlings were produced. Many of the seedlings sampled were at least 5 years old, based on the number of winter bud scars. In future studies, it would be desirable to focus on newly-recruited seedlings following a mast year, before large changes in distribution occur due to mortality. Because oaks lack a seed bank, first-year seedlings can also be assumed to have living parents. It would also be possible to incorporate as potential parents in-plot individuals which died prior to the beginning of the study, but strong constraints would be needed to prevent these nearby-but-ungenotyped individuals from being identified as parents of all poorly-matched seedlings, perhaps by eliminating as potential offspring seedlings originating after the tree's death or by using genotypes of potential offspring to reconstruct the dead tree's genotype. Finally, it may be that there is simply a high rate of geneflow, mostly through pollen but sometimes by seed, which would be consistent with the low levels of landscape-scale genetic structure observed for many oak species (Schwarzmann and Gerhold 1991, Craft and Ashley 2007).

My parentage estimates suggest a high rate of hybridization; 14.2% of Duke Forest seedlings had a parent that was classed as a different species – for example, a "*Q. velutina*" mother assigned to a "*Q. rubra*" seedling. Heterospecific parent-offspring pairs do not necessarily represent true F1 hybridization events: alleles identical in length between species but not identical-by-descent, repeated back-crossing, or misclassification of young seedlings could also result in such matches. However, very similar allele frequencies between co-occuring red oak species suggest that interspecific gene flow is not uncommon in the Duke Forest. Gene flow across long distances and between species would be expected to boost local genetic variation but decrease differences between populations and species. The issue of hybridization is discussed further in **Chapter 5**.

Both long-distance gene flow and hybridization may make oak species less vulnerable than some plant species to loss of adaptive variation due to habitat fragmentation or the strong selective pressure imposed by rapid environmental change. Lack of an appropriate disturbance regime (Abrams 1992), high levels of herbivory (Pedersen and Wallis 2004), and introduced pathogens such as sudden oak death (Rizzo and Garbelotto 2003) pose more immediate threats, and may – through their effects on population size – lead to the erosion of genetic diversity. However, while pollen dispersal from southern populations or more heat-and-drought-tolerant congenerics could introduce useful variation to northern populations, high pollen flow from central populations could also inhibit local adaptation in marginal habitats (Kirkpatrick and Barton 1997, Rehfeldt et al. 1999, Lopez et al. 2007). Several conifer species have been found to exhibit strong local adaptation to climate despite extensive gene flow and recent population expansion (Mimura and Aitken In press) - the same may be true of oaks. Genetic resources for climate-related traits are currently less well-developed for *Quercus,* but due to efforts such as the Fagaceae genome project (http://www.fagaceae.org/) that may soon change. Understanding the distribution of variation in such traits, and the potential impact of dispersal and altered population connectivity on adaptive variation, will be critical to managing forests in an era of rapid environmental change.

Chapter 3. Contrasting Patterns of Dispersal and Gene Flow in Two Populations of Red Oak

Introduction

The spatial scale of effective seed and pollen dispersal – those dispersal events that result in germination and the establishment of a seedling - is important for both local dynamics and for species' potential to respond to climate change. The distribution of a tree species within a forest stand and across a landscape is affected both by the availability of sites suitable for its growth and survival and by its ability to reach those sites. Similarly, the distribution of genetic diversity depends on the scale of effective seed and pollen movement, and in turn affects the potential for local adaptation. Differences in the local environment can potentially lead to large differences in effective dispersal between populations. In order to predict the effect of different management strategies or of environmental perturbations such as climate change on forest trees, it is important to understand how dispersal patterns differ across a species' range, and what environmental factors drive these differences.

In this chapter, I estimate the scale of current seed and pollen movement in two mixed-species populations of red oak using the hierarchical Bayesian parentage and dispersal model developed for hermaphroditic plant species in **Chapter 2.** I also investigate whether differences in disperser abundance are associated with differences in seed dispersal scale, and whether the dense clumps of seedlings observed at one site represent siblings clustered near the mother tree or whether good recruitment conditions result in the clumping of unrelated individuals. In order to understand longer-term patterns of dispersal at these sites, I measure the strength of spatial autocorrelation in genotype indicating the degree of clumping of related individuals, also known as spatial genetic structure (Troupin et al. 2006). Finally, I consider the implications of these findings for population dynamics and responses to climate change in oaks.

Oaks: Conservation and management challenges

The declines in oak recruitment observed in many North American forests have been attributed to decreases in fire frequency and increases in deer density (Abrams 1992, McDonald et al. 2002, Aldrich et al. 2005b). However, seed limitation due to low fecundity or limited seed movement could also make reaching scarce recruitment sites more difficult (Crow 1988, Garcia and Houle 2005). Limited dispersal can also contribute to slow range shifts in response to climate change (Clark et al. 2001, Davis and Shaw 2001). Species that experience both reduced local reproduction and that are unable to disperse to newly favorable areas are vulnerable to the loss of genetic diversity and even local extinction. Genetic diversity and the potential for local adaptation might be maintained in oaks by contributions from wind-dispersed pollen or bird-dispersed acorns, but evidence for the extent of long-distance seed and pollen movement in oaks is conflicting (Johnson and Webb 1989, Dow and Ashley 1996, Sork et al. 2002, Fernandez et al. 2006, Fernandez and Sork 2007).

Interactions with dispersers

Differences in disperser density or activity between sites can have a strong impact on effective seed dispersal distances. Squirrels and blue jays are the most important dispersers of oaks in southeastern oak-hickory forests (Vander Wall 2001), as both species “scatterhoard” acorns in shallow 1-2 nut caches. Other rodents, such as mice and chipmunks, tend to “larder-hoard” in deep burrows or tree holes, from which germination would be difficult or impossible. Large animals can be effective dispersers for some large seeded trees, particularly those with fleshy fruits. However, deer, turkeys, and bears tend to destroy acorns, rather than swallowing them whole (Vander Wall 2001). Blue jays are known from previous studies to be present at both sites, and gray squirrels are also common. Informal observations during oak tissue collections suggested that squirrel densities at Duke Forest may be higher than at Coweeta. Clumping of seedlings downhill of adult trees at the latter site suggests that dispersal at Coweeta may be more limited. The density of seedlings within these clumps is quite high – as many as 7.5 per square meter.

Site history

The distribution of genetic variation across a landscape is affected by patterns of reproduction and mortality. Inbreeding and restricted seed dispersal distances tend to create positive autocorrelation in genotype (positive spatial genetic structure) at fine spatial scales, while long-distance seed and pollen dispersal and self-thinning tend to reduce this spatial genetic structure (Loveless and Hamrick 1984). Site history and dispersal processes also interact to shape spatial genetic structure (Jones et al. 2006, Troupin et al. 2006). Old-growth oak forests have been shown to exhibit spatial genetic structure that is positive at short distances and negative at long distances (Aldrich et al. 2005b), due to individuals recruiting near and mating with relatives over multiple generations. Recolonization of a disturbed site from distant seed sources leads to very low genetic structure in the first colonists, while matings between a limited number of colonists may lead to inbreeding and strong genetic structure in subsequent generations (Sezen et al. 2005). All available data on site history and past management practices was obtained from Duke Forest and the Coweeta LTER, respectively. I supplemented this information with all available tree-core records (41 trees at Duke forest, 37 at Coweeta) to infer the history of oak recruitment at each site.

At Duke Forest, much of the land was cleared for farms. These were later abandoned in the early 20th century, allowing the forest to regenerate. According to Duke Forest records, the eastern side of my plot was purchased by the university in 1945, and in 1952 consisted mostly of loblolly pine that had established in an old field prior to the site's purchase. Loblolly pine tends to be one of the first tree species to colonize abandoned agricultural fields in North Carolina (Oosting 1942). The age of pines at the Duke Forest site suggest an approximate abandonment date of 1912-1921. All oaks would have been cleared from this area when the field was established. The south-eastern corner, which contains more large hardwoods, may have been used as a selectively cut woodlot. The west side the plot was not purchased until 1991. Patches of early-successional vegetation (relatively young *Pinus* and *Liquidambar*) toward the north, south, and extreme west of this section indicate partial forest clearance, while the small sizes of oaks found in the center of the western half suggest some selective cutting. Tree-core data from red oaks show that one of the 41 sampled trees recruited in 1892, but most of the others were added between 1920 and 1955 (figure 3A), which is consistent with the expected pattern of succession and with the hypothesis that many of these trees recruited from dispersed seed.

At Coweeta, available information on site history was less detailed. Pre-historic management in the area included periodic use of fire, and early settlers are believed to have grazed livestock in the forest, both of which would have promoted a more open structure than is observed today. Logging was commonly practiced prior to the 1920's, but complete forest clearance was rare. Before becoming part of the national forest in the early 20th century, most of the stems >12 cm were cut from the area that includes my plot. The oldest trees in the sample of 37 from Coweeta recruited in 1835 and 1870, which would have made them approximately 65 and 30 years old at the time of the selective harvest in the early 1900's, suggesting that some adult trees may have been retained. Such trees may have served as seed sources and promoted the creation or retention of small-scale genetic structure observed. Most of the others were added post-harvest, between 1905 and 1940 (figure 3B). The spike in recruitment during the 1930's may be connected to the removal of competition from American Chestnut due to chestnut blight (Keever 1953, Stephenson 1986). Fire has been excluded for the past century, and the site has been largely undisturbed since the 1920's. Simulated wind-throw gaps were created at both sites in the year 2000.

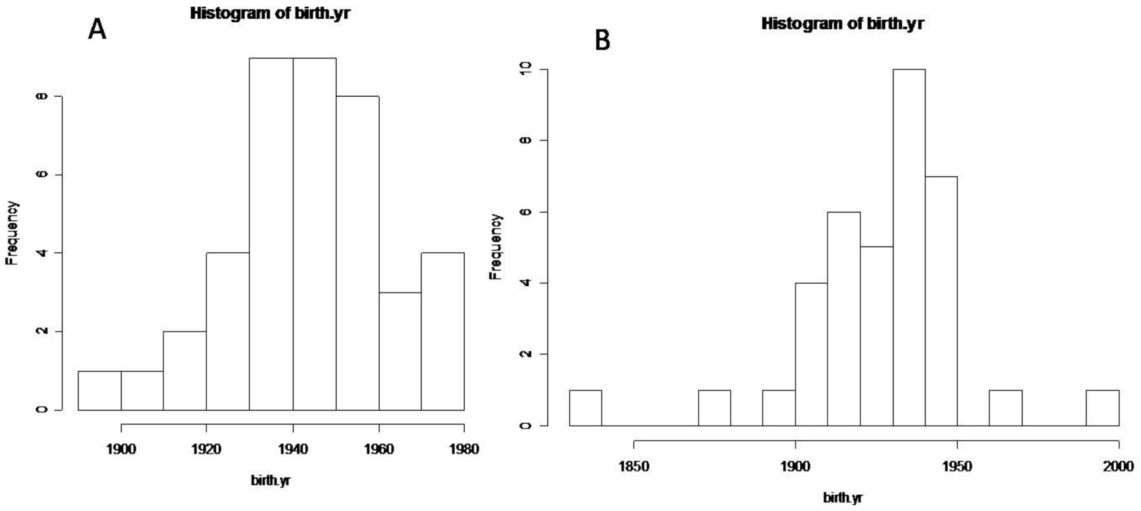

Figure 3: Year of recruitment for trees at Duke Forest (A) and Coweeta (B).

Hypotheses

Seed dispersal: I hypothesize that higher squirrel densities will be confirmed at Duke Forest, and that this site will exhibit greater effective seed dispersal distances than Coweeta. Further, I expect that the dense clumps of seedlings observed at Coweeta downhill from large adults will be found to consist primarily of closely-related individuals originating from those nearby trees. However, if these clumps consist mostly of unrelated seedlings, this would suggest that these dense aggregations are the result of highly favorable recruitment conditions rather than limited dispersal.

Pollen dispersal: Because oaks are wind-pollinated, I hypothesize that pollen dispersal distances will be similarly high at both sites, and that many seedlings will have fathers located outside the mapped stands.

Historical gene flow and spatial genetic structure: Oaks currently in the canopy at Duke Forest likely originated from scattered seed sources outside the formerly cleared area. I would therefore expect to see very little spatial genetic structure in the adult trees. Because more local seed sources were retained at Coweeta, I expect this population to exhibit stronger spatial genetic structure. Because self-thinning gradually reduces the clumping of related individuals, I expect seedlings to exhibit more spatial genetic structure than adults. Low genetic structure in both age classes would suggest that gene flow is extensive within a site – and has been so for at least two generations.

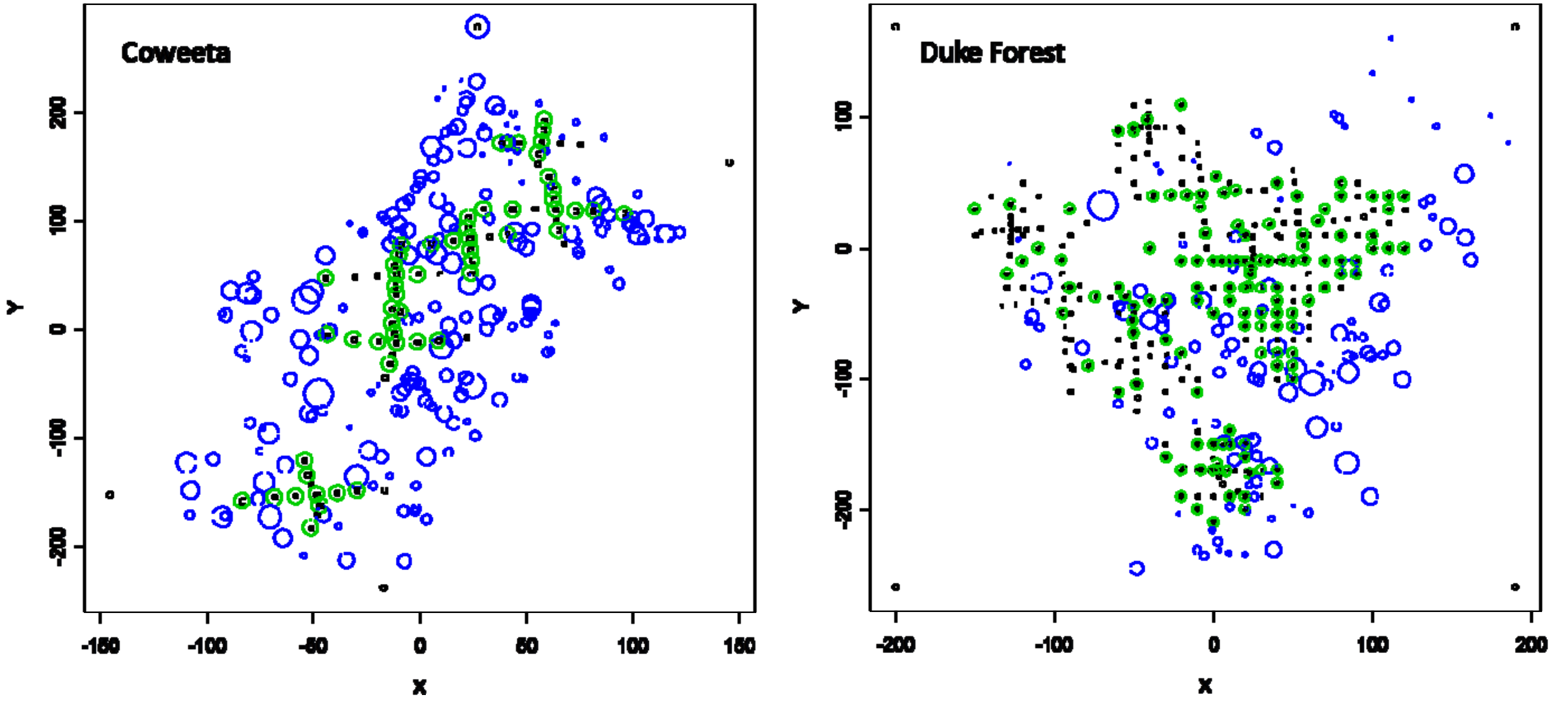

Figure 4: Maps of Coweeta and Duke Forest sites.

Blue circles – adult trees, size proportional to diameter. Green circles – seedlings. Black squares – seedling census plots, size proportional to plot area.

Methods

Dispersal and parentage analysis:

In this analysis, I make use of data from mixed-species populations of red oak (*Q. rubra*, *Q. velutina*, *Q. falcata*, and *Q. coccinea*) at Duke Forest and Coweeta (figure 4), as described in **Chapter 1**, and apply to them the parentage and dispersal model developed in **Chapter 2**. A 2D-t "dispersal kernel" (function describing the probability of travelling a given distance *d* is employed for both seed and pollen. The 2D-t is a fat-tailed kernel, its height and width determined by parameter *u*:

$$p(d) = \frac{1}{\pi u \left(1 + \frac{d^2}{u}\right)^2}$$

Because genetic differentiation between species within a site was found to be very low (**Chapter 5**), all three species at each site are included in the analysis as potential parents and offspring.

Disperser density:

Distance-sampling based on walking fixed transects is an effective and cost-efficient method of estimating density for species such as grey squirrel (Hein 1997). I established a series of transects at each site (eight, with a total length of 540 m, at Coweeta; nine, with a total length of 740 m, at the larger Duke Forest Site). On each sampling date, an observer walked these transects in the morning and evening, scanning ground and trees for squirrels or blue jays, and recording the perpendicular distance of any animals observed from the transect. Following a pilot survey at Duke Forest in

October 2009, subsequent survey dates were chosen so that environmental conditions (temperature, phenology) would be similar at the two sites. Surveys were conducted after leaf-fall, and visibility exceeded 20-50 m. All surveys occurred when squirrels are expected to be most active: early and late in the day during and after the period of oak and hickory seedfall (Hein 1997). Estimates of squirrel density based on distance-sampling data were calculated using the program DISTANCE 6 (Thomas et al. 2010). Three functions for decay of detection probability were compared: uniform/cosine - the recommended omnibus model (Buckland et al. 1993), half-normal/hermite-polynomial, and uniform/simple-polynomial. AIC was used to discriminate between models.

Spatial genetic structure analysis:

As a complement to my parentage and dispersal analyses, I examined microsatellite data using the program Genalex (Smouse and Peakall 1999). This program, developed specifically for multi-locus multi-allele data, calculates a genetic correlation coefficient *r* for individuals within a series of distance classes, and constructs confidence intervals using a bootstrap approach. Adults and seedlings at both sites were analyzed separately. This program also allows calculation of pairwise relatedness, which was used to investigate the relationships within and between seedling plots. Relatedness is also, confusingly, known as r – I will, therefore, refer to it hereafter as R_s. Siblings share, on average, half of their genome. Their expected R_s is 0.5, while the expected R_s for half-sibs is 0.25, and for 1st cousins 0.125. I calculate pair-wise relatedness for all seedlings at both sites, and examine the relationship between R_s and distance.

Results

Parentage and dispersal

As described in **Chapter 1**, MCMC was used to estimate the posterior distributions for parentage and dispersal parameters. Because pedigree is inferred, I refer to the most frequently identified parent pair (after model convergence) as the “estimated parents” for each seedling. Simulations indicate that, on average, true mothers and fathers are identified 86% of the time, while in a further 11% of seedlings the correct parent pair is identified but mother-father identities are inverted.

At Duke Forest, 35 seedlings out of the 219 sampled (16%) were estimated to have both parents within the plot, 43 (19.6%) to have an out-of-plot mother, 60 (27.4%) to have an out-of-plot father, and 81 (37%) to have neither parent in the plot. At Coweeta, 41 seedlings out of the 179 sampled (22.9%) were estimated to have two in-plot parents, 32 (17.9%) to have an out-of-plot mother, 92 (51.1%) to have an out-of-plot father, and 14 (7.8%) to have two out-of-plot parents.

At Duke Forest, the estimate for the seed dispersal parameter u_s is 6300 (se 460), corresponding to a mean dispersal distance of 124.7 meters. At Coweeta, the seed dispersal parameter was estimated to be 92 (se 19), corresponding to a mean dispersal distance of 15 m – only slightly longer than that estimated from seed trap data (u = 34.9, mean distance=9.27 m)(Clark et al. In press). These estimates take into account both distances to in-plot parents and the possible distance between seedlings and unobserved out-of-plot parents, which are inferred as part of the model. Estimates of the pollen dispersal parameter u_p were high for both sites: 12,900 at Duke Forest (se 510), corresponding to a mean dispersal distance of 178.2 m, and 8,600 at Coweeta (se 580), corresponding to a mean dispersal distance of 145.7 m. Due to the scale of pollen dispersal, these latter values are somewhat crude: recall that 59-64% of seedlings are estimated to have fathers outside the mapped study area. This greater uncertainty is reflected in wider credible intervals around the mean estimate. However, study sites were sufficiently large to capture a substantial number of fathers. The distribution of distances between seedlings and in-plot mothers, and between in-plot mothers and fathers, are shown for Coweeta and Duke Forest in figures 5 and 6, respectively.

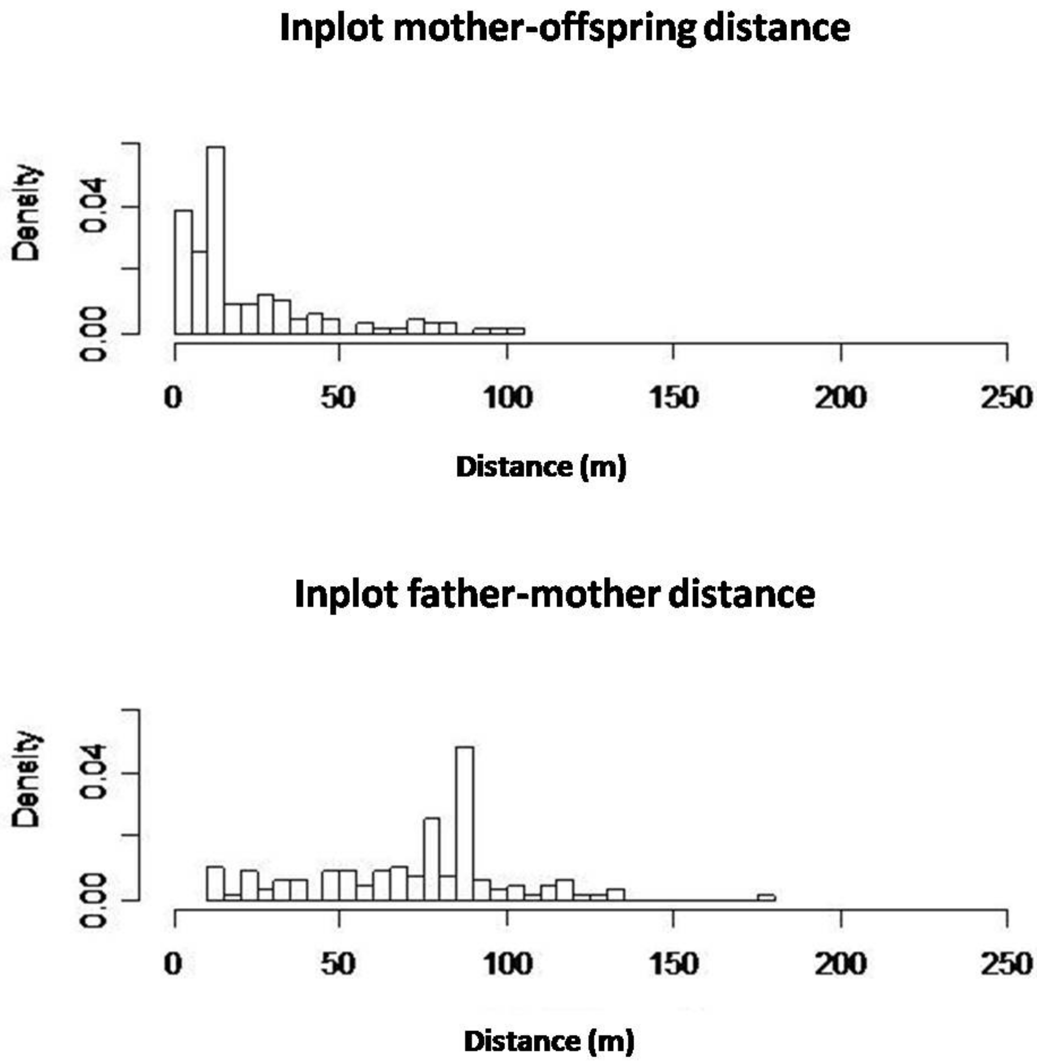

Figure 5: Density of in-plot mother-offspring (a) and father-mother (b) distances at Coweeta

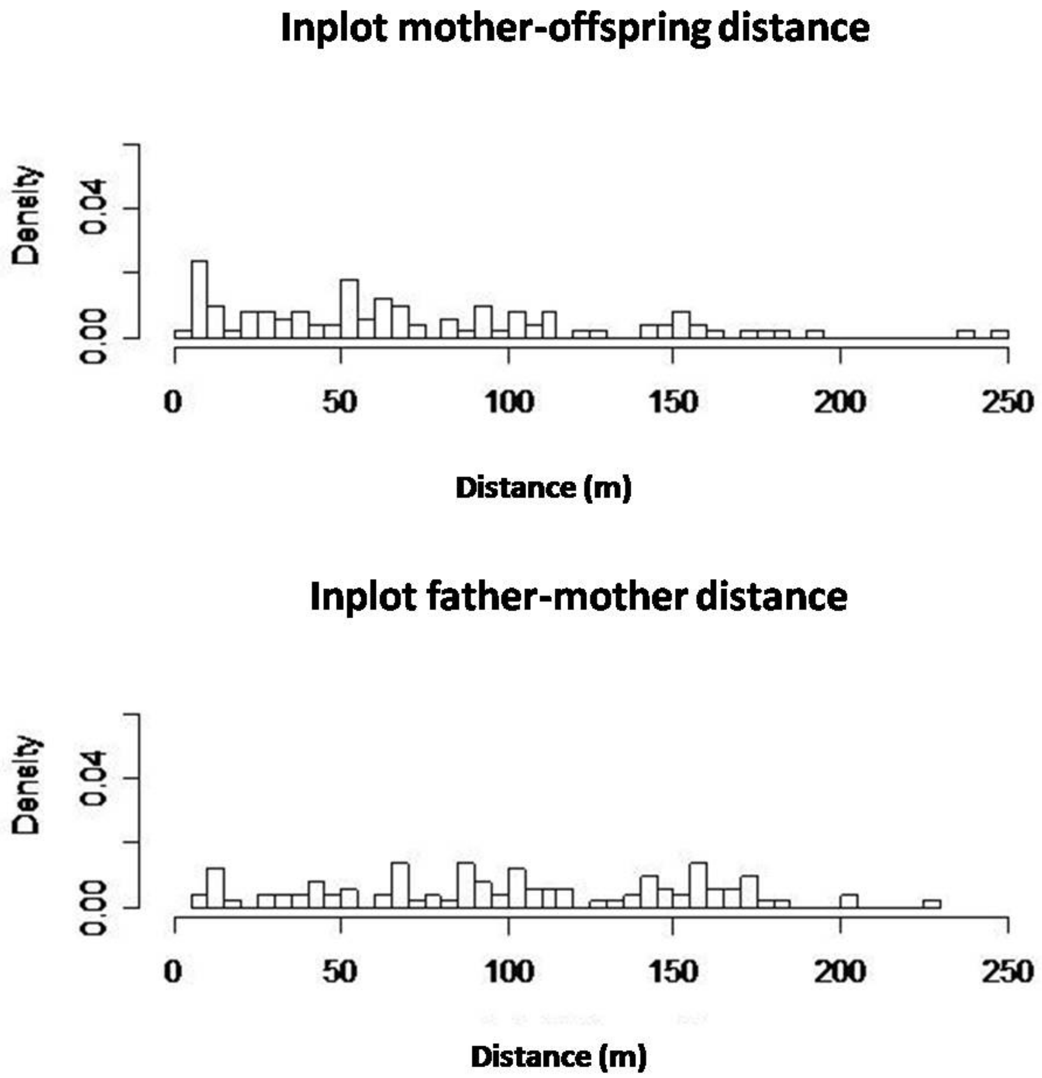

Figure 6: Density of in-plot mother-offspring (a) and father-mother (b) distances at Duke Forest.

Disperser density and activity:

Although longer dispersal distances were observed at Duke Forest, contrary to my predictions disperser activity/density does not appear to be substantially higher at this site. The number of squirrels encountered per hour of observation was similar at the two sites, while the number encountered per meter of transect was actually higher at Coweeta (table 3). No blue-jay activity was observed on the survey days. However, previous studies indicate that jays are common at both sites. Breeding bird surveys conducted at Coweeta revealed blue jay average densities of 7.25 pairs/km^2 in 1993 and 21.6 pairs/km^2 in 1967 in four undisturbed habitats in which oaks are commonly found: oak-hickory, pitch pine-oak, cove hardwood, and northern hardwood forest (Tramer 1993). Repeated surveys conducted in the Blackwood division of Duke Forest reveal that blue jays are present year-round (www.duke.edu/~jspippen/birds/dukeforestsurvey.htm). Assuming that jay calls can be detected to ~200 m from the survey transect (J. Pippen, pers. comm.), the average density of jays in the Duke Forest is ~4.46 birds/km^2. However, as non-calling birds may not be detected this may be something of an underestimate. Chipmunks were common at Coweeta but, as they seldom function as effective dispersers, they were not counted.

Table 3: Squirrel counts

	Date	Time of day	Distance (m)	Total obs time (min)	#squirrels	Squirrels per km	Squirrels per hour
Coweeta	11/06/09	7:19 – 9:13	541	82	4	7.39	2.93
	11/06/09	14:51-16:11	541	61	3	5.55	2.95
	11/07/09	7:15-8:36	541	61	2	3.69	1.97
	11/07/09	9:06-10:14	541	54	3	5.55	3.33
	11/07/09	15:00-16:19	541	61	2	3.69	1.97
	Mean					**5.17**	**2.63**
Duke Forest	*(pilot) 10/21/09*	*8:02-10:01*	*750*	*77*	*7*	*9.33*	*5.45*
	12/01/09	7:37-9:08	740	67	2	2.70	1.79
	12/01/09	9:53-11:16	740	66	3	4.05	2.73
	12/01/09	15:13-16:31	740	65	2	2.70	1.85
	12/03/09	7:38-8:54	740	66	1	1.35	0.91
	12/03/09	14:57-16:11	740	64	4	5.41	3.75
	mean (without pilot)					**4.26 (3.24)**	**2.75 (2.2)**

The recommended uniform/cosine decay function was favored by AIC at both sites (though at Coweeta the uniform/simple-polynomial performed as well). Although the best-model estimate of squirrel density at Coweeta was lower than at Duke Forest (43.7 vs 64.2 squirrels/km^2), confidence intervals overlapped considerably (table 4). As some squirrels near the transect may have gone undetected because they were hidden by dense brush or resting in tree holes, these estimates should be regarded as conservative.

Table 4: DISTANCE results

Site	**Distance function**	**AIC**	**Density (squirrels/km^2)**	**Upper-Lower CI**
Coweeta	Uniform/cosine	104.19	**43.7**	**28.6 - 66.8**
	Half-normal/Hermite	105.06	59.6	31.7-111.0
	Uniform/Polynomial	104.10	**43.7**	**28.6 – 66.8**
Duke Forest	Uniform/cosine	147.4	**64.17**	**31.7-129.9**
	Half-normal/Hermite	149.01	58.2	28.7-117.8
	Uniform/Polynomial	148.66	42.7	1.4-85.1

Spatial genetic structure

For Duke Forest adults, there was no observed autocorrelation in genotype for any distance class, indicating that relatives were not clustered (figure 7 A). By contrast, adults at Coweeta exhibited significant positive spatial structure at < 20 m, indicating that related individuals (with similar genotypes) were more likely to be found within this distance (figure 7 B). Seedlings at both sites exhibited spatial genetic structure at short distances: < 10 m at Coweeta, and < 20 m at Duke Forest (figure 7 C&D).

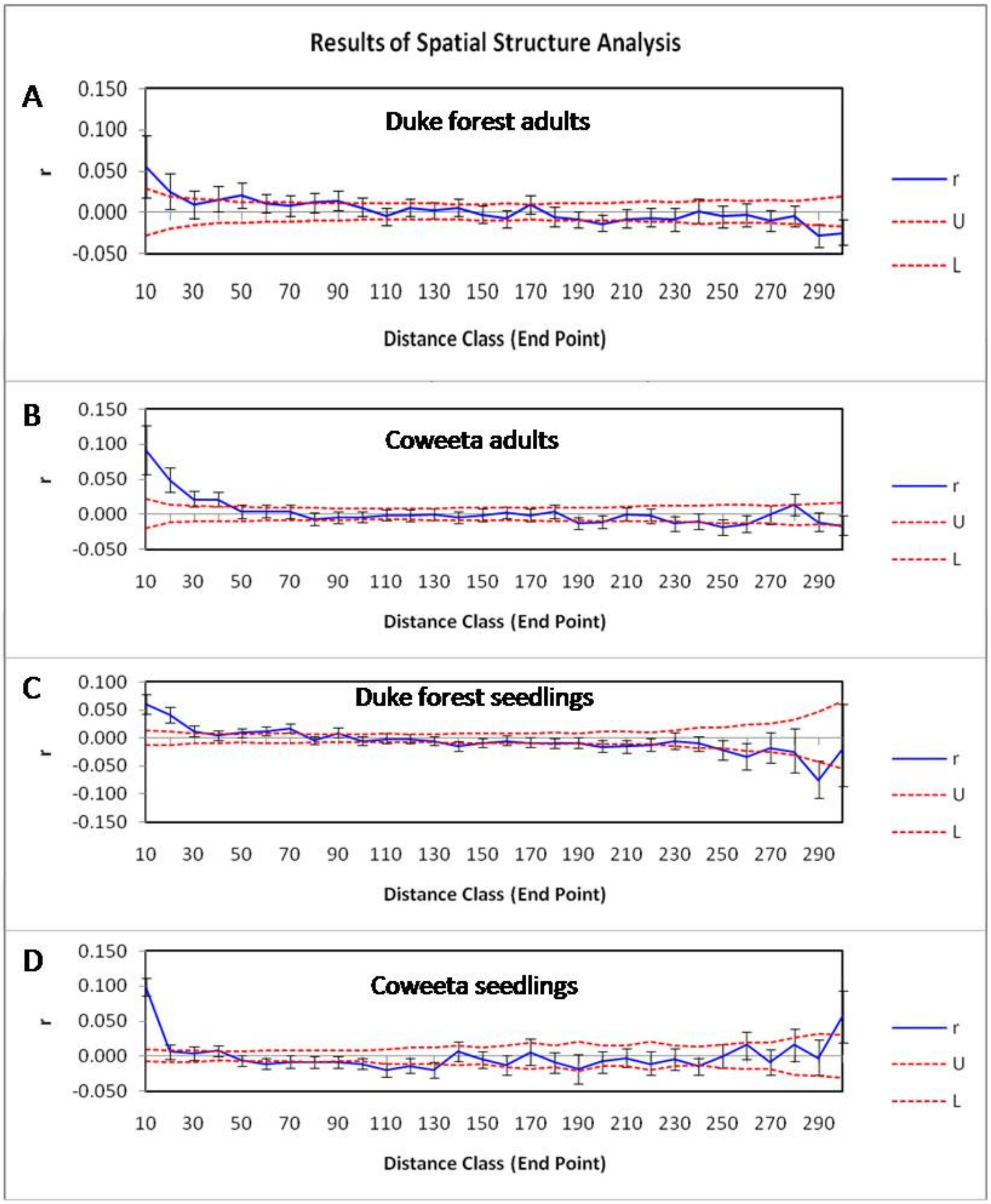

Figure 7: Spatial genetic structure

Blue lines – estimated genetic correlation coefficient r; red dashed lines 95% bootstrapped confidence intervals around zero (999 reps).

At Coweeta, closely related seedlings tended to be found near one another – often within the same plot. The median distance between seedlings at this site was 81.2 m. Out of 15,931 pairwise comparisons, 78 seedling pairs (0.5%) were related at the sib-to-half-sib level (Rs > 0.25), and the median distance between such individuals was 32.7 m; the large proportion occurring in the 0-10 m distance class reflects clumps of sibs and half sibs within the same or neighboring plots (figure 8A). The median distance between the 483 pairs (3%) related at the half-sib-to-cousin level ($0.125 < R_s < 0.25$) was 54.3 m. The proportion of pairs with any given degree of relatedness dropped off steeply with distance. At Duke Forest, on the other hand, related seedlings were distributed over a wider area (figure 8B). The median distance between individuals where Rs > 0.25 (0.6% of 23,872 total pairs) was 90.1 m, and was 95.1 m for the 3.6% where $0.125 < R_s < 0.25$, compared to an overall median inter-seedling distance of 110.4 m. This is consistent with the longer seed dispersal distances estimated for Duke Forest.

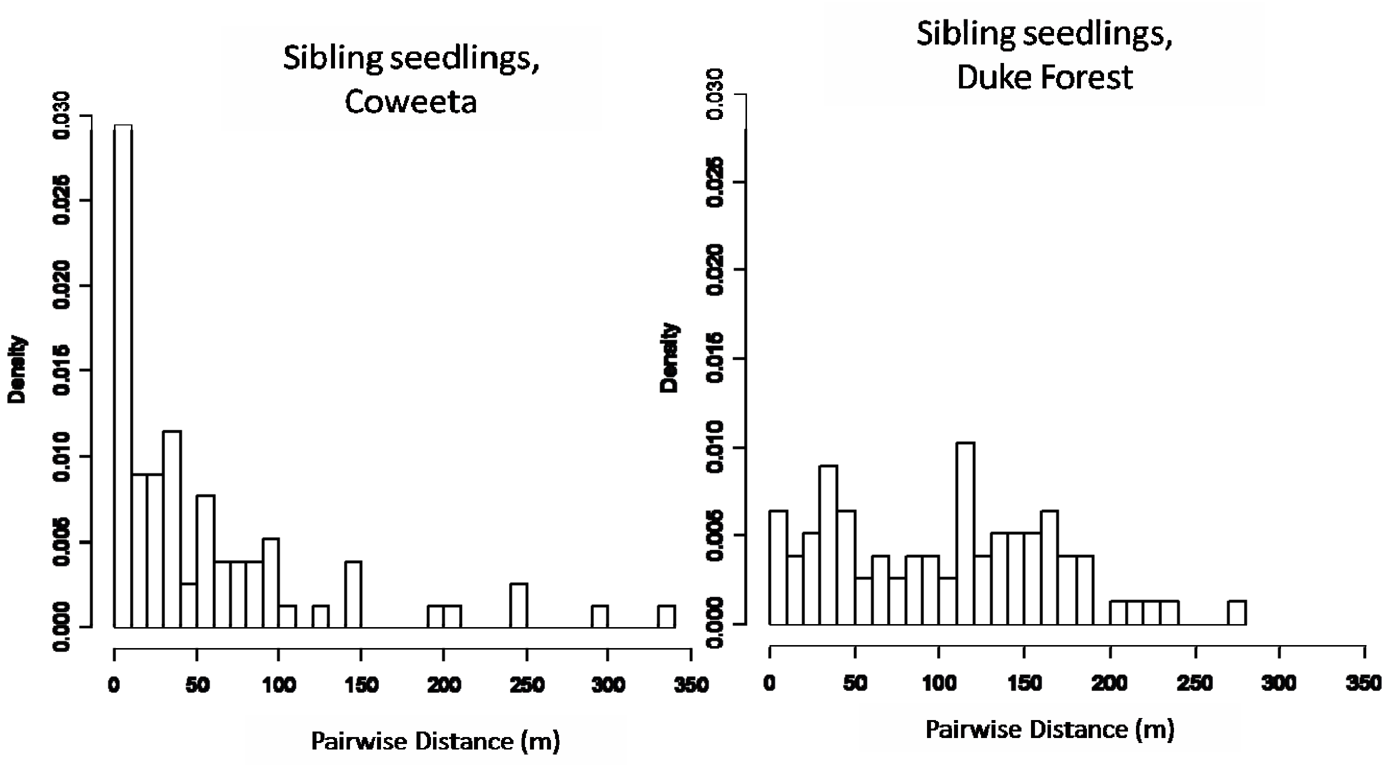

Figure 8: Spatial distribution of seedlings related at the sib-to-half-sib level ($R_s > 0.25$) at Coweeta (left) and Duke Forest (right)

Discussion

Parentage analyses reveal that the oak populations at the Duke Forest and Coweeta sites differ greatly in the scale of effective seed dispersal. As hypothesized, the clumps of seedlings observed at Coweeta were found to consist primarily of sib and half-sib seedlings. Contrary to my expectations, the difference in effective seed dispersal could not be explained by differences in disperser density/activity, as the number of squirrels observed over a given time period under similar environmental conditions, and the estimated density of squirrels, were similar at both sites. Blue jays are suspected to be important agents of long-distance acorn dispersal (Johnson et al. 1997), so higher jay density or activity could lead to substantial increases in dispersal. However, no jays were observed on the survey days, and earlier bird counts suggest that this species is found in similar numbers at both sites.

A number of other mechanisms may be hypothesized to explain the observed differences in effective seed dispersal. First, high deer densities at Duke Forest (**Chapter 1**) could result in the destruction of the (easily visible) clumps of seed or seedlings near-parent trees, leading to higher mean effective dispersal distances through increased mortality at short distances. Density-dependent insect predators or fungal pathogens could have a similar effect (Janzen 1970). However, a high proportion of seedlings at Duke Forest were >100 m from their mothers while few offspring traveled such distances at Coweeta. Such a pattern cannot be explained by density-dependent mechanisms. The steep topography at Coweeta might discourage long-distance transport of acorns by squirrels, but the agility of these animals makes this unlikely. Finally, the history of mast years could produce different patterns of seedling establishment. Seed production and seedling recruitment was quite low during the three years of the study, and many of the sampled seedlings are estimated to be > 5 years old. The estimated dispersal kernels therefore integrate over multiple years of recruitment and survival. It is possible that a heavy mast year at Coweeta which satiated acorn predators and dispersers (Moore et al. 2007) could have produced the observed pattern. Either way, animal-dispersal appears to strongly influence effective dispersal at Duke Forest, while recruitment at Coweeta seems to be more dominated by gravity.

The observation that sites differ widely in effective seed dispersal could have important consequences for both population dynamics and the distribution of genetic diversity. However, detailed population genetic studies of trees rarely include multiple sites (Burczyk et al. 2004), and attempts are often made to extrapolate from genetic or non-genetic single-site dispersal estimates to landscape-level processes. A few other studies have included parentage analyses of two heavily-sampled sites (Schnabel et al. 1998), while others have sampled less intensively at a landscape scale (Craft and Ashley 2007, Fernandez and Sork 2007), but there is necessarily a tradeoff between sampling within and between populations. Genetic information can add greatly to our understanding of cryptic reproductive processes, and it is to be hoped that reductions in genotyping costs will soon enable more cross-site comparisons. Until then, extrapolations from genetic studies at a single site must, like those based on purely ecological data, be approached with caution. The ability of oaks to reach suitable recruitment sites and to shift their range in response to climate change will depend on whether more populations behave like Duke Forest, where seed dispersal distances are long, or like Coweeta, where offspring cluster close to their parents.

Both sites exhibit high genetic diversity at neutral markers for all age classes (14-35 alleles per locus). Long pollen dispersal distances, such as those estimated for both Duke Forest and Coweeta, are believed to help maintain the high genetic diversity and the low levels of isolation-by-distance observed in many oak species (Ducousso et al. 1993, Nakanishi et al. 2004, Germein and Kesseli 2006, Craft and Ashley 2007). Wind dispersed pollen, as well as occasional long-distance transport of acorns, may help to maintain genetic diversity and adaptive potential of oak populations even in the face of strong selective pressures such as climate change – at least in continuous forest. Oak pollen degrades quickly in UV light (Schueler et al. 2005), and reductions in oak density due to fragmentation (Knapp et al. 2001, Fernandez et al. 2006) or low recruitment (Sork et al. 2002) have been shown to result in pollen limitation and reduced allelic diversity in seedlings.

While many studies of spatial genetic structure in trees have found a pattern of significantly positive spatial structure at short distances which becomes negative at long distances (Streiff et al. 1998, Asuka et al. 2004, Aldrich et al. 2005b), I observed no positive or negative correlation at distance classes above 30 m for any site or age class. The weak spatial genetic structure for both populations is consistent with extensive gene flow over several generations (Loveless and Hamrick 1984, Sato et al. 2006). These patterns of spatial genetic structure are also consistent with the history of these sites. At the time many of the current canopy dominants were establishing at Duke Forest, there were few large red oaks remaining nearby as seed sources. Mixing of multiple distant and local seed sources would be expected to produce extremely low levels of genetic structure in this first generation, and that is what we observe. At Coweeta, though there was some selective harvesting, individuals <12 cm in diameter were not cut and some adult trees appear to have been retained, which would have preserved much of the original genetic structure of the site. Although we observe less spatial genetic structure in the adult trees at Coweeta than Aldrich et al. (2005b) found in their old-growth oak forest, it was still greater at Coweeta than at Duke Forest. Both sites are undergoing secondary succession, and it is possible that the clustering of related genotypes will become more evident as local seed sources continue to dominate over multiple generations and the opportunity for inbreeding increases (Troupin et al. 2006).

My results indicate that at least some oak populations are less dispersal limited than previously thought, suggesting that regeneration and population expansion may be best promoted by reducing barriers to survival such as unfavorable disturbance frequencies and high deer densities. Results of analyses using the SLIP forest simulator (**Chapter 6**) show that, under competitive conditions, longer dispersal distances increase the probability of population persistence even in the absence of large-scale disturbances. This supports the hypothesis that long-distance seed dispersal, such as that observed at Duke Forest, enables oaks to reach scarce recruitment sites frequently enough that the abundance of such sites, rather than seed limitation, becomes the most important factor in the dynamics of these populations. Dispersal estimates from the Duke Forest population yield average annual migration rates much higher than gravity dispersal alone. In competitive conditions, Duke Forest estimates did allow population persistence and spread while seed-trap based estimates did not, though spread rates were significantly lower (**Chapter 6**). Populations similar to Coweeta would fall in between these extremes.

These results suggest that dispersal limitation may be less of a constraint on the response of oaks to climate change and other environmental perturbations than it might appear. Long-distance seed movement (in some populations) can allow oaks to reach even scarce recruitment sites, and to colonize newly suitable areas, while long-distance pollen movement helps to maintain genetic diversity and adaptive potential. However, in other populations environmental factors may restrict the possibility for LDD. Their vulnerability will increase where populations are fragmented or where population densities decrease due to predator, drought, or competition. Recruitment failures in oaks due to competition from faster-growing or more shade-tolerant species are well documented (Larsen 1953, McDonald et al. 2002, Aldrich et al. 2005a), while high deer densities have reduced seedling recruitment in many forest species (Rooney 2001, Pedersen and Wallis 2004). In addition, severe drought, which is predicted to become more common in the southeastern US under climate change (Ibanez et al. 2008), can make adult red oaks more vulnerable to pathogens and other stresses (Oak et al. 1986, Clinton et al. 1993). The loss of oaks as a dominant taxon in eastern deciduous forests would have negative consequences for a variety of wildlife species, including squirrels, turkeys, and deer, which depend on acorns as a winter food source (Johnson et al. 1995, McShea et al. 2006). To successfully reach the canopy, red oaks seem to require light from canopy disturbances, as well as periodic reductions in understory competition from low-intensity fires or other ground-level disturbances (Buckley et al. 1998, Elliott et al. 1999, Gribko et al. 2002). Careful management for disturbance frequency and herbivore density may be needed to promote both local population persistence and resilience to environmental change in oaks.

Chapter 4: Inequality of reproductive success in forest trees - its causes and consequences

Introduction

In trees, reproductive success is generally assessed in terms of seed production (for example, Fuchs et al. 2003, Schoen & Stewart 1986) yet, in evolutionary or population-dynamics terms, an individual cannot be said to have successfully reproduced until its offspring are themselves of reproductive age. The degree of inequality in reproductive success has important implications for ecological and evolutionary dynamics. In particular, unequal reproductive success reduces effective population size (*Ne*), affecting the potential strength of natural selection and the probability that genetic variation will be lost to random drift. Most methods for calculating *Ne* assume that reproductive success is normally distributed, yet data from a variety of plant and animal species suggests that reproductive success in most organisms is highly skewed, with a few individuals producing a large proportion of the offspring and the majority of individuals contributing few or none (Clark et al. 2001, Gonzalez-Martinez et al. 2006, Herrera and Jovani 2010). In this chapter, I make use of genetic and long-term ecological data to estimate offspring in the seedling class (hereafter reproductive success, R) and average annual percent germination for individual trees in two mixed-species populations of red oak (*Q. rubra*, *Q. velutina, Q. falcata*, and *Q. coccinea)* in North Carolina, taking into account uncertainty in annual seed production. I then examine the distribution of R and test whether individual reproductive success (R_i) and average germination fraction (Q_i) are positively associated with annual seed production, age, size, or growth rate. Finally, I discuss the implications of the distribution of reproductive success for effective population size and the conservation of genetic diversity.

Variation in individual reproductive success is vital to evolutionary dynamics. The mechanism of natural selection requires that some phenotypes (and their associated genotypes) survive and reproduce at a higher rate than others (Darwin 1859). In addition, variation in reproductive success itself can influence the effectiveness of natural selection. The effective population size, N_e, is the size of an ideal population (random mating, equal reproductive output, constant size) that would show the same genetic properties as the population under consideration. Effective population size can differ markedly from actual population size if some of the individuals are immature or senescent (overlapping generations), if a few individuals contribute a disproportionately to the allele frequencies of the next generation (unequal reproductive success), as well as in dioecious versus monoecious species (Gillespie 2004). This quantity is important for the evolutionary potential of a species or population because the rate at which genetic diversity increases due to mutation is proportional to effective population size, and the rate at which alleles are lost to random drift is inversely proportional to N_e. Selection will overcome drift if $2N_es >> 1$, where s is the strength of selection, while genetic drift is expected to dominate when $2N_es << 1$. Many different (sometimes incompatible) ways of calculating N_e have been developed (Emigh and Pollak 1979, Eguiarte et al. 1993, Nunney and Elam 1994, Dodd and Silvertown 2000), but none specifically account for species in which fecundity is highly variable, mating is distance-dependent, *and* generations overlap extensively, as is the case for most tree species.

Calculations of N_e generally begin with the assumption that adults contribute equally to the next generation, with modifications allowing for departures from this assumption. Three of these equations are given in table 5, where N is the actual population size, μ and σ^2 are the mean and variance in reproductive success, T is generation time, and F_{is} is the inbreeding coefficient for the population (based on observed versus expected heterozygosity). Even when parameters are set so that variation in reproductive success is the only variable that determines N_e (one year to maturity, no inbreeding, average R = 1), they yield quite different estimates.

Table 5: Three methods of calculating the effect of variation in R on N_e

For **N = 100**, T = 1, F_{is} = 0, μ = 1, $\boldsymbol{\sigma^2}$ **= 3**…		
Reference	**Equation**	**Ne estimate**
Gillespie (2004)	$N_e = \frac{N}{\sigma^2}$	33.3
Heywood (1986)	$N_e = \frac{N}{(1+F_{is})(\sigma^2/\mu^2)+1}$	25
Hill (1972)	$N_e = \frac{4(N-2)T}{\sigma^2+2}$	78.4

Not only are results inconsistent, relying on just the mean and variance of offspring number assumes that reproductive success follows a distribution that is well described by its variance, such as normal or Poisson distribution. However, most natural populations of both plants and animals exhibit highly skewed distributions of reproductive success

(Meagher and Thompson 1987, Clutton-Brock 1988, Schnabel et al. 1998, Gonzalez-Martinez et al. 2006, Herrera and Jovani 2010).

It has been proposed that both fecundity and lifetime reproductive success follow a lognormal distribution due to the multiplicative effect of interacting environmental factors (Herrera and Jovani 2010). Even if many individuals do not produce seed or established offspring, this may be true conditional on reproductive success > 0. That is, given that an individual reaches reproductive maturity and produces at least one offspring, reproductive success might be lognormal, while in the population as a whole there could be an excess of zeroes resulting from a) individuals that produce no seed or b) individuals that produce seed but have no surviving offspring.

Given that reproductive success is likely to be complex, it may not be well characterized by a simple index such as variance. More general measures of dispersion include the Gini coefficient (G). In the context of reproductive success, the Gini coefficient is the ratio of the area between the equality line and the cumulative distribution function of adult contributions to the seedling pool known as the Lorenz curve (area A in figure 9) and the total area under the equality line (A+B). Thus,

$$G = 1 - 2\int_0^1 L(x)dx$$

where L(x) is the Lorenz curve. The Gini coefficient is most often used to describe economic inequality, but has also been used to measure species diversity (Weber 1997), plant architecture (shoot size distributions)(Larson and Whitham 1997), and plant size hierarchies (Weiner and Solbrig 1984).

When all adults have equal reproductive success G=0 (thin line), while if one adult produces all offspring (thick line) G = 1 (thick line). Populations with an intermediate degree of inequality will fall somewhere between these curves. For instance, in the case where there are 33 adults, and 16 produce no offspring, 10 produce 1% each, 4 produce 5% each, and the last 3 produce 10%, 20%, and 40%, respectively, we obtain the dashed line in figure 9 and a Gini coefficient of 0.82. A simple relationship between G and expected changes in allele frequency, if it exists, would allow a more sensible calculation of N_e for populations with varying degrees of inequality in reproductive success.

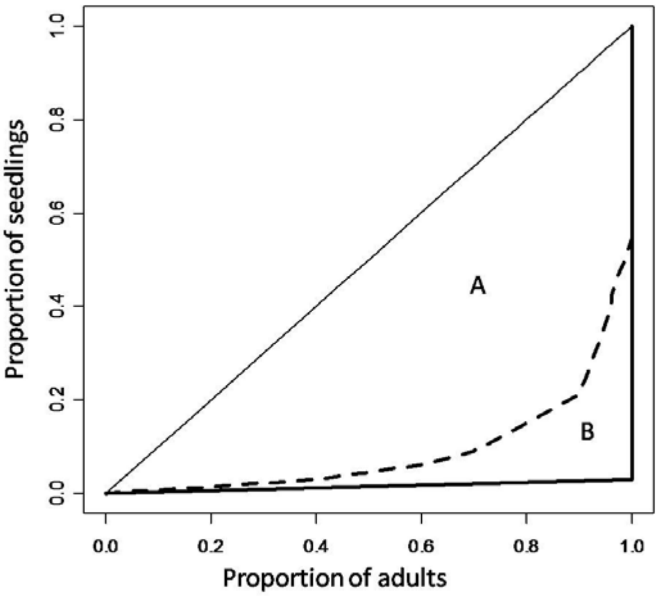

Figure 9: Calculation of the Gini coefficient

Thin line – G=0; thick line – G=1; dashed line – G =0.82

Because most tree species are long-lived and produce annually variable (but often large) numbers of seeds that may take years or decades to reach maturity, it is virtually impossible to count how many adult individuals a particular tree contributes to the population. However, with modern genetic methods, it is possible to identify the parents of tree seedlings (**Chapter 2**). Mortality at the seed-to-seedling transition is extremely high in most tree species; by estimating the number of juveniles produced by individual trees in relation to their seed production, age, vigor, or other characteristics, we can gain a better understanding of what factors contribute to reproductive success, and of the true effective population size.

Surprisingly, despite increasing interest in the use of genetic markers to reconstruct pedigrees in woody plants (Dow and Ashley 1996, Godoy and Jordano 2001, Asuka et al. 2005, Hardesty et al. 2006), few studies have used pedigree information to examine variation in reproductive success. Schnabel et al. (1998) estimated maternity of honeylocust (*Gleditsia triacanthos*) seedlings at two sites based on a maximum-likelihood analysis of allozyme data, but this calculation was based on sampled seedlings, not *total* seedling production for individual mothers, nor were estimates of seed production available for comparison. Gonzalez-Martinez et al (2006) estimated relative reproductive success in maritime pine (*Pinus pinaster*) using the Seedling Neighbourhood Model, in which the relative reproductive success of the *j*-th female in the 50-m neighborhood of the *i*-th juvenile is linked to covariates (including inter-plant distance) through a log-linear model. Relative sapling production in these pines was positively associated with diameter and number of cones. However, because the relationship between these characteristics and seed production was not taken into account in the sapling analyses, it is not clear whether a seed from a large tree has a higher chance of establishment, or whether large trees simply produce more seed, increasing the probability that at least one will survive to the seedling stage.

Reproductive success is likely to be positively associated with fecundity (defined hereafter as annual seed production). However, density-dependent mortality (Janzen 1970, Connell 1978) could reduce the fraction of seeds that germinate from highly fecund mothers. If this is the case, then density-dependent mortality would tend to equalize reproductive output. Age of the mother is likely to be associated with reproductive success, because old individuals have been producing seed for more years than young individuals, allowing more opportunities for seedling establishment. Diameter is sometimes used as a proxy for age (Schnabel et al. 1998), though this should be approached with caution due to individual and site differences in growth (McMahon et al. 2009). Data from tree-cores is more reliable. More vigorous trees may have more resources available for provisioning seed, which could lead to higher establishment success (Gonzalez-Martinez et al. 2006). On the other hand, if there are tradeoffs between growth and seed production or seed provisioning (Clark et al. In press), maternal growth rate might be negatively correlated with seedling productions.

Here, I test several hypotheses using data from my two red oak populations. First, I hypothesize that reproductive success will be highly skewed, and that annual seed production and the number of seedlings R will follow a lognormal distribution with the addition of a large zero class. I hypothesize that trees with high annual fecundity will also exhibit high reproductive success, but that highly fecund trees will also have a lower average germination fraction. I also expect that age, diameter, and increment growth, will also be associated with high reproductive success. Finally, using simulations, I explore the relationship between inequality in reproductive success, as measured by G, and effective population size.

Methods

At each site, parentage of sampled seedlings and seed and pollen dispersal parameters were estimated using a hierarchical Bayesian model incorporating genetic and ecological data, and accounting for uncertainty in genotype and fecundity, as described in **Chapter 2** and **Chapter 3**. In this analysis of reproductive success, I make use of estimates of the dispersal parameter and parentage based on the mean of the posterior distributions of these quantities. I refer to the total number of seedlings originating from tree *i*, and existing in the seedling bank at the time of the parentage analysis, as R_i, and the average annual germination fraction as Q_i. This quantity represents the fraction of seeds produced in a year that become first year seedlings, and is equal to the average number of first year seedlings divided by the average annual seed production. Individual differences in germination fraction may result from genetic differences, environmental differences affecting seed provisioning, or differences in density-dependent mortality resulting from differences in seed production.

Total current number of seedlings R_i:

I assume that seedling recruitment is conditionally Poisson, such that the likelihood of observing s_{ij} seedlings from mother i in plot j is

$$L = \prod_j \frac{e^{-r_{ij}} r_{ij}^{s_{ij}}}{s_{ij}!} \qquad\qquad r_{ij} = K_{ij} A_j R_i$$

where K_{ij} is the probability of a seed dispersing from mother i to plot j (expressed per m^2), A_j is the area of the plot, and R_i is the total number of seedlings produced by mother i (the quantity we wish to estimate). For a 2D-t dispersal kernel, K_{ij} is given by

$$\frac{1}{\pi u \left(1 + \frac{d_{ij}{}^2}{u}\right)^2}$$

where d_{ij} is the distance between the tree and the plot and u is the seed dispersal parameter (Clark et al. 1999).

If the dispersal parameter and the parentage of each seedling are assumed to be known, the maximum-likelihood estimate of R_i is:

$$\hat{R}_i = \frac{\overline{s_{ij}}}{\overline{K_{ij} A_j}}$$

This distribution is degenerate for parents having no observed seedlings in any plot. Therefore, in the analyses that follow, I focus in individuals with at least one observed offspring. The standard error of the estimate can be approximated from Fisher Information

$$\hat{\sigma}_R = \frac{\hat{R}_i}{\sqrt{\sum_j s_{ij}}}$$

Average annual germination fraction Q_i:

In many mid-to-late-successional tree species, such as oaks, seedlings may remain in the seedling bank for years. In previous experiments at the same study sites in North Carolina, over 50% of first-year *Q. rubra* and *Q. falcata* seedlings survived beyond the 3.5 year study (Ibanez et al. 2008). The seedling pool at any given time therefore contains an accumulation of cohorts. I provide an approximation of annual seedling production from which Q_i can be calculated, recognizing that it is only a rough index for a mean value.

If the number of first year seedlings for tree *i*, Y_i, is constant, then seedlings accumulate up to the current time 0 from some remote time in the past so that

$$R_i = \sum_{t=-\infty}^{0} Y_i S(t)$$

where *S(t)* is a function describing survival from one year to the next, and where the number of seedlings at time *t* is equal to the number of new recruits plus surviving recruits from the previous years. Of course, recruitment varies considerably over time, especially in masting species such as oaks. Y_i simply represents an average number of new offspring over many years. First year seedlings often experience higher mortality than older seedlings so, assuming constant annual mortality after the first year

$$R_i = \sum_{t=1}^{\infty} Y_i S_{new} S_{old}^{t-1}$$

We can solve for Y_i using a geometric series

$$Y_i = \frac{R_i}{S_{new} \sum_{t=1}^{\infty} S_{old}^{t-1}} = \frac{R_i (S_{old} - 1)}{S_{new} (S_{old}^{t} - 1)}$$

and, as *t* grows large,

$$Y_i = \frac{R_i (1 - S_{old})}{S_{new}}$$

The average annual seedling production divided by average annual fecundity F_i yields an estimate of average germination fraction

$$Q_i = \frac{Y_i}{F_i}$$

Over the four years of this study (2006-2009), individual seedlings in permanent census plots were tagged and tracked over time. At Duke Forest, 13 first-year seedlings were identified in 2006, 17 in 2007, and none in 2008. Eighty-five percent of these seedlings survived until their second year. For seedlings >1 year old, the average annual survival probability was 95.9%. At Coweeta, only 6 first-year seedlings were identified between 2006 and 2008. Four survived until their second year, for a first-year survival probability of 66.7%. For older seedlings, the average annual survival probability was 98.2%.

Distribution of reproductive success

To assess the inequality of reproductive success, I calculated the percentage of potential parent trees contributing to the seedling pool as mothers and as fathers, as well as the Gini coefficients for both populations. I then examined the distribution of annual fecundity for all trees within the original mapped stand and fit lognormal distributions to a) average annual fecundity for trees estimated to have a >50% chance of being mature, and b) R_i for individuals having at least one offspring. I chose to focus on trees within the original mapped stand for this analysis because, as long-term records of diameter and seed-trap data are available, we have much more confidence in their fecundity estimates.

Factors contributing to reproductive success

I hypothesized that fecundity would be positively correlated with reproductive success but negatively correlated with germination fraction. Fecundity has been estimated for all trees within the original mapped stands (Clark et al. In press). Parameter estimates derived from this analysis allow us to estimate the probability of maturity (θ) and annual fecundity F_t given maturity for the border trees as well:

$$\theta = \text{inverse logit } (\alpha_0 + \alpha_1 D + \alpha_2 C)$$

$$\ln(F_t|\text{mature}) = \beta_0 + \ln(D_t)\ \beta_1 + \ln(D_t\text{\^{}}2)\ \beta_2 + \ln(C_t)\ \beta_3 + \ln(I_{t-1})\ \beta_4 + \varepsilon$$

where D is diameter, C is canopy area, I is growth increment, and ε is an error term based on individual-, year-, and site-level variability.

I also hypothesized that older and larger trees should have more offspring in the seedling pool. Exact ages are only known for individuals from which increment cores have been taken – 34 of the 44 potential parents within the original mapped stand at Duke Forest, 34 of the 94 potential parents within the original mapped stand at Coweeta. Annual growth increments have been estimated for trees within the original mapped stand (44 at Duke Forest, 94 at Coweeta). Diameter measurements, on the other hand, have been taken for all potential parent trees, both in the original mapped stand and in the additional border area. In the interest of consistency, I will use diameter-at-breast-height measurements for 2006, the year in which the additional border area was surveyed.

Annual growth increment has also been estimated for all trees within the original mapped stands (Clark et al. In press). I hypothesize that if high resource availability enables investment in both growth and reproduction, then the relationship between growth rate and reproductive success should be positive, while if there is a tradeoff between growth and reproduction then the relationship between growth rate and reproductive success should be negative.

Effective population size

To address the issue of effective population size, I first compare the proportion of trees in the population estimated to be parents to the expected N_e based on overlapping generations. Nunney and Elam (1994) calculated that for a population with overlapping generations (and equal reproductive success),

$$N_e = \frac{N}{(2 - T^{-1})}$$

where N is the (apparent) number of adults and T is the generation time. When T is large, as in forest trees, effective population size approaches 50% of observed N.

I then explore the relationship between G and N_e using simulations. One way that N_e can be estimated is through its effect on heterozygosity (H). A population in which every individual had two different alleles at the locus of interest would have H =1; one in which there was only one allele at the locus would have H = 0. The change in heterozygosity from one generation to the next, in the absence of mutation, is expected to be

$$H_1 - H_0 = -\frac{H_0}{2N_e}$$

where H_0 and H_1 are the heterozygosities of the parental and offspring generations. Therefore,

$$N_e = \frac{H_0}{2(H_0 - H_1)} = \frac{1 - \sum_I p_j^2}{2\left(\left(1 - \sum_I p_j^2\right) - \left(1 - \sum_I p_j'^2\right)\right)}$$

where p_j is the frequency of allele j in the parental population, and p'_j is the frequency of allele j in the offspring generation.

I simulate changes in allele frequencies over one generation for a population of a given size N with lognormal reproductive success, focusing on one locus with two alleles. The alleles are initially at equal frequency

$$p_1 = p_2 = 0.5$$

Individuals in the first generation are assigned two-allele genotypes according to a binomial probability distribution. Reproductive success follows a lognormal distribution

$$f(x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}$$

If $\mu = \sigma^2 = 1$, then reproductive success is equal to 1 for all adults. In the simulation, μ is set to one, so that the population size remains constant (on average) from one generation to the next even when reproductive success varies. The Gini coefficient and average N_e is calculated for a range σ^2's between 1 and 10.

Adults are randomly assigned a number of offspring based on the lognormal distribution of reproductive success, and a value for G is calculated. Each individual passes on one of its two alleles (with equal probability) to each of its offspring. This process is repeated 200 times for each σ^2 and N_e is calculated based on the average change in heterozygosity. These values for effective population size are then plotted against the average G for each value of σ^2.

Results

Distribution of reproductive success

As predicted, reproductive success was highly unequal. At Duke Forest 45 of 118 potential parents (38.1%) were estimated to be the mother of at least one seedling, while 50% were estimated to be either a mother or a father. At Coweeta, reproductive success was even more unequally distributed: 40 of 199 potential parents (20.1%) were estimated to be mothers of at least one seedling. Of these, one tree was estimated to be the mother of 14% of all sampled seedlings. Sixty-five (32.7%) were estimated to be a mother or father.

For trees with identified offspring, distributions of R_i were also skewed. At Coweeta, some trees located far from the sampling plots were assigned very high values of R_i (> 4000) and very high standard errors (>3000) by the maximum likelihood approach (Figure 10 B). These individuals I considered to be outliers, and they were removed from subsequent analyses. This effect was not seen at Duke Forest, where seedling sampling plots were more widely and evenly distributed. In figure 10, as in all subsequent figures, vertical bars represent 2 standard deviations to each side of the ML estimate of R_i.

The distribution of R_i and average annual fecundity for mature trees, along with the lognormal curves fitted to these values, are shown in figures 11 and 13. As I hypothesized, the lognormal distribution provides a reasonable fit both for fecundity given maturity and for offspring number given maturity and at least one offspring. Also as predicted, for the population as a whole there is a large zero class (figure 12). In other words, the number of adults producing no offspring is large.

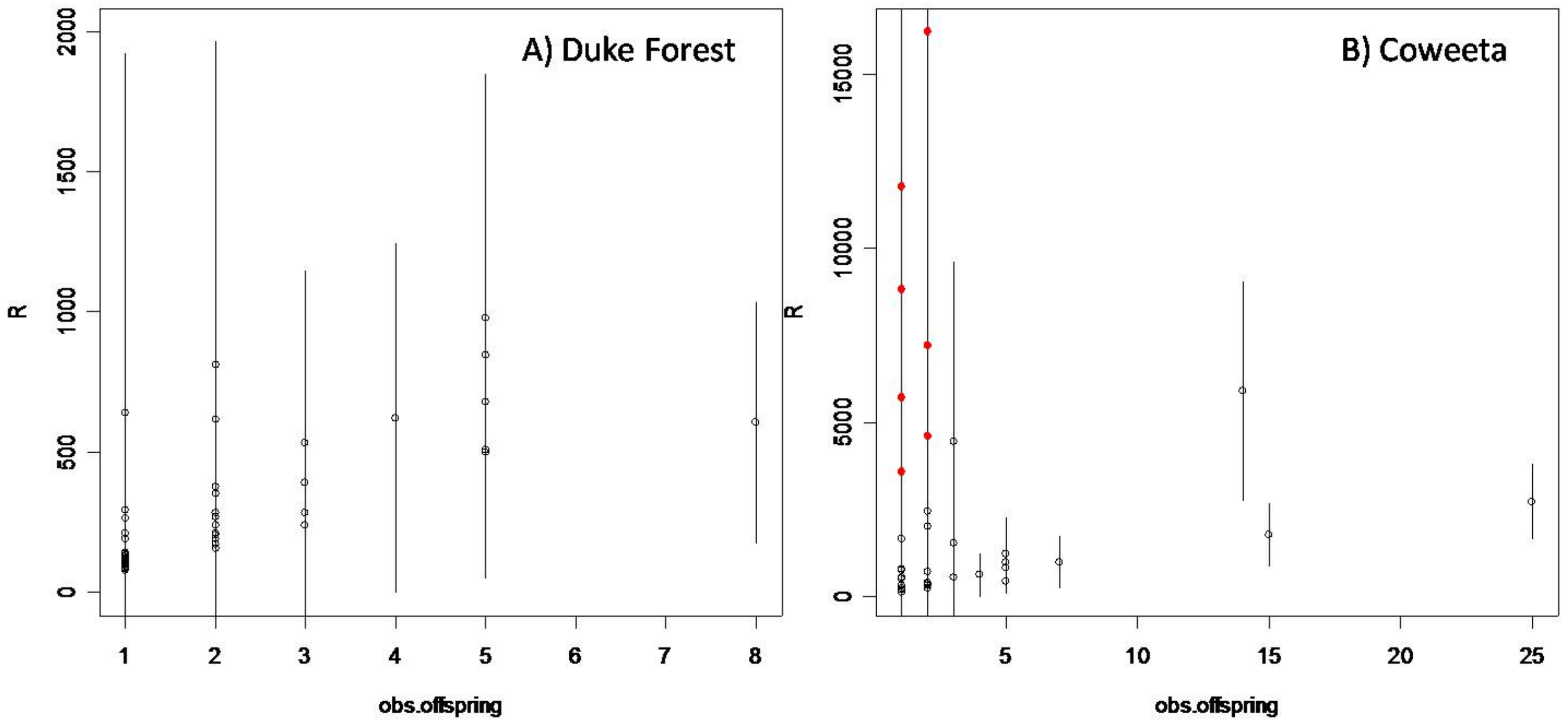

Figure 10: Estimated number of seedlings, Ri, vs. observed number of seedlings at Duke Forest (A) and Coweeta (B). Coweeta outliers (red) removed in subsequent analyses

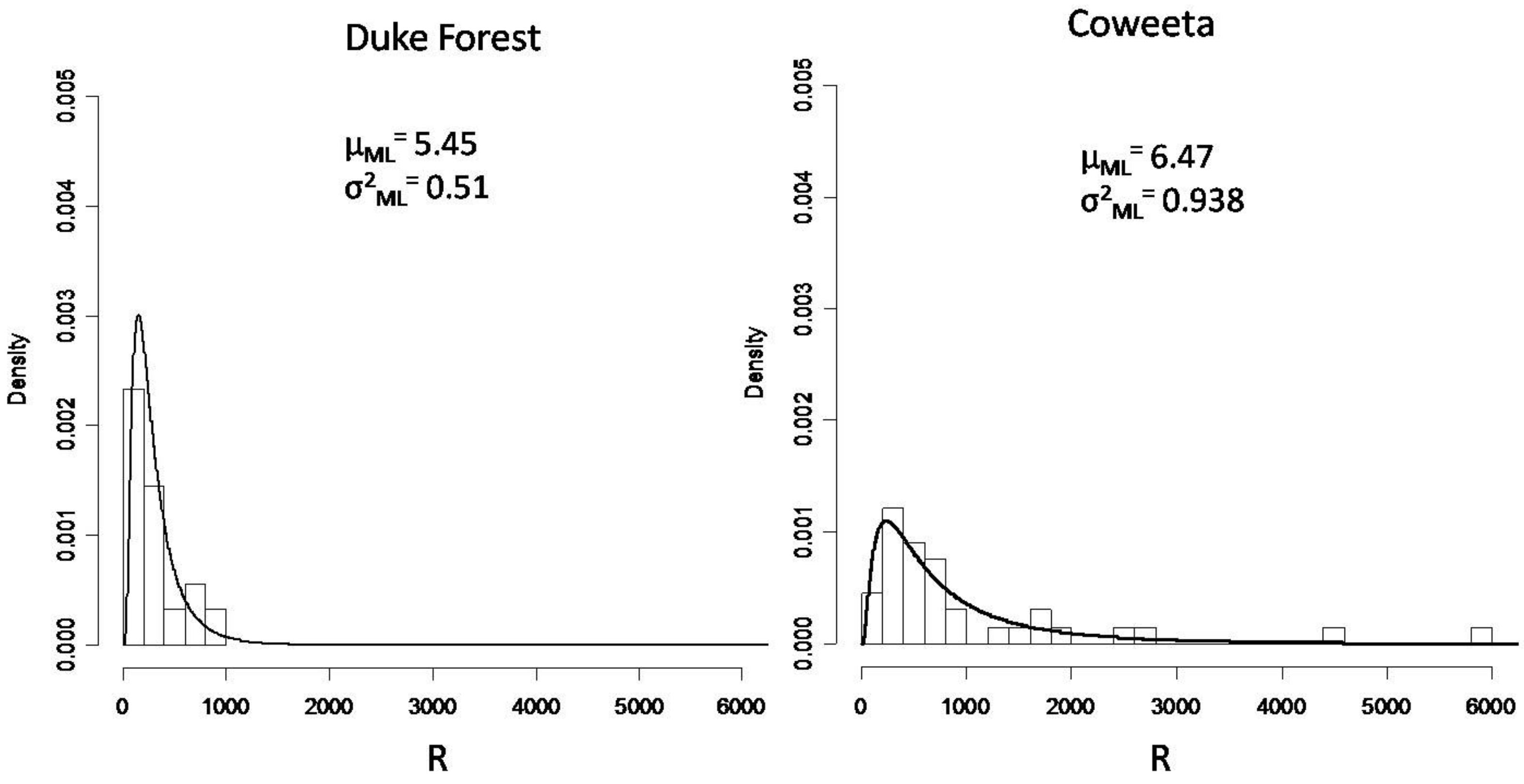

Figure 11: Ri at Duke Forest (left) and Coweeta (right), with fitted lognormal distributions.

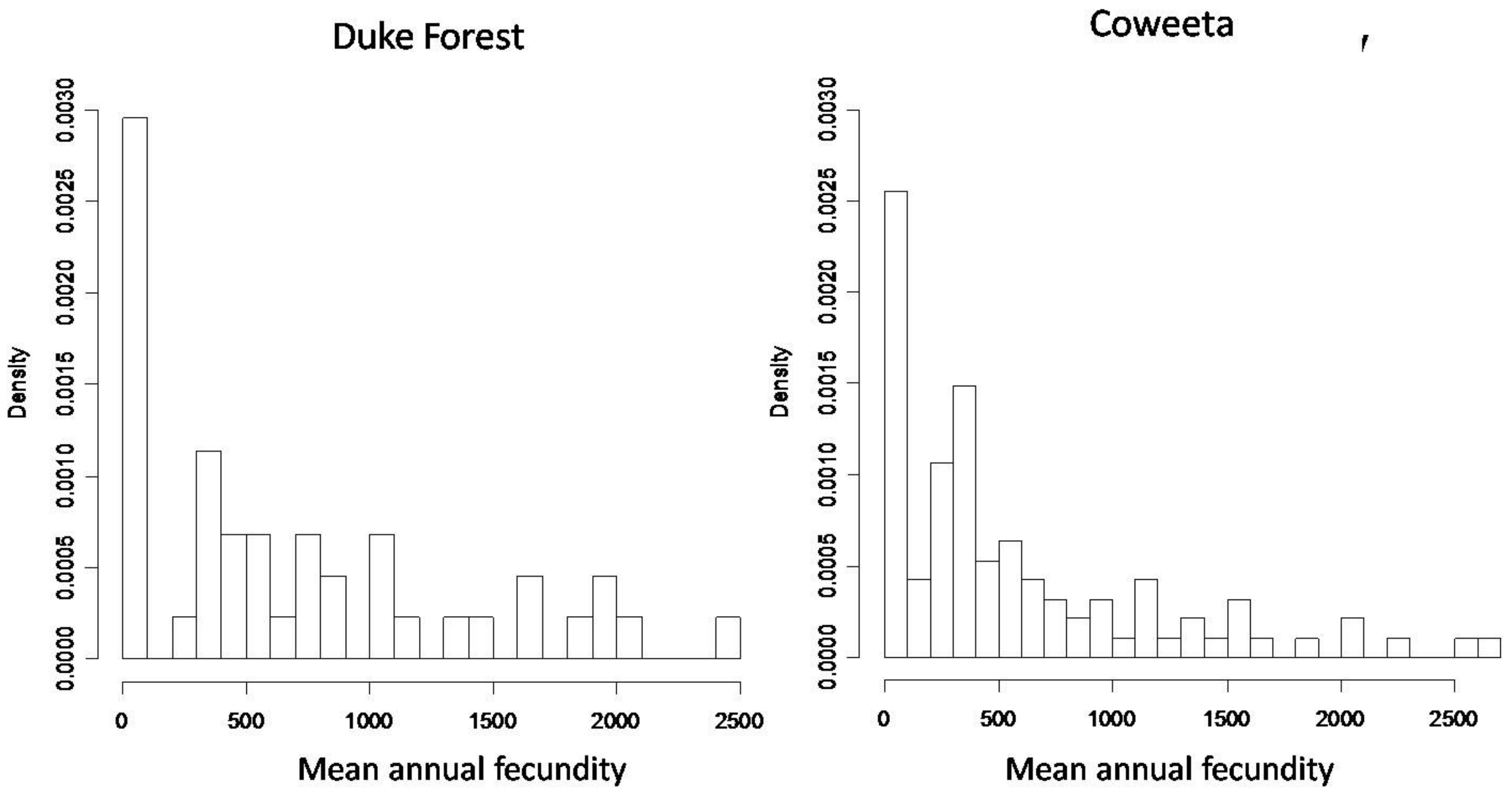

Figure 12: Fecundity for all trees in original mapped stand at Duke Forest and Coweeta. Note excess of zeroes and very small values, which creates a bimodal pattern.

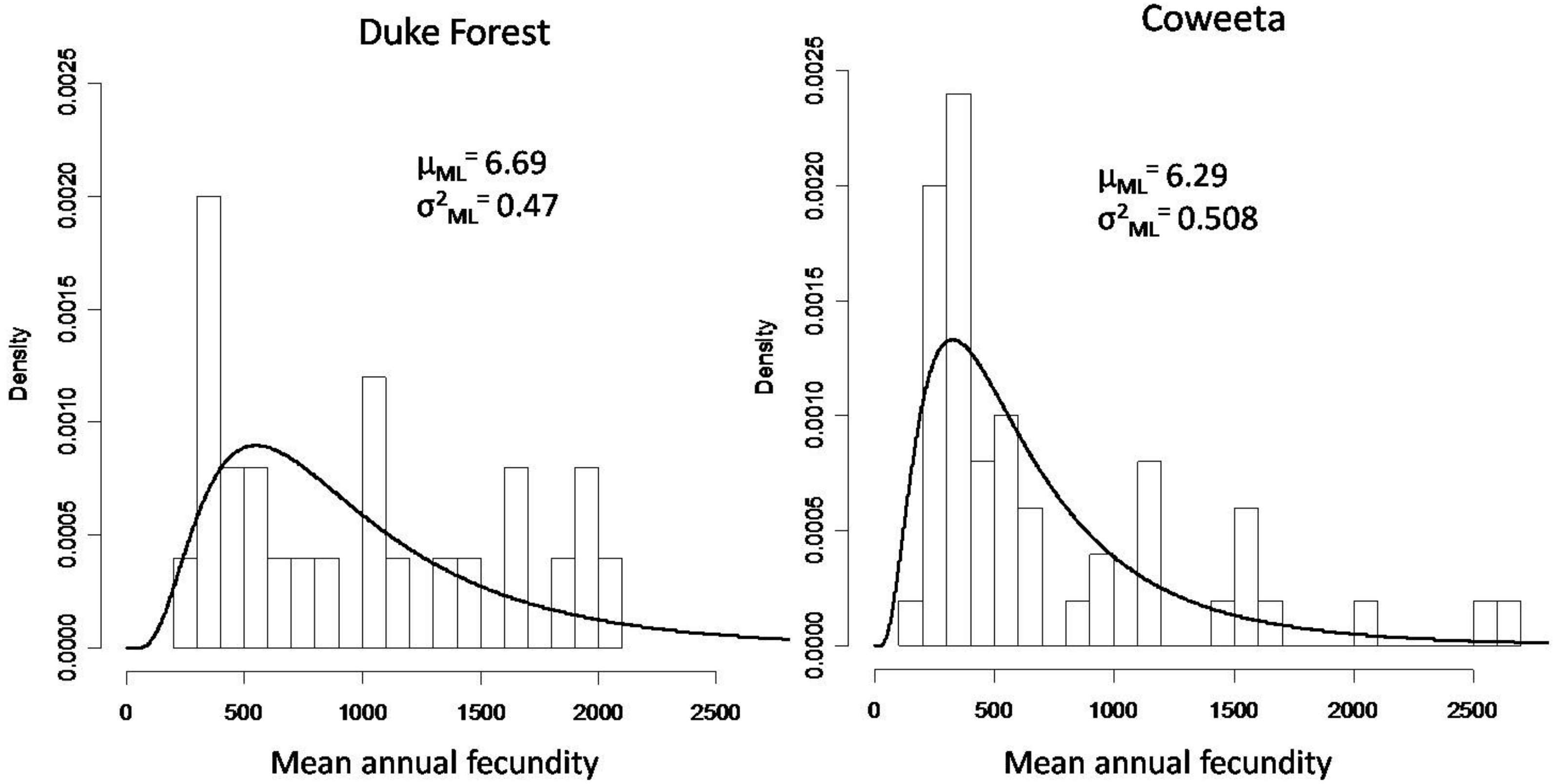

Figure 13: Fecundity for trees with > 50% probability of maturity at Duke Forest (left) and Coweeta (right), with fitted lognormal distributions.

The Gini coefficient was 0.398 for Duke Forest and 0.527 for Coweeta. The associated Lorenz curves can be seen in figure 14 (dashed line – Duke Forest, thick solid line – Coweeta).

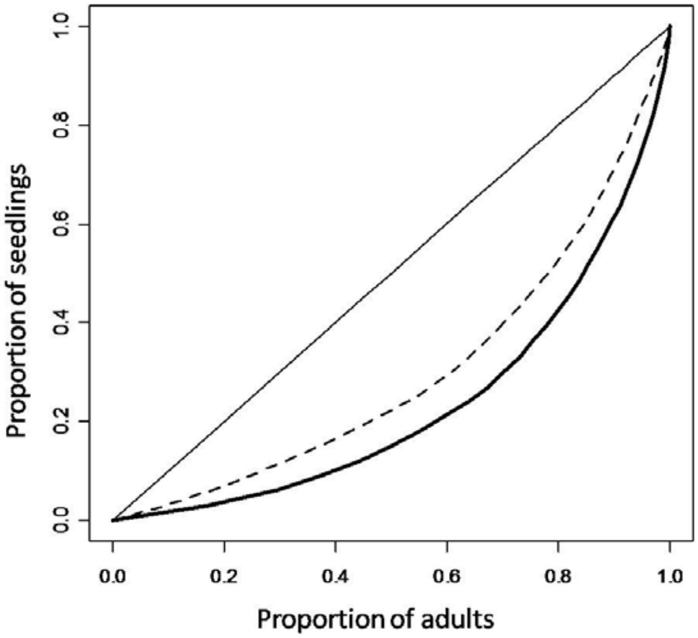

Figure 14: Cumulative density functions of seedling production (given > 0 observed seedlings) for Duke Forest (dashed line) and Coweeta (thick line)

Age and growth Rate

Contrary to my predictions, R_i did not exhibit a linear relationship with either age or growth rate, but rather a hump-shaped pattern, suggesting that intermediate values are associated with higher reproductive success. I therefore performed a quadratic regression in both cases for both sites (figures 15 & 16). There was no relationship between age and growth increment.

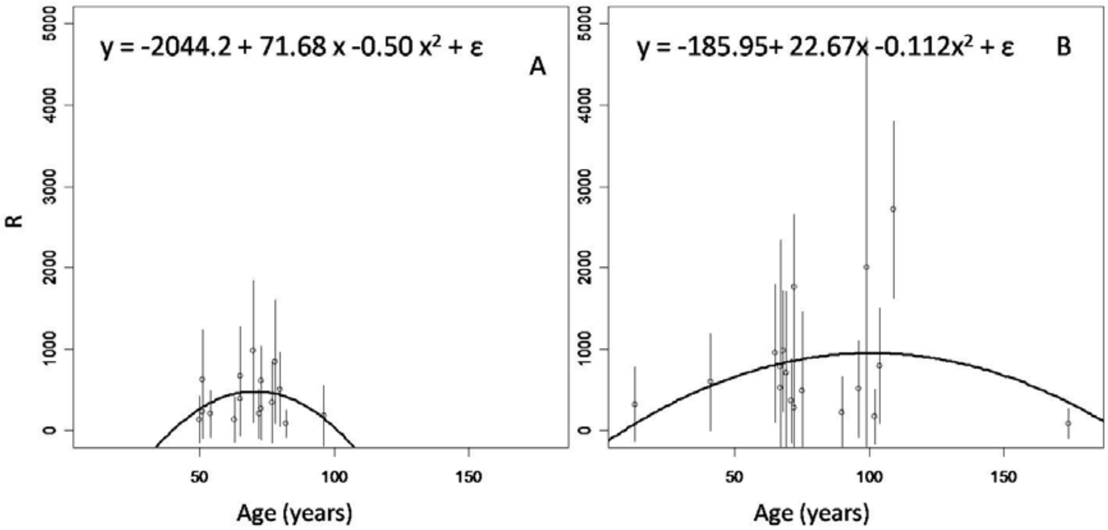

Figure 15: R_i vs. age at Duke Forest (A) and Coweeta (B).

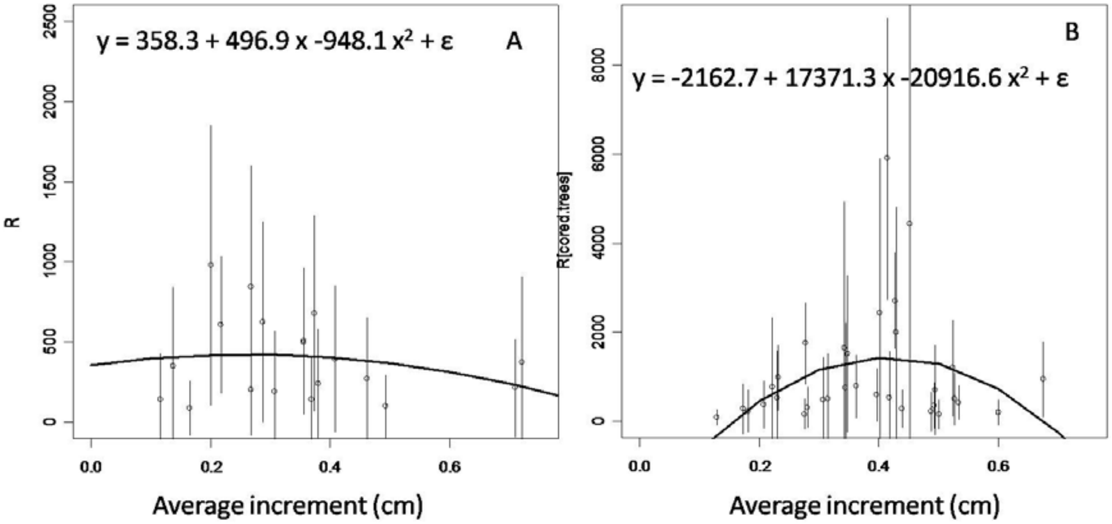

Figure 16: R_i vs average annual growth increment at Duke Forest (A) and Coweeta (B).

Fecundity and diameter

As predicted, reproductive success was moderately to strongly correlated with fecundity (annual seed production) (r = 0.296 at Duke Forest, 0.413 at Coweeta). This relationship is shown in figure 17. On average, less than 2% of the seeds produced per year result in a first-year seedling, but for some individuals this fraction can be as high as 20% (figure 18). Consistent with my hypothesis, germination fraction Q_i for individual trees was negatively correlated with fecundity. Consistent with the weaker correlation between fecundity and reproductive success at Duke Forest, this effect appeared slightly stronger at that site (r = -0.38) compared to Coweeta (r = -0.21) (figure 19). There was no relationship between Q_i and any other variable measured.

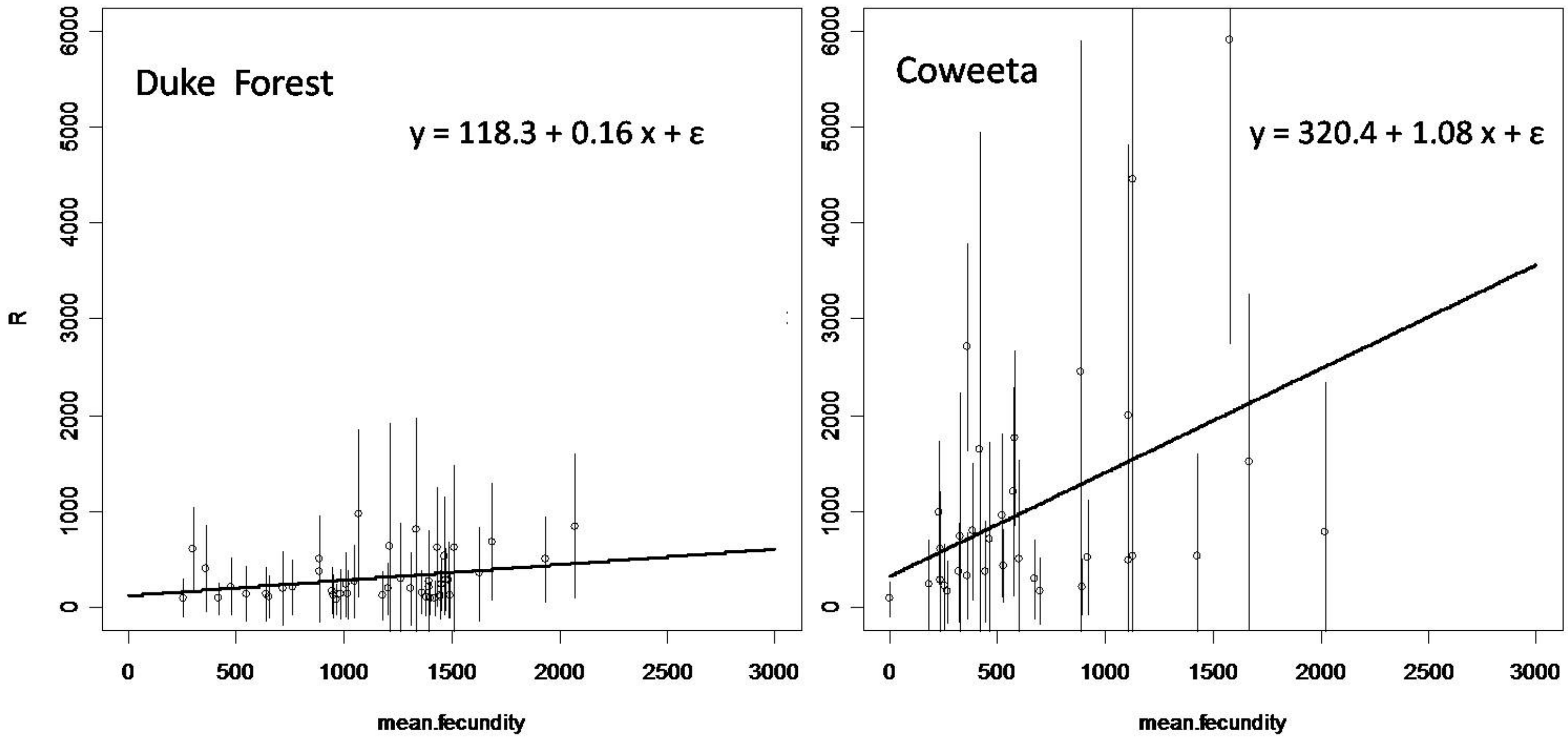

Figure 17: R_i vs. average fecundity at Duke Forest (left) and Coweeta (right).

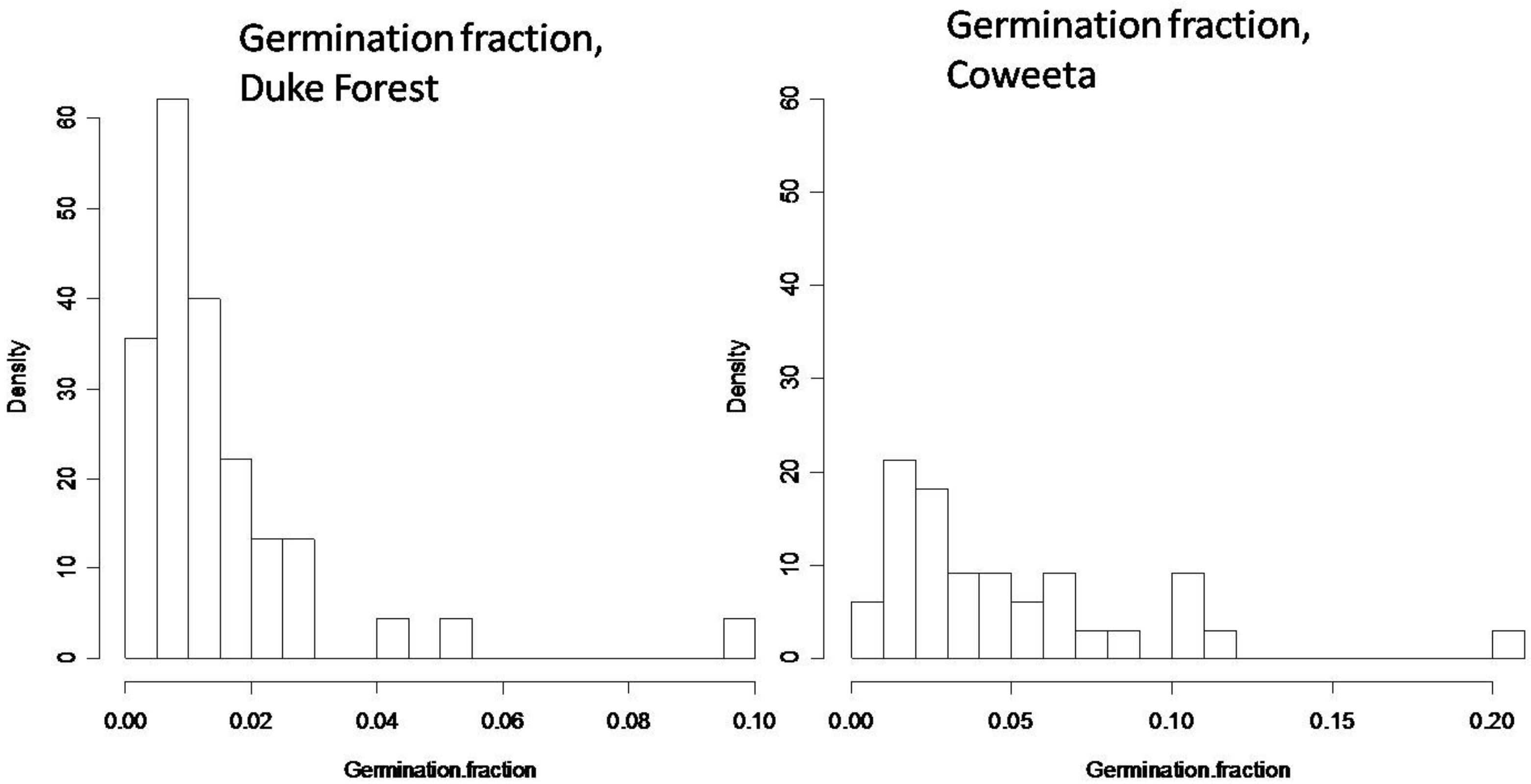

Figure 18: Average annual germination fractions at Duke Forest and Coweeta.

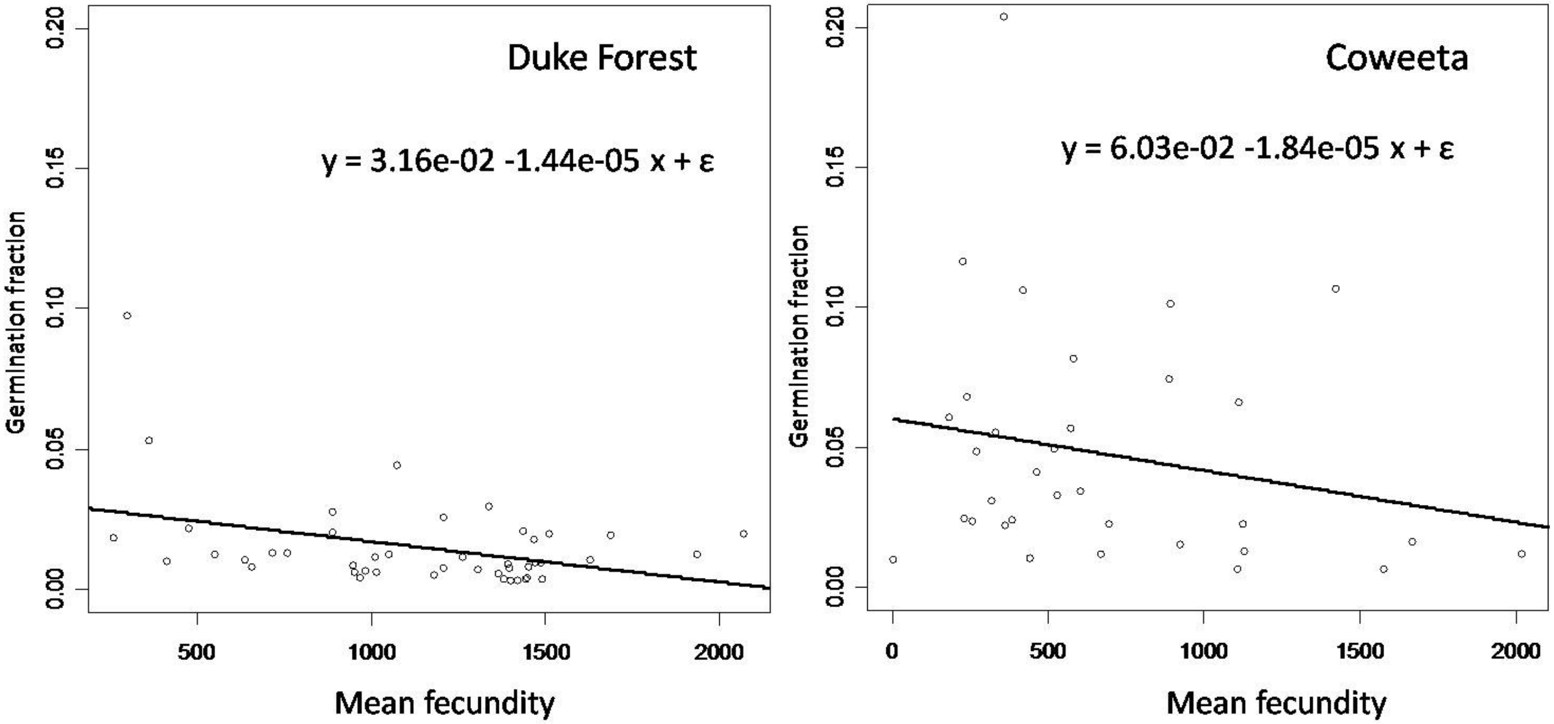

Figure 19: Average germination fraction vs. average fecundity at Duke Forest (left) and Coweeta (right).

Weak to moderate correlations of R_i with diameter were identified ($r = 0.19$ at Duke Forest, $r = 0.25$ at Coweeta). Given the non-linear relationship between age and reproductive success, the linear relationship between diameter and reproductive success is most likely due to the correlation between diameter and fecundity.

Effective population size

Simulations indicate that there is a strong sigmoid relationship between effective population size (or the ratio between effective and actual population size) and reproductive inequality as measured by the Gini coefficient (figure 20). Notice that, as theory requires, $N_e=N$ when $G = 0$ and N_e approaches 1 as G approaches 1. If approximately 50% of trees in a population are reproductive, and if the inequality in reproductive success between these individuals is $0.3 < G < 0.6$ (as data suggests), then effective population size should be between 15% and 40% of the actual population size.

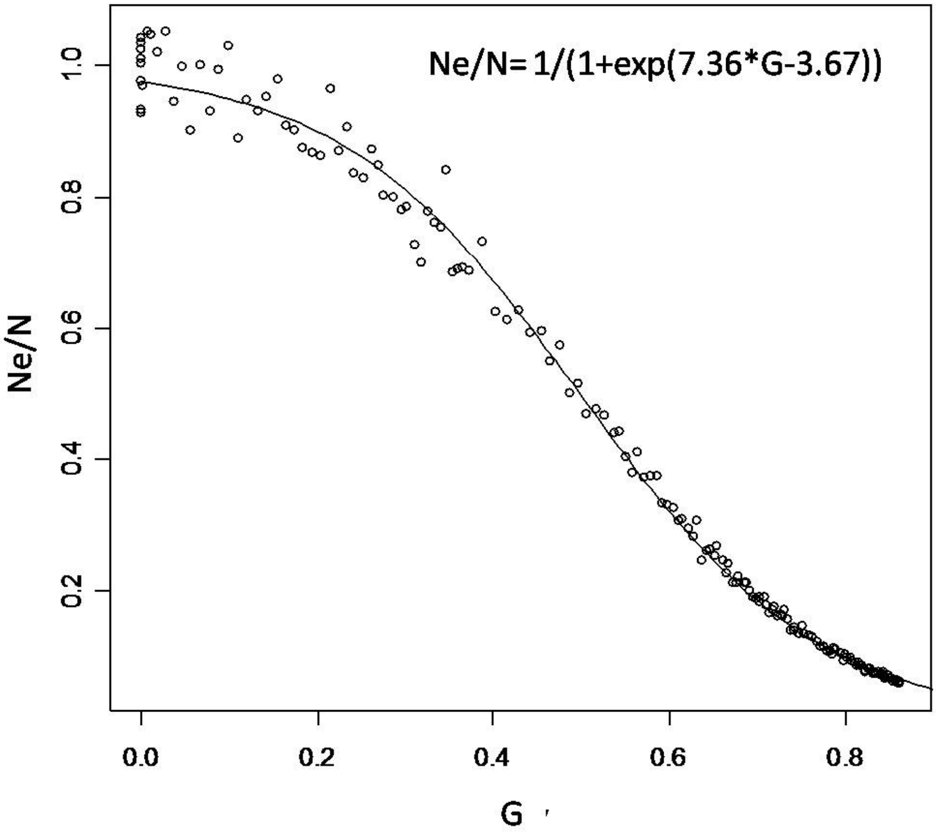

Figure 20: Ratio of N_e to actual population size for simulated populations with varying degrees of reproductive inequality.

Discussion

In plants, reproductive success depends on a) reaching reproductive maturity (in trees, dependent on size and canopy exposure); b) resource availability for flower production (dependent on water, soil nutrients, stored photosynthates); c) the availability of compatible pollen and/or stigmas; d) environmental conditions during seed maturation (available resources, predators); e) patterns of seed dispersal that can affect the probability of reaching suitable recruitment sites and intensity of competition or other density-dependent effects; f) the suitability of offspring locations for long-term survival and reproduction; as well as g) the genetic background and fitness of both parents and offspring. All these processes introduce interacting sources of variation.

I find that the distribution of reproductive success for red oaks at both of my study sites is highly variable, with a few individuals contributing disproportionately to the next generation while many individuals produce few or no offspring. Similar results have been found in previous studies of plant seedling production. Schnabel et al. (1998) found that 46-58% of *Gleditsia* seedlings were produced by just 3 female trees (8.6-8.8% of all mature females), while 47-51% of females each contributed less than 1% of seedlings. In *Pinus pinaster*, 10% of trees produced 50% of all saplings (Gonzalez-Martinez et al. 2006). Meagher and Thompson (1987) also found large differences in reproductive success between individuals of the herbaceous plant *Chamaelirium luteum*, suggesting that this may be a general pattern for all plant life forms. Highly skewed distributions of reproductive success are also common in animal species (Clutton-Brock 1988).

It has been suggested that lifetime reproductive success in both plants and animals follows a lognormal distribution resulting from the multiplicative effects of interacting random environmental effects (Herrera and Jovani 2010). My data suggest that this is true for those individuals that are reproductively active, but not always for the population as a whole. When all individuals are considered, the large number of trees which are immature or which have no established offspring create a large zero class, leading to a zero-inflated lognormal distribution (figure 12).

I observed a hump-shaped distribution between age and reproductive success, rather than the positive relationship hypothesized. Woody plants can exhibit senescence, producing fewer seeds as they age (Herrera and Jovani 2010, Clark et al. In press). For older trees, this reduced fecundity plus the advancement of previously established offspring to larger size classes (and out of the seedling class) may reduce the standing number of seedlings, leading to the observed pattern.

A hump-shaped relationship was also observed between average growth rate and reproductive success, although this was more pronounced at Coweeta than at Duke Forest (figure 16). This suggests the existence of tradeoffs between growth and reproduction. That is, while very slow-growing trees may lack the resources to invest in reproduction, very fast-growing trees may also be diverting resources toward growth and away from reproduction. This is consistent with an analysis of multiple tree species in North Carolina showing that the overall relationship between growth rate and fecundity can range between positive and negative even within a genus (Clark et al. In press). There was no relationship between germination fraction and growth rate, so fast-growing trees do not appear to be able to give their offspring a competitive advantage through seed provisioning. There was also no relationship between age and growth rate.

As predicted, trees with high annual fecundity and large diameter also exhibited high reproductive success. This was true despite the fact that the germination fraction was lower for highly fecund trees, suggesting the operation of density-dependent mortality at the seed or very early seedling stage (Janzen 1970, Connell 1978, Ibanez et al. 2007). This supports Hubbell's contention that density-dependent mortality may weaken the relationship between seed number and seedling number but will not necessarily make it negative (Hubbell 1979).

What does the observed variation in reproductive success mean for effective population size in oaks and the strength of drift versus selection? This and previous studies suggest that, at any given time, roughly 35-60% of the "adult" trees in a population are actively reproducing, and that this is not due to year-to-year variation, but rather to differences in seed production and seedling success that persist over multiple years (Schnabel et al. 1998, Gonzalez-Martinez et al. 2006, Herrera and Jovani 2010). This percentage is approximately equal to the $N/N_e = 1/2$ predicted by Nunney and Elam (1994) for a population with overlapping generations.

Variation in reproductive success between individuals can further reduce effective population size. Gini coefficients for my study populations were between 0.3 and 0.6, indicating moderately strong inequalities in reproductive success, though not total dominance by one or two individuals. Simulations revealed a sigmoidal relationship between effective population size (estimated from inter-generational changes in heterozygosity) and G. These simulations suggest that effective population size for species with a degree of reproductive inequality similar to that observed for my oak populations would be between 30% and 80% of the number of actively reproducing adults. This is of course a rough estimate, and if the species is dioecious or the population is widespread enough for isolation-by-distance to be important, these factors should also be taken into account. The stands investigated in this study are part of a continuous forest, and many seedlings have parents located outside the mapped stands (**Chapter 3**). Gene flow from outside the plot boosts effective population size, and allows the maintenance of high levels of genetic variation (**Appendix 1.1**). However, these results suggest that if these stands were suddenly isolated (by clearing of the surrounding forest, for instance), the reduction in effective population size due to reproductive inequality would make them much more vulnerable to the loss of genetic variation. The effective population size for the 118-individual Duke Forest stand, where half the trees contribute to the seedling pool, would be approximately 40. For the 199-individual Coweeta stand, where only about 40% of trees are parents, effective population size would be approximately 66. Genetic structure data indicates that the co-occuring red oak species at these sites hybridize (**Chapter 5**). If they did not, effective population sizes would be considerably lower.

Oaks and other wind pollinated trees are generally regarded as having relatively high effective population size and thus being relatively well buffered against loss of genetic variation (Hamrick 2004). However, my results suggest that small, fragmented populations of oaks may be vulnerable to the loss of genetic variation through drift, although this loss may take multiple generations (measured in decades to centuries) to become obvious. Fernandez and Sork (2007) have documented such a reduction in allelic diversity of *Quercus humboltii* seedlings relative to adults within forest fragments (<50 trees). Conservationists should thus not take the large population sizes of many temperate trees for granted; variation in reproductive success may effectively make these populations much smaller than they appear.

The large discrepancies between the methods of calculating effective population size and the highly skewed distributions of reproductive success found in many populations illustrate the need to develop better approaches, both theoretical and empirical, to estimating effective population size in forest trees. Ideally, these approaches would include a standardized method of quantifying variation in reproductive success (such as the Gini coefficient) that does not require offspring number to be normally distributed. Simulations can help to identify these relationships. By allowing practitioners to calculate the risk of genetic impoverishment under different sets of assumptions, they may also be useful in the management of species whose life histories are not fully understood.

Chapter 5. Hybridization and genetic structure in two mixed-species red oak populations

Introduction

Understanding the extent of hybridization between species and is important for managing, classifying, and conserving populations. This issue is of particular importance for oaks (*Quercus* spp.), as disagreement over the extent of gene flow between putative species and the implications of such gene flow has persisted for decades (Palmer 1942, Burger 1975, Kashani and Dodd 2002, Coyne and Orr 2004). The common view is that the potential for hybridization between oak species is extensive within but not between sections of the genus (eg. red oaks, section *Lobatae*, or white oaks, section *Quercus*). On the other hand, many oak species appear to fill distinct ecological roles and, while notoriously variable in morphology, they appear to be morphologically cohesive over large spatial scales (Whittemore and Schaal 1991, Bacilieri et al. 1996). The controversy has persisted in part because, until recently, most evidence for hybridization between oak species was purely morphological (Rushton 1993). In the last two decades, genetic information has begun to contribute to the debate, though results to date have been mixed. In this study I investigate the extent of gene flow between co-occuring red oak species at two sites in North Carolina using measurements of genetic structure coupled with parentage analyses.

The earliest records of oak hybrids date to the early 1800's (Rushton 1993). Since that time, many cases of putative natural hybridization have been identified, and deliberate crosses have confirmed that many oak species are inter-fertile (Cottam et al. 1982). While some have argued that widespread hybridization throws doubt on the utility of the biological species concept for oaks (Burger 1975), others point out that the way in which oaks maintain their morphological and ecological distinctness, instead of merging into a few large hybrid swarms, suggests the existence of reproductive barriers (Coyne and Orr 2004). "Intermediate" phenotypes can also occur even where genetic differentiation is strong, due to the high levels of intra-specific morphological variation in some oak species (Craft et al. 2002). On the other hand, introgression can be cryptic, with no obvious effects on morphology (Dodd and Afzal-Rafii 2004).

Some prezygotic barriers do appear to exist in oaks, as the success rate for deliberate interspecific crosses is typically lower than for intraspecific crosses (Rushton 1993); nevertheless, in one of the most extensive oak hybridization trials, Cottam et al (1982) recorded 43 different successful interspecific hybrid combinations out of 137 attempted. In some cases, hybrids exhibit reduced fertility compared to their parental species (Rushton 1993), but fitness effects are subtle, once hybrid offspring are produced – given the right conditions for growth, hybrid trees generally appear healthy, and produce viable seed and pollen (Salvini et al. 2009).

The fact that a cross can be produced in the arboretum does not necessarily mean that it will be successful in the wild. Early studies relied on the existence of intermediate morphologies or, later, biochemical characteristics to identify putative hybrids (Rushton

1993, Dodd and Afzal-Rafii 2004), but it was unclear whether such characters reflected gene flow or convergent evolution. Genetic markers have more recently been applied to the problem, and have revealed a wide range of gene flow patterns. Some studies have shown levels of genetic differentiation that suggests a high level of reproductive isolation (Craft et al. 2002, Muir and Schlotterer 2005), while others have found evidence of cryptic introgression in individuals that appear morphologically pure (Aldrich et al. 2003b, Dodd and Afzal-Rafii 2004), and still others have found evidence of introgression of chloroplast haplotypes but not of nuclear markers (Whittemore and Schaal 1991). Together, these studies suggest that oak hybridization does occur in the wild, in at least some cases, but that natural selection generally maintains the suites of identifiable morphological and biochemical species traits.

Differences in flowering time can constitute a partial reproductive barrier (Rushton 1993). However, all oaks are wind pollinated, bloom in spring, and from year to year can have widely varying flowering times depending on climatic conditions (Ducousso et al. 1993). A stronger filter may be selection against trees bearing combinations of traits ill-suited to parental habitats (Anderson 1948). Dodd and Afzal-Rafii (2004) found that among individuals with classic *Q. wislizeni* morphology there was evidence of extensive introgression from three other species, and that the proportion that different species contributed to the genome was correlated with environmental characteristics of the site (temperature and moisture). This finding supports the idea that selection for suites of traits advantageous in a given local environment helps maintain the distinctiveness of oak species, or at least limits the extent of introgression from species adapted to different environments.

If the local climate determines flowering time, and is one of the factors selecting for particular combinations of traits that help to differentiate species, what happens to patterns of gene flow between oak species when the environment changes? In some cases, climate change might strengthen reproductive isolation – by lengthening the growing season and allowing differentiation in flowering time, for instance (Rushton 1993) – while in other cases introgression from related species might be favored. For example, in the southeastern US, northern red oak (*Q. rubra*) frequently co-occurs with southern red oak (*Q. falcata*) and black oak (*Q. velutina*), both of which occur in more xeric microhabitats and have more southerly range limits (Little 1980). If this region becomes warmer and more arid in the future, as predicted by climate models (Mearns et al. 2003), then introgression from these species could introduce useful genetic variation to *Q. rubra* while, conversely, introgression from *Q. rubra* might be disadvantageous to *Q. falcata* and *Q. velutina*. It has also been suggested that hybridization of some oak species with relatives with better dispersal abilities, followed by later backcrosses that 'resurrect' the original phenotype, may have sped up the migration of some oak species during past periods of climate change (Petit et al. 2003). With or without global change, understanding patterns of gene flow between species is important for the management of forest resources. Forest managers interested in promoting certain characteristics – such as the production of knot-free timber, mast for wildlife, or resistance to pests or disease – may find their efforts unexpectedly helped or hindered by cryptic gene flow (Elena-

Rossello et al. 1992, Aldrich et al. 2003b). There are also conservation implications of hybridization, as the ecological and genetic distinctness of rare species and sub-species can be endangered if they hybridize with a common relative (Allendorf et al. 2001, Mank et al. 2004, Lorenzo et al. 2009).

Red oaks (section *Lobatae*) are an important component of the North American flora, with 200 endemic species (Kashani and Dodd 2002). Temperate red oaks differ from co-occuring white oaks (section *Quercus*) in a number of physical and life-history characteristics. Red oak acorns mature in two years rather than one (Schwarzmann and Gerhold 1991) and, while white oak acorns germinate immediately in autumn, tannin-rich red oak acorns require a chilling period to sprout. This in turn leads squirrels to cache them intact, while white oak acorns are usually consumed immediately or buried with embryo excised (Steele et al. 2004). These and other differences suggest that both intra- and inter-specific patterns of gene flow may differ considerably between red and white oaks. While the genetics of white oaks have been extensively studied, particularly in European species, few studies have examined hybridization and genetic differentiation in red oaks despite indications that species barriers in this group may be much weaker (Guttman and Weigt 1989, Kashani and Dodd 2002, Aldrich et al. 2003b).

In this study, I measure the degree of neutral genetic differentiation between co-occurring red oak species at two sites in North Carolina, pairing these genetic structure measures with parentage analyses of seedlings to investigate the occurrence of ongoing inter-specific gene flow. If species never or rarely hybridize, then we would expect to observe a high degree of genetic differentiation between morphologically-defined species groups. If, on the other hand, hybridization is common enough that species regularly exchange genetic material, genetic structure measures should show no significant differentiation between "species", and/or genetic structuring within stands that does not correspond to morphologically-defined species. Additionally, if hybridization is ongoing, we would expect this to be reflected in parentage analysis results.

Methods

Data collection from the Duke Forest and Coweeta populations is described in **Chapter 1**. At Duke Forest, three co-occuring red oak species are considered: *Quercus rubra* (northern red oak), *Q. velutina* (black oak), and *Q. falcata* (southern red oak). At Coweeta, *Q. falcata* is not present, but *Q. coccinea* (scarlet oak) is common. Hybridization between *Q. rubra*, *Q. velutina*, and *Q. coccinea* has been reported (Jensen 1977, Tomlinson et al. 2000), while *Q. rubra* and *Q. falcata* are believed to be very closely related (Guttman and Weigt 1989).

We identified trees using a whole-tree silvics approach, similar to that described by Aldrich et al. (2003b) (figure 21). "*Quercus rubra*" individuals have large smooth, thin leaves with 7-11 bristle-tipped lobes all the way around the margin; grey bark that in adult trees is deeply cracked and furrowed, giving the impression of dark and light stripes; and large acorns(1.5-2.8 cm) with shallow cups of closely overlapping scales. "*Quercus coccinea*" individuals are similar, but the leaf lobes are narrower and deeper, divided nearly to the midvein; the pale stripes on the bark are less obvious; they tend to retain dead branches in their crowns; and their acorns are egg-shaped, 1.2-2.5 cm long. "*Quercus velutina*" individuals have thicker, hairier (pubescent) leaves with shallower lobes than *Q. rubra*; their bark is blackish and deeply cracked and furrowed; and their acorns are elliptical (1.5-1.9 cm long) and lighter in color than *Q. rubra's*, the cup scales long, loose, and pointed. Finally, "*Q. falcata*" individuals have thick leaves with a mat of brownish hairs underneath and 3 primary lobes near the tip; medium-grey ridged bark; and small (1.2-1.5 cm), round, dark brown acorns (Little 1980). Nearly all adults could be easily classed according to these features. A few individuals were more difficult to classify than others, but were assigned to the species with which they shared the most features.

Figure 21: Typical adult leaves and acorns.
Left: Adult leaves for *Q. coccinea* (top), *Q. falcata* (left), *Q. rubra* (middle), and *Q. velutina* (right). Right: Acorns of *Q. rubra* (top row), *Q. velutina* (middle), and *Q. falcata* (bottom).

All four oak species are broadly distributed in eastern North America (Little 1980). *Quercus rubra* has the widest range, reaching from Ontario to Alabama. Southern red oak (*Q. falcata*), as the name suggests, has a more southern range, from southern Indiana and

New Jersey to northern Florida. *Quercus velutina* does not grow as far north as *Q. rubra*, but is otherwise similarly distributed. *Quercus coccinea* has perhaps the most restricted distribution, being concentrated around the Appalachian Mountains from SW Maine to Georgia. *Quercus falcata*, *Q. velutina*, and *Q. coccinea* all tolerate drier soils than *Q. rubra*. Nevertheless, at both sites species are frequently found growing side-by-side (figure 22).

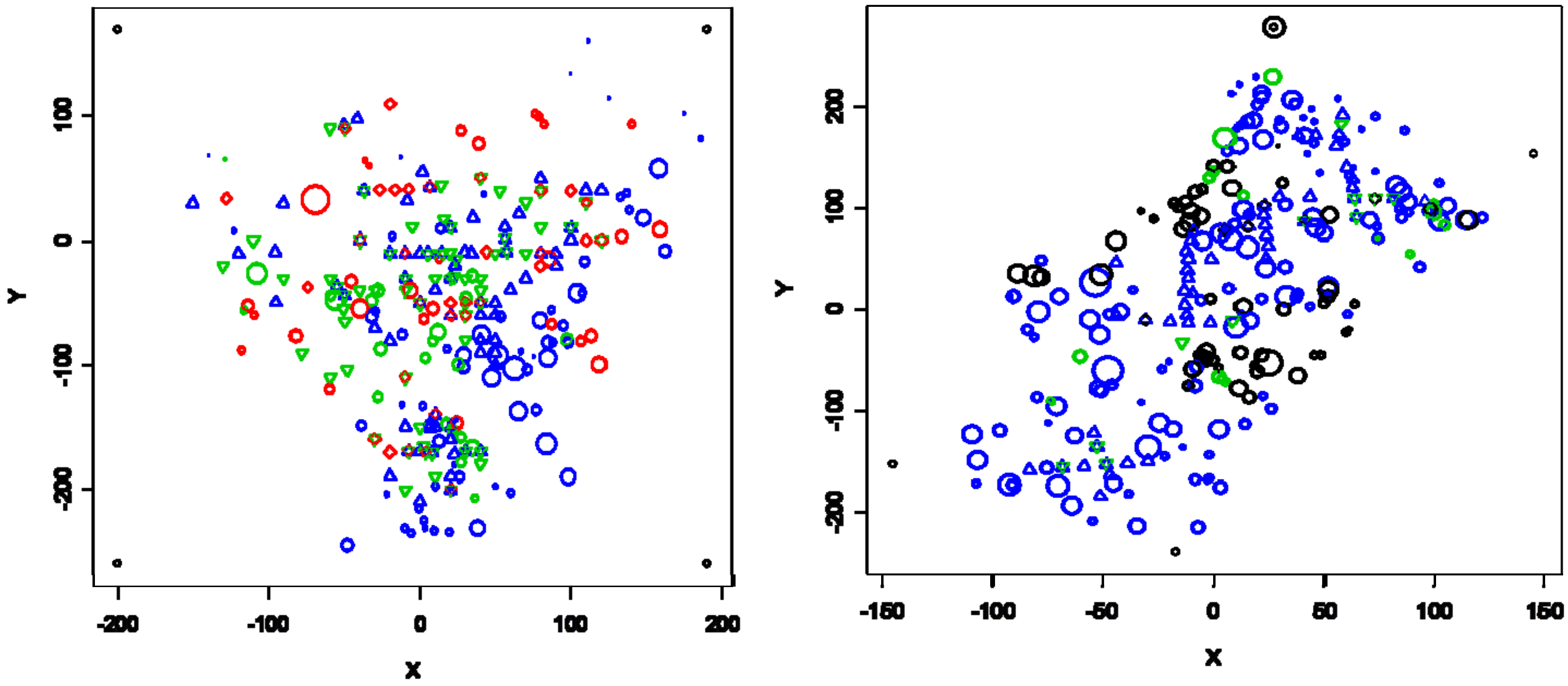

Figure 22: Locations of adults and sampled seedlings of the 4 red oak species at Duke Forest (left panel) and Coweeta (right panel).

Circles proportional to adult diameter. Blue – *Q. rubra* (seedlings upright triangles); Red- Q. falcata (seedlings diamonds); Green – *Q. velutina* (seedlings inverted triangles); Black – *Q. coccinea* (seedlings diamonds).

Adult genotype data was analyzed using the program STRUCTURE 2.2 (Pritchard et al. 2000), which uses Bayesian inference to evaluate the probability that sampled genotypes come from a given number of populations K. If populations (in this case species) are reproductively isolated, over time genetic drift will cause their allele frequencies to differ. Hybridization, on the other hand, increases the similarity of allele frequencies. This analysis therefore reflects long-term patterns of gene flow. A run of 10,000 MCMC steps after an initial burn-in of 10,000 proved to be sufficient for parameter convergence. I used the admixture model, which allows that individuals may derive portions of their genomes from more than one population. Since shared ancestry may introduce correlations in allele frequencies between populations, I ran the data using both correlated and uncorrelated frequency assumptions. Three runs were conducted for each site at K=1 through K=5.

F-statistic, R-statistic, and AMOVA analyses for adult data were conducted using Genalex (Peakall and Smouse 2006). The fixation index F describes the reduction in heterozygosity relative to Hardy-Weinberg expectations in a set of populations. F can be subdivided into Fis, which describes the reduction in heterozygosity within populations (in this case, species) due to inbreeding, and Fst, which describes the deficit in heterozygosity between populations due to genetic drift. High values of Fst relative to Fis are expected for populations that are strongly differentiated. Fst values smaller than about 0.01 are generally considered biologically insignificant, indicating that subdivisions act as one inter-breeding population (Aldrich et al. 1998, Gillespie 2004, Sato et al. 2006, Craft and Ashley 2007). R-statistics (Rst, Ris) refer to similar measures developed specifically for microsatellites, to take into account that they are scored by length and are believed to mutate in a stepwise fashion (Slatkin 1995). I compared Rst and Fst values for between-species comparisons to differentiation between *Q. rubra* adults at the two sites. Duke Forest and Coweeta are approximately 300 km apart. Because of the low degree of between population genetic isolation generally exhibited by oaks and other wind-pollinated species at this scale (Schwarzmann and Gerhold 1991, Sato et al. 2006, Craft and Ashley 2007), if species are reproductively isolated within-site between-species differentiation is expected to be greater than between-site within-species differentiation.

Table 6: Sample size

Species	**Adults, Duke Forest**	**Seedlings, Duke Forest**	**Adults, Coweeta**	**Seedlings, Coweeta**
Q. rubra	68	96	129	158
Q. velutina	22	85	15	13
Q. falcata	28	38	NA	NA
Q. coccinea	NA	NA	54	7

Results of the parentage and dispersal model detailed in **Chapter 3** provide evidence for hybridization that is revisited here. Potential hybrids are identified as seedlings for which one member of the most-frequently identified parent pair is identified as a different species from the other parent or from the seedling.

Results

F statistics and AMOVA:

Between-site differentiation was low for adult *Q. rubra* (Fst = 0.107, Rst = 0.001). Almost no significant between-species structure was detected at either site, overall or for pairwise comparisons (tables 7-8).

Table 7: Overall F and R statistics, Duke Forest and Coweeta

	Duke Forest		**Coweeta**	
	Value	**P**	**Value**	**P**
Fst	0	0.469	0.003	0.093
Fis	0.209	0.001**	0.237	0.001**
% of variation between-pop, among indiv, within indiv	**0%** 21% 79%		**0%** 24% 76%	
Rst	0.006	0.202	0	0.957
Ris	0.954	0.001**	0.979	0.001**
% of variation between-pop, among indiv, within indiv	**1%** 95% 5%		**0%** 98% 2%	

The only exceptions are for the pairwise Rst of *Q. rubra* and *Q. falcata* at Duke Forest and the pairwise Fst of *Q.velutina* and *Q. coccinea* at Coweeta (table 8) but, even if statistically significant, these values were still extremely low. The Fst of 0.004 between *Q. rubra* and *Q. falcata*, for instance, is much smaller than the value of 0.017 between *Q. rubra* at Duke Forest and *Q. rubra* at Coweeta. AMOVA analyses showed that most of the molecular variance occurred among and within individuals, with less than 1% attributed to between-species differences (table 7).

Table 8: Pairwise Fst and Rst. P values in parentheses.

Duke Forest	***Pairwise Fst (Pval)***	***Pairwise Rst (Pval)***
Q. rubra - Q.velutina	0 (*0.433*)	0 *(0.411)*
Q. velutina - Q.falcata	0 (*0.455*)	0.01 *(0.193)*
Q.rubra - Q. falcata	0 (*0.461*)	**0.016** *(0.034*)*
Coweeta	***Pairwise Fst (Pval)***	***Pairwise Rst (Pval)***
Q. rubra - Q.velutina	0.001 (*0.282*)	0 *(0.396)*
Q. velutina - Q.coccinea	0.000 (*0.464*)	0 *(0.431)*
Q.rubra - Q. coccinea	**0.004** (*0.002**)	0 *(0.389)*

STRUCTURE:

If the three species at each site are reproductively isolated, then strong genetic structure is expected, with clusters of genetically similar individuals corresponding to species identities. Like the F statistic analyses, STRUCTURE analyses did not support this hypothesis. Within each site heterozygote frequencies did not conform to Hardy-Weinberg equilibrium, so results indicated the existence of population sub-structure. However, support for any particular K>1 was not high (table 9).

Table 9: STRUCTURE RESULTS, lnP(data|K)

Adult trees only, for three runs at each population (K) value. Highest likelihoods shown in bold.

Duke Forest					
	K=1	K=2	K=3	K=4	K=5
Correlated	-5333.2	-5042.4	-5045.6	**-5020.1**	-5149.1
	-5332.2	-5039.1	-5076.0	-5019.6	**-5010.3**
	-5334.4	-5038.3	**-5025.4**	-5186.9	-5174.6
Uncorrelated	-5333.4	-5063.0	-4978.9	-4966.3	**-4914.1**
	-5333.1	-5056.7	-4984.6	-4947.9	**-4889.5**
	-5333.8	-5059.7	-4979.8	-4942.3	**-4933.2**
Coweeta					
	K=1	K=2	K=3	K=4	K=5
Correlated	-5988.6	-5970.9	-5924.5	-5878.1	**-5719.1**
	-5985.5	-5957.2	-5990.6	-5798.7	**-5782.8**
	-5988.0	-5975.4	-5896.3	**-5846.4**	-5887.8
Uncorrelated	-5988.6	-5911.3	-5897.0	-5841.8	**-5805.7**
	-5989.3	-5906.9	-5969.4	-5846.6	**-5801.4**
	-5988.4	-5910.8	-5970.9	-5854.4	**-5816.4**

Assumptions of correlated allele frequencies between species led to slightly smaller numbers of "populations" being favored than if allele frequencies were independent. Genetic clusters did not, for the most part, correspond closely to morphological species, and most individuals were inferred to be of mixed ancestry. At Duke Forest, individuals identified as *Q. rubra* or *Q. velutina* mostly fell between populations 1 and 3 (when K = 3). Individuals identified as *Q. falcata* were assigned primarily to population 2 (as were a few *Q. velutina*), but a few *Q. falcata* also fell between 1 and 3 (figure 24); these individuals were all >10 cm dbh, and so are not the product of very recent hybridization. At Coweeta, differentiation between species was even lower, though there was some tendency for *Q. coccinea* to be assigned to population

2 (figures 23 &24). Results were similar when all individuals (adults and seedlings) were included.

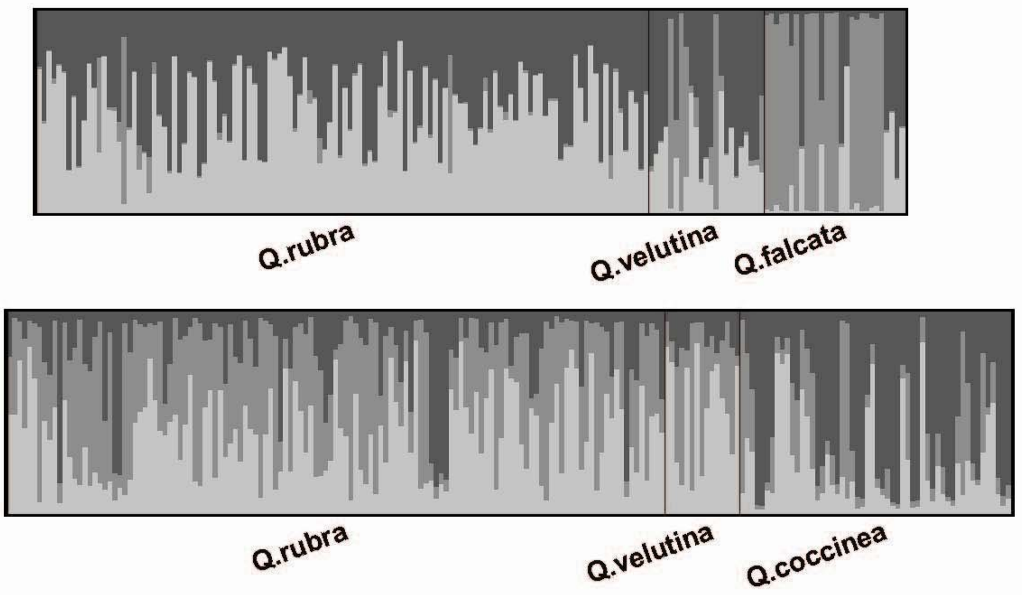

Figure 23: Population structure at Duke Forest (top) and Coweeta (bottom) for K=3.

Correlated allele frequencies. Vertical bars represent individuals, the size of the shaded bands reflecting the proportion of the genome attributed to genetic clusters 1 (light grey), 2 (grey) or 3 (dark grey). Blocks on the x axis correspond to morphological species.

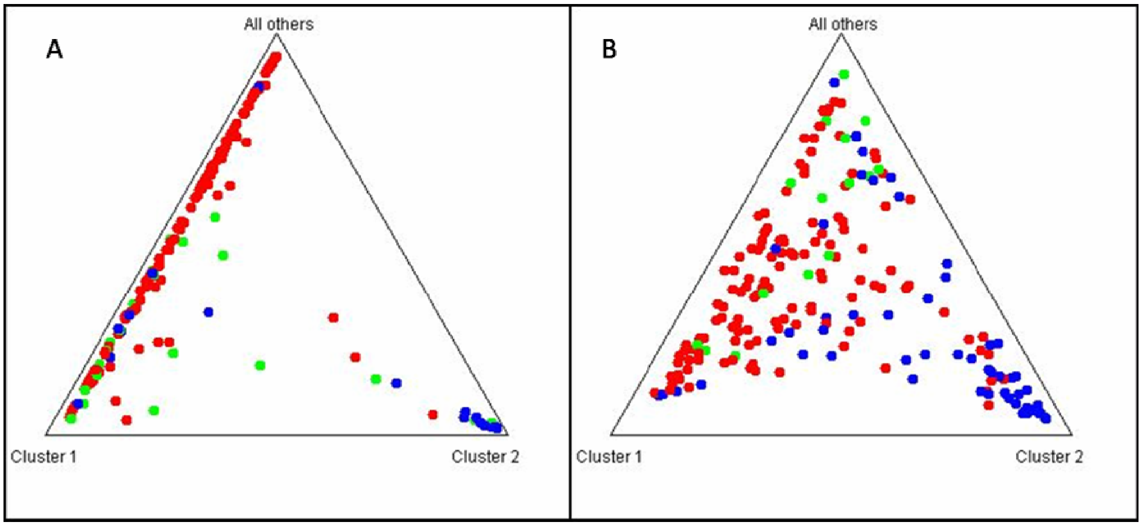

Figure 24: Triangle plot of population structure at Duke Forest (a) and Coweeta (b)

Correlated frequencies, K=3. *Q. rubra* = red; *Q. velutina* = green. At Duke Forest *Q. falcata* = blue, at Coweeat *Q. coccinea* = blue. If each morphological species classification corresponded perfectly to a genetic grouping, all individuals of the species would cluster in the corresponding point of the triangle. Instead, most individuals are estimated to be of mixed ancestry, and are fall in the center or along an edge.

Parentage:

Thirty-one out of 219 seedlings (14.2%) at Duke Forest and 37 out of 179 seedlings (20.7%) at Coweeta were estimated to have a parent belonging to a different species than the seedling in question. The results, broken down by species, can be seen in table 10. Numbers in bold indicate how many seedlings were *not* linked to any heterospecific parent. Note that this is by far the largest category (among seedlings for which at least one parent was identified) for nearly all species. All *Q. velutina* seedlings at Coweeta were linked to heterospecific parents, but this may be an effect of small sample size. Cases in which at least one heterospecific parent was identified represent putative hybridization events. Notice that the most frequent pairings include *Q. falcata* to *Q. velutina* at Duke Forest but very few *Q. rubra* to *Q. falcata*, and that the most differentiated species at Coweeta, *Q. coccinea*, also has the highest percentage of intraspecific matches (albeit the smallest sample size). Notice that the most frequent pairings include *Q. falcata* to *Q. velutina* at Duke Forest but very few *Q. rubra* to *Q. falcata*. *Quercus falcata* and *Q. coccinea*, which were the most highly differentiated species in terms of allele frequencies, also exhibit the highest percentage of seedlings without identified heterospecific parents.

Table 10: Percentage of seedlings by species with only conspecific parents identified (bold), at least one heterospecific parent identified, or neither parent identified

Duke Forest	***Q. rub.* parent**	***Q. velu.* parent**	***Q. fal.* parent**	**Parent unid.**
***Q. rubra* sdl** (N=96)	**47 (48.9%)**	3 (3.1%)	6 (6.2%)	40 (41.7%)
***Q. velutina* sdl** (N=85)	4 (4.7%)	**37 (43.5%)**	11 (12.9%)	33 (38.8%)
***Q. falcata* sdl** (N=38)	1 (2.6%)	6 (15.8%)	**23 (60.5%)**	8 (21%)
Coweeta	***Q. rub.* parent**	***Q. velu.* parent**	***Q. coc.* parent**	**Parent unid.**
***Q. rubra* sdl** (N=158)	**121 (76.6%)**	0 (0%)	25 (15.8%)	12 (7.6%)
***Q. velutina* sdl** (N=13)	7 (53.8%)	**0 (0%)**	4 (30.8%)	2 (15.4%)
***Q. coccinea* sdl** (N=7)	0 (0%)	1 (14.3%)	**6 (85.7%)**	0 (0%)

Discussion

At both sites, there is little between-species genetic structure, as revealed by low Fst/Rst values relative to between-population within-species comparisons and in STRUCTURE analyses. Although shared alleles between species can result from common ancestry or repeated mutations (Muir and Schlotterer 2005), the similarity in allele frequencies at all six loci for co-occurring species at these sites is too high to be the result of these processes alone. Allowing for correlated allele frequencies increases the sensitivity of STRUCTURE to any population differentiation that may exist. While the correlated-frequency model tended to estimate a smaller number of subdivisions than did the uncorrelated-frequency model, in both cases these genetic subdivisions did not correspond to morphologically-defined species (figure 23). Genetic structure within these populations may reflect clusters of related individuals or other spatial structuring more than species differences.

When populations are strongly isolated, STRUCTURE can identify the correct number and composition of populations using as little as 5-7 microsatellite loci (Pritchard et al. 2000). Craft et al. (2002) detected significant differentiation between *Q. lobata* and *Q. douglasii* using only 4 microsatellite markers, indicating very strong reproductive isolation. Similarly, Salvini et al.(2009) detected clear genetic differentiation between individuals of non-intermediate morphology at 5 microsatellite loci in an interbreeding population of *Q. petraea* and *Q. pubescens*. My results, on the other hand, are consistent with those of Aldrich et al (2003b) for red oaks in Indiana, which were based on 15 microsatellite loci. In a study of three co-occuring white oaks, Craft and Ashley (2007) found that STRUCTURE and PCA analyses (based on 5 microsatellite loci) did not support the existence of well-defined genetic clusters corresponding to traditional species, despite the fact that Fst analyses indicated significant differentiation between *Q. alba* and *Q. macrocarpa/Q. bicolor* (but no significant differentiation between *Q. macrocarpa* and *Q. bicolor*, which are known to introgress). They concluded that these species (and the 5 populations in which they were found) seemed to act as a single interbreeding population across the northeast corner of Illinois. Within-species between-site Fst values of 0.0067 to 0.116 were observed in this study (sites within 150 km of one another) while Fst values for *Q. alba* vs. *Q. macrocarpa* or *Q. bicolor* ranged from 0.06 to 0.107. The degree of inter-species genetic differentiation was even lower for my populations. I conclude that there is no evidence of strong reproductive isolation between the oak species at either site, although the slight tendency of *Q. falcata* individuals to cluster together may indicate that this species is less likely to hybridize with *Q. rubra* than either is to hybridize with *Q. velutina*.

The low differentiation between species supports the inclusion of all individuals as potential parents and offspring in parentage analyses, as does the observation that seeds taken from known mother trees in this population sometimes yield seedlings of a different morphospecies from the mother (pers. obs., pers. comm. Inez Ibanez). Parentage analyses suggest hybridization rates of up to 20.7% at Coweeta and up to 14.2% at Duke Forest, which are high but within the range of previously reported values. Reported hybridization rates among white oaks range from over 25% to almost zero (Table 11). Hybridization between North American red oak species has not been extensively studied. Those studies that do exist are summarized in table 12. None of these studies examined parentage, focusing instead on adult trees selected according to their “pure” or “intermediate” morphology.

Table 11: Published studies of hybridization in white oaks. In bold – the well studied *Q. petraea*/*Q. robur* system

Citation	Species	Life Stage	Hybridization	Comments
Craft and Ashley 2006	*Q. alba* *Q. bicolor* *Q.macrocarpa*	adults	Lots (%'s not calculated) – little differentiation	STRUCTURE and PC applied to multiple sites, IL
Salvini et al. 2009	*Q. pubescens* *Q. petraea*	seeds	♀ *Q. pubescens* – 25% ♀ *Q. petraea* – 5%	Mixed stand, Italy
Valbuena-Carabana et al. 2007	*Q. pyrenaica* *Q. petraea*	adults	22%, 9%, 6%	3 mixed populations, Spain
Bacilieri et al. 1996	***Q. petraea*** ***Q. robur***	seed from "typical" mothers	♀ *Q. robur* – 27 to 48% ♀ *Q. petraea* ~ 0%	Mixed species stand, France
Whittemore and Schall, 1991	*Q. alba* *Q.macrocarpa* *Q. michauxii* *Q. stellata* *Q. virginiana*	"typical" adults	Shared chloroplast haplotypes within localities: *Q.a* ,*Q. ma*, *Q.s* & *Q. mi* *Q. s* &*Q. v*	Sites in WI, MO, AR, TX, KS, MN, Mexico
Curtu et al. 2007	*Q. pubescens* *Q. frainetto*	adults	16.2%	Mixed species stand, Romania
Streiff et al. 1999	***Q. petraea*** ***Q. robur***	seed from "typical" mothers	7%	Mixed stand, France
Craft et al. 2002	*Q. lobata* *Q. douglasii*	adults	4.6%	Used STRUCTURE for mixed stand, CA
Burgarella et al. 2009	*Q. ilex* *Q. suber*	adults	<2%	used STRUCTURE for adults across zone of overlap, Europe
Curtu et al. 2007	*Q. frainetto* *Q. robur*	adults	1.7%	Mixed species stand, Romania
Muir and Schlotterer, 2005	***Q. petraea*** ***Q. robur***	"typical" adults	~0%, based on strong differentiation	Used STRUCTURE for adults from 6 sites across Europe

Table 12: Published hybridization studies for red oaks

Citation	**Species**	**Life Stage**	**Hybridization**	**Comments**
Aldrich et al. 2003	*Q. rubra* *Q. shumardii* *Q. palustris*	adults	Lots (%'s not calculated) – little differentiation	STRUCTURE analysis for old growth forest, IN
Dodd and Afzal-Rafii, 2004	*Q. wislezni* *Q. parvula* *Q. agrifolia* *Q. kelloggii*	"typical" adult *Q. wislezeni*	<u>Hybrid Index</u> (1 = pure *Q.w*; 0 = pure other): *Q. parvula* – 0.66 *Q. agrifolia* – 0.88 *Q. kelloggii* ~ 1	Sites across CA. Fixed differences in AFLP's only evident between *Q. kelloggii* and *Q. wisezeni* (Kashani and Dodd 2002)
Penaloza-Ramirez et al. 2010	*Q.hypoleucoides* *Q. scytophylla* *Q. sideroxyla*	adults	2 and 3-species hybrids. But species morph. & genetically distinct	Mexico. Intermediate individuals found in zones of contact.
Tovar-Sanchez & Oyama 2004	*Q. crassifolia* *Q. crassipes*	adults	Hybrids occur in zones of overlap. 4.5% of trees ID'd as *Q. crassifolia* and 8.6% ID'd as *Q. crassipes* shown to be hybrids	Seven hybrid zones and two pure stands, Mexico

Both low genetic differentiation and the high number of heterospecific parent-offspring pairs identified for both Duke Forest and Coweeta populations suggest extensive ongoing gene flow. Nevertheless, some caveats must be mentioned. First, artificial hybridization studies suggest that inter-specific pollination may have low success, but I did not have enough information to attach weights to this process in the parentage model. Second, because seedlings are morphologically variable and characters such as bark texture and acorn size cannot be examined, there is the possibility that a putative hybridization event merely represents a mis-identified seedling. Young seedlings of *Quercus rubra* versus *Q. coccinea* and *Q. falcata* versus *Q. velutina* can be difficult to distinguish if their leaves have been damaged by herbivores or are not fully developed. Third, repeated mutations can cause alleles to be identical in length even if not identical-by-descent. Finally, parentage analysis of seedlings will only reflect the proportion of hybrids at an early stage – selection over the 20-30 years until maturity may reduce their frequency among reproductive adults. On the other hand, not including heterospecifics as potential parents significantly increases the number of seedlings without parents within the mapped area. When *Q. falcata* individuals are excluded from the Duke Forest analysis the proportion of seedlings with two identified in-plot parents declines 7.5% and the proportion with no plausible in-plot parent increases 8%, which is significant given the fairly low percentage of seedlings with two identified parents (**Chapter 2**). It should also be remembered that the low genetic differentiation between morphologically defined taxa at these sites means that two "*Q. falcata*" adults might give rise to a seedling resembling *Q. velutina* if they themselves are the product of past introgression, and that likewise a "hybrid seedling" produced by a mating between parents identified as *Q. rubra* and *Q. coccinea* may not be a true F1. The issue of age and species identification could be resolved by planting large numbers of seedlings from known source trees and following them until their first adult leaves are produced (2-3 years, in a high-light environment).

Only a handful of studies to date have examined hybridization in red oaks and, to my knowledge, this is the first to employ parentage analysis to examine current hybridization rates. Salvini et al. (2009) employed a similar strategy of pairing genetic structure measures with parentage analyses in their examination of two potentially hybridizing white oaks, *Q. petraea* and *Q. pubescens*. However, their analysis focused on seeds collected from a small number of known mother seeds and therefore measured *initial* hybrid frequency. This study, in contrast, identifies large numbers of putative hybrid seedlings. Whether they will survive to reproduce is uncertain, yet the extremely low genetic differentiation between species at this site suggests that individuals of mixed genetic background have attained canopy status in the past. The high rates of hybridization suggested by my results and by those of previous studies indicate the need for closer examination of inter-specific gene flow in North American red oaks.

Oak species may exhibit four patterns with regard to hybridization: 1) complete reproductive isolation; 2) formation of a hybrid zone, where F1 hybrids are restricted to a narrow geographic area (perhaps along an ecotone); 3) introgression over a more widespread area, but where parental types are still identifiable; or 4) the formation of a hybrid swarm, where the majority of the population exhibit intermediate morphologies and/or mixed genetic characteristics. Morphological data suggests that this fourth type is rare – even in forests containing multiple oak species, taxa do tend to maintain some distinctness. However, complete isolation between closely related sympatric species may also be rather rare. Hybrid zones are most likely to be exhibited by species with strong differences in habitat requirement, where intermediate habitats exist between the parental ranges, such as has been observed between the sand-loving *Q. harvardii* and *Q. mohriana*, which occurs on nearby limestone cliffs (Muller 1952). In contrast, where species occur intermixed in a habitat that exhibits smaller-scale heterogeneity, such as in my study sites, introgression may be facilitated by the presence of suitable micro-habitats near trees bearing the hybrid offspring (Rushton 1993, Valbuena-Carabana et al. 2007), as well as by the proximity of heterospecifics. The rarity of individuals with intermediate phenotypes suggests that selection favors individuals with phenotypes resembling the parental types. Selection against intermediate types, however, may not form a severe barrier to the establishment of F1 hybrids, as the relationship between percent admixture and morphology is often weak in oaks (Valbuena-Carabana et al. 2007).

If hybridization is widespread in North American red oaks, this could have a number of consequences for their ecological and evolutionary dynamics, particularly under conditions of rapid environmental change. Effective population size is increased by hybridization, and large effective population size reduces the importance of genetic drift relative to selection on allele frequencies (Gillespie 2004). While wind-pollinated species such as oaks generally have high effective population sizes, they are not immune to the effects of fragmentation (Sork et al. 2002, Fernandez and Sork 2007); hybridization could maintain gene flow in highly fragmented landscapes by allowing heterospecifics to act as “stepping-stones”(Whittemore and Schaal 1991), and may even assist migration or range expansion through back-crossing (Petit et al. 2003). Hybridization may introduce genetic variation useful for local adaptation. In *Quercus wislizeni*, introgression from moisture-loving species was common in moist environments while in dry environments signatures of introgression from drought-tolerant species were more frequently detected (Dodd and Afzal-Rafii 2004). It is possible that hybridization between *Q. rubra* and the more drought-tolerant *Q. velutina*, as suggested by this study, might introduce traits into the former that would be favored in more arid environments; on the other hand, gene flow in the opposite direction might hamper local adaptation in *Q. velutina*, especially where this species is in the minority (Salvini et al. 2009). However, manipulative experiments would be necessary to test these hypotheses.

Given the slow growth and long lifespan of most oak species, the study of hybridization under natural conditions is challenging. Nevertheless, it is a subject deserving of further investigation, particularly given the potential for climate change to shift the selective pressures affecting the fitness of hybrids. Several classes of experiments would be particularly useful:

First, the fitness of hybrid offspring under natural conditions, the heritability of particular species traits, and the performance of pure and hybrid individuals under projected future conditions could all be assessed in common-garden experiments using seedlings of known parentage. Experiments of this nature would help to address the apparent contradiction between the moderate-to-high levels of introgression suggested by this and other studies and the stability of morphology observed in many oak species. Seedlings could either be produced through deliberate crosses, or by planting seeds from known mothers and using paternity analysis to identify the fathers. Although frequently used to assess heritability of traits or species and genotype performance under shared environmental conditions (Rehfeldt et al. 1999, Anekonda et al. 2002, Beckage and Clark 2003), the common garden approach has seldom been used to assess the effects of oak hybridization. In order to detect subtle differences in adult growth and fitness, plantings would have to be maintained for >20 years, which in many cases may be prohibitive, but more dramatic effects on early growth and survival should be evident within the span of a Ph.D. thesis.

Second, it is important to compare patterns of hybridization across environmental gradients and across a species' geographic range. It has been suggested that some oak species complexes exhibit latitudinal clines in hybridization (Rushton 1993) – which would make sense if longer growing seasons allow for greater differentiation in bloom time. Multi-year studies measuring flowering time and the percentage of seeds with heterospecific paternity in relation to spring temperatures would help to clarify the role of climate in reproductive isolation. We may also hypothesize that stronger local ecological differentiation between species would lead to less successful hybridization. In addition, while some species, notably *Q. petraea* and *Q. robur*, have been extensively studied, other oak taxa/species complexes, including the North American red oaks, have been largely neglected.

Finally, comparison of species and population differentiation at neutral vs. functional loci will help to identify which traits may be particularly important in a given local environment. Wind-dispersed tree species often show patterns of local adaptation in functional loci despite evidence of widespread geneflow in neutral loci (Mimura and Aitken 2010). Individuals heterozygous at particular functional loci may be at a disadvantage in either parental environment; backcrossed individuals homozygous at these loci, but bearing foreign alleles at other less important sites, may not suffer any loss of fitness. The development of genomic resources for Fagaceous species (http://www.fagaceae.org/) should aid in identifying species-specific genetic differences.

Chapter 6. Sensitivity of tree migration rate to variation in demographic rates and the presence of competitors

Introduction

During past periods of climate change, such as at the end of the last glacial period, organisms responded in large part through shifting their geographic distributions (Davis and Shaw 2001, Petit et al. 2002). While the magnitude of anthropogenic climate change is predicted to be similar - a 2.4-6.4°C increase in mean global temperature (IPCC 2007) - current warming is predicted to occur 10-100 times faster (Huntley 1991, Overpeck et al. 1991). This has given rise to concerns that many species, especially woody plant species with long life cycles, will not be able keep pace (Dyer 1995, Pitelka 1997, Neilson et al. 2005). Predicting species responses to climate change is challenging, due to uncertainty in estimates of dispersal ability and rates of climate change, the potential impacts of habitat distribution and biotic interactions, and the variation in population dynamics resulting from demographic stochasticity. In this study, I investigate the sensitivity of predicted migration rate for red oak (*Quercus rubra)*, a species generally regarded as highly dispersal limited, to changes in demographic processes and the presence of competitors and ask whether the longer dispersal distances indicated by genetic data (**Chapter 2**) are sufficient to allow species such as oaks to keep pace with predicted rates of climate change.

Dispersal distance is a key factor limiting the spread of plant species – no plant can grow where its seeds never arrive. Long-distance seed dispersal in particular is important for the colonization of new areas. Studies of the potential impacts of climate change on biodiversity usually consider scenarios in which plants either have an unlimited ability to disperse to new habitat or do not disperse at all (Pearson 2006). The truth surely falls somewhere between these extremes, but estimating real migration rates for plants is extremely challenging.

Pollen records suggest that the maximum migration rates for most North American and European tree species during deglaciation ranged from 100-2000 m/year (Huntley 1991, Pearson 2006), yet seed movement within extant forest communities is dominated by short-distance dispersal events. This disparity has come to be known as *Reid's paradox*, after calculations by Clement Reid (1899) showing that the spread of oaks into Great Britain at the end of the Pleistocene would require seemingly impossible seed dispersal distances. In fact, the situation is not so paradoxical if, instead of spreading through simple diffusion, plants tend to have "fat-tailed" (leptokurtic) dispersal kernels. A higher probability of long-distance dispersal (LDD), even if these events are rare, allows a population to move by leaps, as colonists are established far in advance of the main population front (Clark 1998, Higgins and Richardson 1999, Clark et al. 2001).

Empirical evidence indicates that seed dispersal in many tree species is indeed best described by leptokurtic probability distributions (Clark 1998, Clark et al. 1999, Streiff et al. 1999, Goto et al. 2006, Hardesty et al. 2006). As suggested by Reid (1899), even heavy-seeded species such as oaks can be transported long distances (>1 km) by birds (Darley-Hill and Johnson 1981, Johnson and Webb 1989, Purves et al. 2007). The dispersal distances required for post-glacial population expansion are further reduced if chloroplast haplotype data indicating the existence of small northern refugia are taken into account (McLachlan et al. 2005). In addition, recent estimates of contemporary gene flow in several tree species has suggested that dispersal distances may be longer than had previously been estimated (Bacles et al. 2006, Hardesty et al. 2006). My analysis for red oaks (*Quercus rubra, Q. velutina, Q. falcata,* and *Q. coccinea*) in North Carolina (**Chapter 3**) found that in some populations mean effective seed dispersal distances based on genetic data were more than ten times higher than estimates based on seed-trap data. On the other hand, it is uncertain whether these longer dispersal distances are sufficient to allow for rapid population spread when other biotic and abiotic factors are considered.

Many factors besides dispersal distance and climate influence invasion success and the realized rate of range expansion for a plant species. Where suitable habitat is fragmented, either because of natural heterogeneity (in the distribution of suitable soil types, for instance) or because of human activity, migration may be slowed because many seeds will land in areas unsuitable for growth. When a seed does reach a site suitable for establishment, it usually faces competition for resources from other plant species. Life history also plays an important role – all things being equal, early maturation and high fecundity (characteristics common in invasive species (Petit et al. 2004)) tend to lead to faster population spread. Finally, demographic stochasticity can lead to wide variation in migration speed even under similar environmental conditions.

Numerous simulation studies have been conducted to investigate the impact of habitat fragmentation on migration rate. While the model, distribution of suitable habitat, and form of the dispersal kernel all impact the threshold at which habitat loss severely impacts migration rate, these studies agree that habitat fragmentation can decrease migration rate, sometimes by an order of magnitude (Schwartz 1993, Malanson and Cairns 1997, Collingham and Huntley 2000, Higgins et al. 2003). Large barriers and long distances between fragments seem to have a particularly strong effect (Malanson and Cairns 1997). On the other hand, because certain animal dispersers will fly further in a fragmented landscape, wider spacing of habitat patches (up to a point) may actually increase effective dispersal distance in some plant species (Dyer 1995, Purves et al. 2007).

Few studies have attempted to model the effect of competition on range shifts. In part, this is because competitive interactions depend heavily on species identity and environmental conditions, making generalization difficult. Yet competition is known to have a strong influence on plant population dynamics, including the invasibility of an established community by new species (Cadenasso and Pickett 2001, Martin and Marks 2006). Some authors have suggested that the lack of response of some tree lines to warming may be partially due to competition between non-forest vegetation and tree seedlings (Masek 2001). Dullinger et al. (2004) studied the expansion of mugo pine shrubs (*Pinus mugo*) in an alpine area under different assumptions of dispersal probabilities, degree of climate change, and resistance of herbaceous communities to shrub establishment. They concluded the shape of the dispersal curve and spatial patterns of competition can affect range expansion dynamics as strongly as variation in future mean annual temperature. A recent simulation study by McMahon et al. (Submitted) found that migrating populations that readily "leapfrog" across an empty landscape barely move when disturbance is absent and the landscape is filled with a second "species" having identical demographic parameters because the local seed rain is so high that immigrant seeds have a very low probability both of finding a site vacated by a dead adult and of winning the competition for that site. Real forests consist of species with different environmental responses, and experience large-scale disturbances that increase spatial heterogeneity in resource levels, which should increase the probability that an immigrant can establish, but competition may still present an important barrier. There are currently no published studies modeling tree migration into an area dominated by a realistic, multi-species tree community.

Life history characteristics, including survival probabilities, fecundity, and time to sexual maturity may also affect migration rates (Malanson and Cairns 1997, Clark et al. 2001). If a colonist has a high probability of survival, matures quickly, and/or produces many offspring, it is more likely to contribute to the population's next leap forward. Many of these demographic processes are likely to be altered by global change. Loblolly pine (*Pinus taeda*) has been shown to mature at a smaller size and produce more seed when exposed to high CO_2 concentrations, such as are projected for the next century (LaDeau and Clark 2001). It is unknown how many species might respond similarly, but those that do are likely to have an advantage in reaching suitable recruitment sites. Fecundity in many temperate tree species also appears to be more sensitive to spring temperature and summer drought than are adult growth and survival (Clark Submitted). However, arid summer conditions (predicted to become more common in the southeastern US) have in some areas been shown to negatively impact seedling survival in both local species and potential migrants (Ibanez et al. 2008).

Demographic stochasticity can also limit migration, even where conditions for establishment are favorable. Clark et al. (2001) showed that projected migration rates are significantly lower when variation in reproductive success of the colonizing seed (resulting from stochastic survival) is taken into account. Although only variation in survival was considered in that analysis, seed production and growth can also vary substantially between individuals and between years, and this variation may reduce potential migration rates even further. In an experiment using flour beetles, Melbourne and Hastings (2009) found highly variable spread rates even under strictly controlled conditions, due to endogenous variation.

Using the SLIP forest simulator, which incorporates individual and annual variation in growth, survival, and reproduction, I address the following questions:

1) What is the sensitivity of dispersal rate and population growth to changes in mean dispersal distance relative to proportional changes in other life-history traits?
2) Given the effective seed dispersal probabilities estimated for red oak (*Q. rubra*) using genetic techniques, what is the maximum rate of population

spread for this species through an unoccupied landscape versus a landscape occupied by a diverse tree community?

3) How do potential rates of population spread for oaks estimated here compare to those required to track current climate change?

Methods

I used the SLIP forest simulator, described in two papers by Govindarajan et al. (2004, 2007), to model population growth and spread. SLIP is an individual-based, spatially-explicit forest simulator. Efficient algorithms, such as the use of monopole approximation for seed dispersal from distant sources, and the use of graphics hardware for quick parallel computation of light availability for individuals, allow this model to accommodate much larger landscapes and numbers of trees than its predecessors. SORTIE, the first spatially-explicit forest model, was constrained by light and dispersal calculations to areas of 9 ha; SLIP can accomodate total areas of up to 105 ha (1024 m x 1024 m) (Govindarajan et al. 2004). Landscapes are heterogeneous with respect to light, which varies according to the distribution and size of an individual's neighbors. Seed and 1st year seedlings are modeled as densities, whereas later stages (older seedlings, saplings, and adults) are treated as individuals, with distinct physical locations and attributes (diameter, canopy area, etc.). Individuals compete for light and space.

Equations describing growth, fecundity, and mortality are based on well-known principles in forest ecology (for instance, that fecundity depends on the size of an individual, and whether it is mature), and are parameterized for multiple species using over 270,000 tree-years of data from 11 plots in North Carolina. Unlike other forest simulators, SLIP incorporates uncertainty estimated at all stages of the hierarchical Bayesian model used to derive parameter estimates (McMahon et al. 2009). SLIP has been demonstrated to accurately recreate a number of forest dynamic processes, including tree ring increment over two centuries (McMahon et al. 2009), and has been used to investigate the role of individual variation (Clark et al. 2007), species sensitivity to climate (Clark Submitted), and intra- vs. inter-specific competition (Clark et al. In press) in species coexistence.

Sensitivity of migration to demographic processes

In order to determine the relative importance of dispersal distance for migration rate relative to other demographic characteristics, I conducted a series of sensitivity analyses. In baseline runs, the migrating species was assigned the demographic parameters estimated for *Quercus rubra* based on non-genetic measurements, including seed-trap data (Clark et al. In press). In “treatment” runs, mean dispersal distance (Disp±), fecundity (Fec±), and size at ½ chance of maturity were increased or decreased by 50%. The probability of maturity at a given size is altered either by changing the slope of the relationship (Mat 1±) or the intercept (Mat2±) (**Appendix 2.1**). I then determined whether changes in migration rate, distance of the furthest-forward individual from the origin, and final population size were greater than or less than 50%. We would expect that increasing dispersal distance and fecundity and decreasing size at maturity would lead to more rapid migration and population growth. Combinations tested are shown in table 13.

Table 13: Parameter sets for sensitivity analysis

	u	α_0	β_0	β_1	β_2
Baseline	34.9	3.77	-2.89	0.0819	0.0579
Disp+	**78.36**	3.77	-2.89	0.0819	0.0579
Disp-	**4.64**	3.77	-2.89	0.0819	0.0579
Fec+	34.9	**4.175**	-2.89	0.0819	0.0579
Fec-	34.9	**3.077**	-2.89	0.0819	0.0579
Mat1-	34.9	3.77	-2.89	**0.1638**	0.0579
Mat1+	34.9	3.77	-2.89	**0.0546**	0.0579
Mat2-	34.9	3.77	**-1.445**	0.0819	**0.02895**
Mat2+	34.9	3.77	**-4.335**	0.0819	**0.08685**

There were a total of 9 treatments, including baseline. Landscapes consisted of an empty 1024 m x 1024 m square. Runs were initialized with 10 adult trees placed at random within a 20m x 20m square at the lower left corner of the landscape. Adult sizes were drawn from a uniform distribution 20 cm to 45 cm (**Appendix 2.2**). This size range was chosen to simulate a small group of recently mature or soon-to-be mature trees. Each simulation was run 3 times under 3 different initial conditions, a total of 9 times for each set of parameters over a 50-year timespan. Sizes and positions of all saplings and adult trees were recorded at 5 year intervals. I identified the furthest-forward individual at any given timestep. Average annual spread was calculated over the 30 years of the simulation; over the full 50 years the main edge of the population advanced too close to the edges of the plot in some simulations, resulting in unrecorded dispersal beyond the edge of the plot.

Oak migration for genetic dispersal estimates under non-competitive conditions

Estimates of effective seed dispersal distances were obtained for a population of red oak located in Duke Forest in the North Carolina Piedmont, as described in **Chapter 2**. The dispersal parameter u estimated using genetic data from Duke Forest was 6300, corresponding to a mean dispersal distance of 124.7 m. This dispersal kernel is based on seedling data, and thus describes *effective*, rather than initial, seed dispersal. However, it does capture secondary dispersal by animals, which estimates based on seed-trap data do not. If seedling survival is not a function of distance from the mother tree (currently an assumption of SLIP), then we would expect the initial seed dispersal kernel to have a similar shape to that based on seedling distribution. All seedlings at the study site were within 60 m of an adult oak (not necessarily their parent), and annual survival was high, so distance-dependent mortality can be assumed to be fairly weak. Using seedling genetic data may somewhat underestimate the frequency of short-distance dispersal events, if initially high seedling densities near parent trees lead to increased mortality, while precise detection of long-distance dispersal events depends on the size of the mapped stand (12.5 ha, in this case). Nevertheless, I regard this estimate as a reasonably accurate one for a high-dispersal site, where many offspring were found to be located >100 m from their parent and many were found to lack an in-plot parent altogether (**Chapter 3**).

By contrast, the “baseline” dispersal parameter, derived from seed-trap data, is 39.4, corresponding to a mean dispersal distance of only 9.27 m. Acorns caught in seed traps have been scattered by gravity, but have not yet been dispersed by squirrels or jays. In testing the effect of this increased dispersal estimate, all other parameters were held constant. As in the previous experiment, simulations were conducted using an empty 1024 m x 1024 m landscape, with 10 adult trees placed in the lower left corner, and run-times of 50 years. Results were compared to the “baseline” treatment of experiment 1.

Oak migration under competitive conditions

Rather than spreading through an empty landscape, as in the first two experiments, in this simulation the migrating population must spread through a forested landscape (figure 25). A 512m x 512 m forest landscape was generated by “planting” a complete copy of the mapped stand from Duke Forest into an empty landscape and allowing trees to generate seedlings and saplings, gradually filling in the open space over a period of 220 years. Three “background forests” were generated in this manner. All background trees were removed from the 20 m x 20 m at the lower left to make room for ten individuals of the migrating species (figure 25).

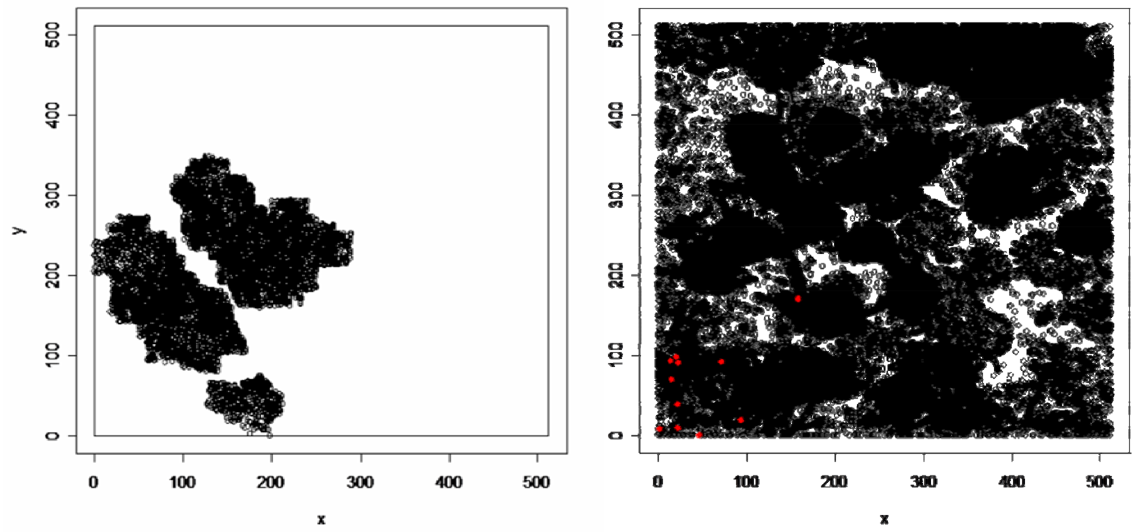

Figure 25: a) Duke Forest trees "planted" in empty landscape; b) example "competition with disturbance" landscape year 150. Migrant trees in red.

Simulations were run 9 times (3 "starting forests" x 3 initial migrant configurations) under two disturbance conditions: no disturbance (other than individual mortality) and periodic large-scale disturbance. In the "disturbance" simulations, every 30 years beginning at year 1 all trees were removed from several randomly-placed circular patches 60 m in diameter. The number of patches was randomly generated from a uniform distribution (6 to 25), so that between 6.5% and 27% of the canopy area was removed at 30 year intervals. This disturbance frequency and percent disturbance was chosen to approximate the historical frequency of large-scale wind disturbance in east-coast deciduous forests (Runkle 1982, Greenberg and McNab 1998, Lorimer and White 2003). Seedlings and saplings were left intact, as previous studies have shown that most regeneration in canopy gaps comes from seedling and sapling banks that survive the disturbance event (Holladay et al. 2005, Dietze 2006, Shure et al. 2006). Because I was simulating spread through a continuous forest, edge-light was set to sub-canopy levels, whereas in previous non-competitive situations landscape edges received full sunlight. Simulations were run for 151 years.

Results

Sensitivity of migration to demographic processes

Results of the sensitivity simulations are shown in table 14. The measure which showed the greatest response to changes in vital rates was population growth. All populations were still undergoing exponential growth at the end of the simulation. Increasing mean dispersal distance by 50% resulted in an average final population more than three times greater than the baseline, while reducing mean dispersal distance by 50% led to an 86% decrease in mean final population size. Responses of population growth to changes in fecundity were approximately proportional. Decreasing size at maturity led to a less-than-proportional increase in average final population size. Increasing size at maturity decreased population size proportionally for Mat1+ but increased it slightly for Mat2+.

Table 14: Results of sensitivity analysis

Param set		#adults, yr 50	Max distance, yr 50 (m)	Migration (m/yr), yr 1-30
Baseline	Range	111 – 4,481	654,6 - 1,267.9	5.9 – 31.2
	Mean	1,901	950.5	18.8
Disp+	Range	139-10,874	838-1,172.3	13-34.7
	Mean	6,470	994.6	24.4
	% Δ	**+240.4****	**+4.6**	**+29.9**
Disp-	Range	13 – 1,435	168.8 – 843.3	4.8 – 27.3
	Mean	264	537.2	13.7
	% Δ	**-86.1***	**-43.5**	**-27**
Fec+	Range	45 – 10,822	599.2 – 1,208	9.6 – 26.4
	Mean	3,146	875.4	15
	% Δ	**+65.5***	**-7.9**	**-20.2**
Fec -	Range	60 – 3,242	295.5 – 1,114	5.9 – 28.4
	Mean	949	728.5	15.4
	% Δ	**-50.1**	**-23.4**	**-18.1**
Mat1+	Range	29 – 2,210	499.5 – 1,346	10.1 – 44
	Mean	1,003	811.2	18.1
	% Δ	**-47.2**	**-14.7**	**-4**
Mat1-	Range	54 – 5,095	500.1 – 1,070.8	8.2 – 20.8
	Mean	2,097	824.4	16.4
	% Δ	**+10.3**	**-13.3**	**-12.6**
Mat2+	Range	49 – 7,353	412.9 – 1,330.5	11.1 – 20.7
	Mean	2,058	900	14.6
	% Δ	**+8.3**	**-5.3**	**-22.1**
Mat2-	Range	43-4,935	477.6 – 1,193	6.8 – 37.0
	Mean	2,611	823.2	17.4
	% Δ	**+37.4**	**-13.4**	**-0.08**

The response of annual migration rate was somewhat less than proportional for all parameter sets. Again, dispersal distance had the strongest effect. Simulations in which fecundity increased or size at maturity decreased saw a reduction in spread rate, contrary to expectations. The furthest distance attained during the 50 year simulation was only moderately affected by changes in vital rates, with decreased dispersal having the largest effect, followed by decreased fecundity.

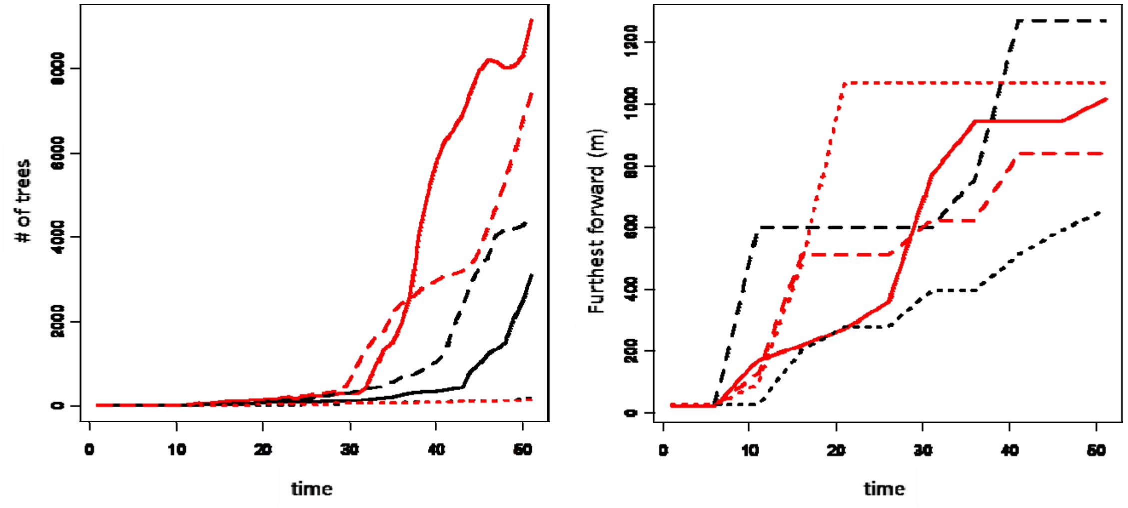

Figure 26: Total population size (left) and distance of furthest-forward individual from origin (right) over six separate runs.
Black – baseline parameters. Red – Disp +. Solid – initial condition 1. Dashed – initial condition 2. Dotted – initial condition 3.

There was broad overlap in all three measures between sets of parameters, reflecting the effect of demographic stochasticity. Results differed between sets of initial conditions for some measures. The third initial population, consisting of slightly smaller individuals than the other two (mean DBH 30.12 cm, vs 32.26 & 34.53 cm), never grew larger than 200 individuals, while the highest population size attained by initial conditions 1 and 2 were 12,817 and 10,874 – both under high-dispersal conditions. However, this did not affect the ability of the third initial population to expand spatially:

annual spread rates and furthest distance attained overlap broadly for all three sets of initial conditions (Fig 25).

Oak migration for genetic dispersal estimates under non-competitive conditions

As suggested by the sensitivity analysis, the larger estimated mean dispersal distance estimated from genetic data had a strong effect on all three measures, but particularly on final population size. Final population sizes ranged from 203,614 to 435,539 (mean 298,187): a more than150-fold increase, compared to a roughly 14-fold increase in mean dispersal distance. Annual migration rate over the first 30 years increased 124%, to an average of 42.1 m/year. In fact, this may be something of an underestimate of migration speed, as within 30 years the population had spread close to the edges of the plot (mean distance from the origin 1,287 m). Average migration rate over the first two decades was 55.7 m/yr. By the end of 50 years, the population had filled more than a quarter of the landscape, and the furthest-forward individuals had reached the upper right corner (average maximum distance of 1,433 m from the origin, compared to maximum possible distance of 1,448 m).

Oak migration under competitive conditions

The presence of competitors severely impacted simulated oak population dynamics. The most striking change was in population growth. Instead of increasing exponentially, the number of adult migrants was, on average, less than the starting number after 150 years (table 15); although the number of trees periodically climbed above ten, numbers of adults > 25 were not observed, because other species fill most of the available spaces in the canopy. In one simulation with disturbance, the number of migrants dropped to zero by year 120 (extinction).

Table 15: Results for individual model runs – competition. Over full 150 years.

Param set	**IC**	**Final # adults**			**Max distance(m)**			**Migration rate (m/yr)**		
		f1	**f2**	**f3**	**f1**	**f2**	**f3**	**f1**	**f2**	**f3**
No disturb	**1**	7	17	6	160.1	260.98	449.2	0.91	1.58	2.84
	2	6	4	1	327.9	309.1	108.2	2.0	1.91	0.57
	3	4	8	6	99.8	576.7	188.5	1.49	3.68	1.09
	mean	**6.55**			**274.9**			**1.79**		
	sd	**4.41**			**159.8**			**0.98**		
Disturb	**1**	5	5	18	309.3	375.6	270.7	1.9	2.35	1.65
	2	1	10	8	89.6	247.1	382	0.45	1.5	2.4
	3	0	8	6	0	135.2	68.96	0	0.78	0.29
	mean	**6.78**			**176.9**			**1.26**		
	sd	**5.3**			**129.1**			**0.091**		

Nevertheless, in most case the migrants were able to advance through the landscape. This advance was somewhat erratic, because not only were there fewer opportunities for establishment, but the furthest-forward individual often died before a new tree established ahead of it, causing the “edge” of the population to fall back. Consequently, the average annual advance was considerably lower than in the absence of competition (tables 15 & 16). Canopy disturbance did not lead to faster migration rates over the long run, although within the first 50 years it did seem to aid population establishment (table 16).

Table 16: Competition vs. no competition (averages over all runs).

Param set	Average density (trees/km^2), yr 50	Furthest forward individual (m), yr 50	Migration rate (m/yr), yr 1 to 30
Competition – Disturbance -	284,373.3	1433.0	42.1
Competition + Disturbance –	29.8	224.7	2.82
Competition + Disturbance +	40.4	248.8	3.9

Discussion

Changes in phenology and geographic ranges consistent with predicted responses to climate change have been observed in many taxa, including woody plants (Walther et al. 2002). But will these responses be sufficient to allow species to track climate change? Here I discuss the results of my simulations in the context of the challenges inherent in estimating both migration ability and the speed of migration necessary to keep species ranges in equilibrium with a changing climate.

My results confirm that increased dispersal distances derived from genetic data can result in faster rates of spread and population growth. Population growth rate is much more sensitive to changes in demographic parameters than are the two measures of population spread rate; the long-distance jumps allowed by a fat-tailed dispersal kernel enable even small populations to rapidly expand their spatial extent. The effect of dispersal on population growth was striking. Long dispersal distances increase the probability that a seed will reach an open site where, relative to the shaded environment near its mother, it can grow more quickly and begin contributing to population growth at an earlier age. In the presence of competition, population growth and migration rate were severely reduced. However, the longer dispersal distances indicated by genetic data did permit *Quercus rubra* to persist and spread under competitive conditions. Although trees with "baseline" *Quercus rubra* parameters (ie, short dispersal) were initially abundant in the forest matrix, they went extinct in 66.7% of simulations with disturbance and 77.8% of those without (**Appendix 2.2**), indicating that they were unable to reach suitable recruitment sites with sufficient frequency. The long-dispersal *Q. rubra* migrants, starting with only 10 individuals, went extinct in only one out of 18 simulations.

Somewhat surprisingly, the addition of disturbance did not improve migration rate or population growth – average migration rate and final population size were actually slightly lower over the full 150 year simulation. Although red oak seedlings do respond favorably to higher light levels, other species such as red maple tend to respond faster (Clark et al. 2003), which may have negated the positive effect of gaps. Indeed, experiments have shown that competition from other seedlings and saplings can hamper red oak recruitment even under high-light conditions (Buckley et al. 1998, Elliott et al. 1999).

Variation introduced by demographic stochasticity and by small changes in initial conditions results in a broad range of potential end-points for a population. Within a given set of initial conditions and in the absence of competition, mean annual spread rate over 30 years may vary more than 3-fold, while final population size may vary more than 13-fold. Initial conditions also had a strong effect on population dynamics. Very high variability means that while a higher mean dispersal distance may make a species more likely to spread quickly, precise prediction of population growth and spread is impossible and that, by chance, a given population with a high mean dispersal distance may actually spread more slowly than a population with a shorter mean dispersal distance (figure 26). Competition reduced variation in migration rate and population growth, if only because both remained low in all simulations.

Predicting how far and fast a species might *need* to migrate is nearly as complex as predicting how fast it *can* migrate. Anthropogenic climate change is expected to produce a 2.4-6.4° shift in mean global temperature over the next century (IPCC 2007). By contrast, during the last deglaciation, Europe warmed by ~2°C per millennium (Huntley 1991). If, following Huntley (1991), one assumes a temperature gradient corresponding to ~ 1°C change in mean temperature per 125 km change in latitude in temperate regions, and a 4°C change in mean temperature over the next century, then a plant species would have to travel roughly 5,000 m/year to remain on the same temperature isocline. The required speed, however, depends strongly on both topography and current climate. Loarie et al. (2009)estimated that while the global mean migration speed needed to track climate change is 420 m/year, species in mountainous habitat need only migrate 8 m/year, while species in mangrove forests or flooded grasslands would need to migrate 1.26 km/year. Malcolm et al. (2002), using an array of seven global climate models (GCMs) and two global vegetation models (GVMs), defined migration distance as the distance between each cell in the current climate and the nearest cell of the same biome type at year 100. Required migration rates were less than 315 m/yr for 61-71% of grid cells, but a significant proportion of grid cells would require migration rates of over 1,000 m/year. A similar result was obtained by Iverson and Prasad (2001), who estimated potential changes in suitable habitat for 80 common tree species in the United States, and found that the optimum latitude of suitable habitat moves >100 km north by 2100 for 30-35 species under the Hadley and CCC models. McKenney et al. (2007) noted that while the mean centers of future climate envelopes for 130 North American tree species are predicted to shift 700 km northward, overlap in current and future suitable range is extensive for many species; however, range sizes would decline for many in the absence of migration. Required altitudinal shifts, on the other hand, may be as little as 8-10 m per decade (Walther et al. 2002).

Annual migration rates estimated for red oak in this study in the absence of competition ranged from 7 to 26.8 m/yr based on seed-trap estimates and 38 to 67 m/yr based on genetic estimates. In the presence of competition, these rates were reduced to less than 5 m/yr. Compared to the rate of latitudinal migration required to keep up with climate change estimated above, these rates are low, though probably sufficient for altitudinal shifts.

It should be noted that both seed trap and single-site genetic dispersal data may underestimate the frequency of extreme long-distance dispersal events, the former because only initial gravity-dispersal is captured, the second because it is impossible to determine how far individuals originating outside the plot may have travelled. Jays can transport acorns several kilometers (Darley-Hill and Johnson 1981, Purves et al. 2007).

Still, the effect of such events may be limited given the low probability of survival to maturity for any given acorn. Dyer (1995), assuming that oaks could spread into any suitable area within 200 m of an established population and experience two random long-distance bird-mediated colonization events within a generation, estimated a maximum migration rate of only 136 m/yr. However, if seeds of a native species "invading" a distant habitat enjoy a release from natural enemies this could counter the numerical advantage of the residents, allowing swifter spread. Reduction in herbivore damage has been documented in some exotic invasives (Wolfe 2002, Jogesh et al. 2008), but replacement of one set of enemies with another can negate the effect (Agarwal et al. 2005) and the hypothesis has not been tested in the context of climate-driven range expansion. It should also be noted that my simulations (like many of the other migration simulations cited) assumed constant mean climatic conditions. Changes in temperature or precipitation could either promote or retard migration, depending on whether the effect on demographic rates of the migrant relative to the residents was positive or negative.

For some species of oak, climate change is predicted to open up wide swaths of newly-suitable habitat: a 8.7 degree latitudinal shift in the northern limit of *Q. rubra*, and shifts of > 7 degrees for *Q. alba*, *Q. palustris*, and *Q. macrocarpa* (McKenney et al. 2007). However, limited dispersal could prevent these species from realizing that potential. Iverson et al. (2004) calculate that for 5 species, including southern red oak (*Q. falcata*), less than 15% of the habitat predicted to become suitable in the next 100 years will be colonized by that time, if species follow an average migration rate of 500 m/yr, and probability of colonization of an unoccupied cell depends on habitat quality and abundance of the species in nearby cells. Furthermore, even under full-dispersal scenarios, many tree species are predicted to experience contractions in area of suitable habitat, including *Q. nigra* (-54.7%), *Q. phellos* (-55.2%), *Q. coccinea* (-50.1%), and *Q. virginiana* (-81.8%). The creation of novel combinations of temperature, precipitation, and substrate, may also mean that in some areas the most likely potential migrants may not be much better adapted than the current residents to future conditions (Ibanez et al. 2008).

A particular challenge for oaks is that many species exhibit low recruitment success even where they are abundant in the canopy, a problem frequently attributed to inappropriate disturbance (fire) frequency, but which may be exacerbated by high deer herbivory or other factors (Crow 1988, Abrams 1992, Elliott et al. 1999). Migrating oaks may experience a similar lack of recruitment sites, as my “competition” simulations illustrate: opening gaps in the canopy does not necessarily eliminate competition in the understory, and faster-growing species such as maple or tulip-poplar can easily overtop oak seedlings when light resources increase. Forest managers concerned about the response to climate change must continue to pay particular attention to the understory level – reduction in herbivore numbers or changes in management practices (harvesting frequency, controlled burns, etc.) may be necessary to make sure that there is a diverse seedling bank (of current residents as well as potential migrants) available to take advantage of canopy gaps when they occur (Pedersen and Wallis 2004, Spetich 2004).

Although woody plants may seem to be at a disadvantage under rapidly shifting climatic conditions due to their slow maturity and low reproductive rates, their other life history traits, including longevity and the ability of adults to survive periods of adverse climate, may promote the maintenance of populations at the trailing edge that can serve as seed sources, and allow rare long-distance migrants to persist until the rest of the population catches up with them (Malcolm 2002). Such traits may also buffer woody plant species against the loss of genetic diversity (Hamrick 2004). In addition, many tree species can be grown much further north than their current range limit, so cultivated populations (along with small natural "northern refugia", if they exist) could potentially serve as seed sources, much as ice-edge refugia did in the past (McLachlan et al. 2005, Van der Veken et al. 2008).

Climate envelope models provide an estimate of potential future range of a species as defined by abiotic conditions. For many species, these "envelopes" represent the best available approximation of their climatic niche (Pearson and Dawson 2003). However, because climate envelopes are based on static measures of abundance rather than on population dynamics, they may not fully reflect the sensitivity of different tree species to *changes* in climate (Clark Submitted). Where sufficient longitudinal data is available, it would be desirable to re-evaluate the requirements for population persistence – and therefore the magnitude of range shifts that may be necessary - based on demographic responses to environmental conditions.

Improved estimates of dispersal kernels will be essential in predicting what proportion of newly suitable areas (whether defined using climate envelopes or measures of demographic sensitivity) could potentially be colonized within a given timeframe. The assumption of equal migration rates across a species range is likely to be unrealistic (Thuiller et al. 2008), so comparison of dispersal kernels under different environmental conditions (disperser density or topography, for instance) would be extremely helpful. I used here genetic dispersal estimates from a population of red oak in the Piedmont of North Carolina that were nearly 10 times higher than similar genetic estimates for a population in the Southern Appalachians, which was found to conform more closely to seed trap-based dispersal estimates (**Chapter 3**). However, understanding dispersal processes will not be enough. Among the models most frequently used to predict changes in species range, few deal explicitly with dispersal processes and demographic stochasticity, let alone the altered dynamics of leading and trailing edge populations, species interactions, or interactions between climate and land-use (Thuiller et al. 2008). Given all these sources of uncertainty, we cannot hope to precisely predict the future distribution of a species or the composition of a forest, but by quantifying uncertainty, and explicitly taking it into account when making forecasts, we can better understand the range of possible outcomes (McMahon et al. 2009) and the interventions that may be necessary to achieve biodiversity conservation goals.

Chapter 7. Conclusion

Summary

Predicting and managing the ecological and evolutionary responses of plant species to climate change (and other environmental perturbations) will require a better understanding of the scales of dispersal and gene flow in natural populations, of the interactions between dispersal and the various biotic and abiotic factors which may limit the success of immigrant seed and pollen, and of the effect of unequal reproductive success and hybridization on effective population size and the maintenance of genetic diversity and adaptive potential. This dissertation provides new insights into seed dispersal, reproductive success, and inter- and intra-specific gene flow in red oaks; the variation in these processes from site to site; and their influence on genetic structure, population dynamics, and migration potential.

Combining genetic and ecological data through hierarchical Bayesian modeling is a powerful tool for reconstructing pedigrees and estimating dispersal kernels for species, such as oaks, in which reproductive processes are highly cryptic. The model described in **Chapter 2** is the first such model specifically developed for a monoecious plant, the first to treat dispersal both within and from outside a mapped stand as a consistent process, and one of the first to explicitly consider two common sources of genotyping error. I showed that, at Duke Forest, effective seed dispersal distances are considerably greater than the dispersal distances captured by seed trap data, as would be expected for species in which secondary dispersal by animals plays an important role. Furthermore, I show that even in a stand 12 ha in size, a majority of seedlings have fathers located beyond the borders of the mapped plot, suggesting a very high capacity for gene flow via pollen.

In **Chapter 3**, I demonstrate that effective dispersal distances vary significantly between Duke Forest and Coweeta. At the latter site, effective dispersal distances were not much greater than those estimated from seed trap data alone and siblings were frequently clumped near the maternal tree, suggesting that gravity dominates the dispersal processes, whereas at the former, many seed dispersal events of >100 m were detected and related individuals were often widely spaced, indicating substantial secondary dispersal. Differences in dispersal between the two sites could not be explained by differences in squirrel density, and may be due to the influence of topography on disperser behavior, differences in other biotic relationships (such as with herbivores or seed predators), or interactions between the distribution of suitable recruitment sites and mast-year history. Patterns of spatial genetic structure at both sites were consistent with site history and observed patterns of gene flow: Coweeta, where timber harvests were selective and dispersal distances short exhibited stronger spatial genetic structure among adults than Duke Forest, where dispersal distances are long and much of the landscape was cleared for farming.

In **Chapter 4**, I estimated reproductive success for individual trees, tested several hypotheses regarding the causes of variation in reproductive success, and examined its effect on effective population size. This is one of only a handful of studies to date to use pedigree information to estimate reproductive success (defined as seedling, rather than seed, production) in a forest tree. I found that annual seed production is positively associate with reproductive success (*R*), but that highly fecund trees have lower germination rates, suggesting the effect of density dependent mortality. Age exhibited a hump-shaped relationship with *R*, as did growth rate, suggesting that senescence and tradeoffs in resource allocation may play a role. Inequality in reproductive success was high. Reproductive success for individuals with at least one offspring followed a lognormal distribution, as predicted by Herrera and Jovani (2010), but many individuals produced no offspring at all. Simulations revealed a strong sigmoidal relationship between effective population size and the Gini coefficient of the population. Given this relationship, and the observation that ~50% of "adult" trees are reproductively active, the effective population size for red oaks was estimated to be approximately ¼ of the total population size. This has important consequences for the strength of drift versus natural selection for fragmented oak populations: populations of < 50 oaks are likely to be vulnerable to the loss of genetic variation.

Hybridization, too, can have important effects on effective population size, the distribution of genetic variation, and the evolutionary potential of the species involved. In **Chapter 5**, I examined evidence for the extent of hybridization in my two mixed-species stands. This is one of only a handful of studies to have investigated hybridization in North American red oaks using genetic information. I found very little genetic differentiation between morphologically-defined species, as well as a relatively high proportion of seedlings identified by parentage analyses as putative hybrids, suggesting fairly extensive, ongoing geneflow between coexisting red oak species. I suggested that at these sites, where species grow closely intermixed, both pollen transfer and the survival of hybrid offspring may be enhanced, as hybrid individuals with are likely to have similar fitness to those with parental trait combinations. On the other hand, the maintenance of characteristic morphological trait combinations suggests that selection may resist introgression in at least some key functional loci.

In **Chapter 6**, sensitivity analyses of an individual-based simulation model confirmed that dispersal is one of the most important life history traits affecting potential migration rate and population growth, with fecundity ranking second. However, demographic stochasticity and minor differences in initial conditions can lead to a wide variety of population trajectories. I also found that, in the absence of competition, the longer dispersal distances estimated using genetic data from Duke Forest did result in rates of population spread significantly more rapid than seed trap-based estimates. This is one of very few simulation studies of migration potential to have examined the effect of competition from a diverse plant community. Competition severely restricted the potential for migration and population growth, reducing rates of population advance to less than 4 m/yr but, all else being equal, both persistence and migration were enhanced by longer dispersal distances. I found that even under best-case scenarios migration rates estimated for oaks were rather slow compared to projected rates of climate change. However estimates of “necessary” migration rates based on climate envelopes may not accurately reflect the sensitivity of species to climate change, as they are based on abundance rather than population dynamics.

Implications for oak responses to climate change

My results demonstrate that that the extent of gene flow via seed in oaks (as reflected in effective dispersal distances) can be much greater than seed-trap-based estimates had previously indicated. This suggests that many oak populations may be limited much more by the availability of suitable recruitment sites than by dispersal, and that the capacity of oaks – and perhaps other animal-dispersed forest trees - to shift their geographic ranges in response to climate change may be better than is generally thought. Simulation model analyses support both of these hypotheses. However, I also show that oak populations can vary widely in effective dispersal, with patterns of seedling recruitment being similar to gravity-dominated seed shadows in some, while others exhibit longer parent-seedling distances characteristic of dispersal by animals. Improving the availability of suitable recruitment sites (areas with canopy gaps, low understory competition, and low-to-moderate deer herbivory) is likely to promote both local persistence and range shifts in oaks.

Effective pollen dispersal was extensive at both study sites. My results also suggest high levels of hybridization between the focal species, which could have important consequences for responses to global change, if between-species mating either introduce useful genetic variation (increasing adaptive potential) or reduce reproductive success (resulting in lower recruitment). Both long pollen dispersal distances and hybridization are expected to increase effective population size, buffering populations against the loss of genetic diversity to drift or strong selection. This is important given that overlapping generations and variation in reproductive success can reduce N_e significantly.

Future research directions

Several lines of research will be helpful in understanding the role of dispersal and gene flow in the ecological and evolutionary dynamics of oak populations, and the likely responses of oak species to climate change:

1) **The sensitivity of oak demographic rates to climate.** Some tree species exhibit strong "tracking" of climate variables (spring temperature, summer drought) in their fecundity or survival, while others are relatively insensitive; furthermore, these responses are poorly correlated with predicted climate responses based on climate envelopes (Clark Submitted). Demographic sensitivity will determine the extent of the range-shifts and/or local adaptation that would be required to keep a species more or less in equilibrium with climatic conditions, but the long-term data needed to estimate demographic responses are available for relatively few sites and species.
2) **The frequency of hybridization between North American red oaks, and the effects of introgression on fitness and responses to the environment**. Compared to white oaks, hybridization in red oaks is very little studied (Aldrich et al. 2003b, Dodd and Afzal-Rafii 2004, Tovar-Sanchez and Oyama 2004, Penaloza-Ramirez et al. 2010), although lower between-species genetic

differentiation (Guttman and Weigt 1989) may allow more opportunity for hybridization in this group. Similarly, the role of the environment in promoting or limiting inter-specific gene flow is still poorly understood.

3) **Physiological and environmental controls on the amount and timing of pollen production in oaks.** The amount of pollen or male flowers produced can have a strong effect on fitness in plants (Janzen 1975, Klinkhamer and de Jong 1993), yet I was unable to find any published studies relating amount of individual male reproductive effort in oaks to either female reproductive effort or characteristics such as size and canopy area. Similarly, although the timing of flowering may strongly affect the potential for hybridization, few studies have examined the timing of pollen release in co-occuring oak species, either in the context of climate effects or its influence on hybridization.
4) **The role of North American jay species as long-distance dispersers of acorns.** Purves et al. (2007) found that a model including both climate and European jay behavior explained the distribution and abundance of three Spanish oak species significantly better than a model based on climate alone. Blue jays and scrub jays in North America exhibit similar acorn caching behaviors to European jays – shallow burial, a preference for disturbed or early successional habitats (Darley-Hill and Johnson 1981, DeGange et al. 1989, Johnson et al. 1997, Gomez 2003) - and it has been suggested that they may have played an important role in post-glacial range expansions of oaks (Johnson and Webb 1989). The behavior and abundance of this group of birds

may therefore also be important to meta-population dynamics and migration ability of oaks today (Dyer 1995), but more detailed studies of the behavior of North American species and their impact on oak population dynamics are needed.

5) **The effect of escape from natural enemies or lack of mutualists on the competitive ability of long-distance migrants in native plant species.** Escape from predators and pathogens is one reason that has been proposed for the rapid population growth and spread exhibited by species in their invasive range relative to their native range. While this hypothesis has been tested for some (mostly herbaceous) invasive species (Agarwal et al. 2005, Jogesh et al. 2008), it is not known whether enemy escape might increase the probability of establishment for long-distance migrants of native tree species.

6) **More generally, the role of interactions between plants and their various enemies and mutualists in shaping patterns of dispersal and establishment**. In tropical forests, the loss of large frugivores to hunting can have severe impacts on seed dispersal (Terborgh et al. 2008). Likewise, changes in the abundance or behavior of temperate dispersers due to human impacts or geographical and temporal variation could have important effects on the dispersal patterns of their associated tree species, while artificial increases in herbivores such as deer can severely reduce seedling recruitment (Rooney 2001, Ibanez et al. 2008). Changes in climate can also affect the

dynamics of pest and pathogen species (McMahon et al. 2009). These interactions are complex, and deserving of further study.

Conclusions

Above all, I hope readers will come away with a greater understanding of, and appreciation for, the importance of interactions between ecological and evolutionary dynamics and for the link that dispersal and gene flow provides between the two. These processes are still poorly understood even for "charismatic" tree species such as oaks, yet they are likely to be critical to predicting and managing the effect of climate change on forest species. The response of an organism to environmental change depends on how its population dynamics are affected by shifts in the distribution of resources, the intensity of stressors such as drought, and interactions with mutualists, predators, and competitors. The resulting changes in population size or mating and recruitment patterns affect the amount and distribution of genetic variation within a species which in turn can feed back to population dynamics by affecting adaptive potential and population dynamics. Better information about gene flow and the genetic basis of ecological traits will be important in answering many of the questions faced by ecologists and managers in this era of rapid environmental change: "What individual or site characteristics promote reproductive success in species x?" "Will hybridization between sub-species due to range shifts or assisted migration increase or decrease population viability?" "Is invasiveness in exotic species enhanced by different selection pressures within the introduced range?" "What is the effect of changes in disperser community on plant population dynamics?" The next twenty to fifty years will be an exciting time, as ecologists and evolutionary biologists continue to develop the tools and knowledge to understand and manage the effects of global change on species, communities, and ecosystems as those effects unfold.

Appendix 1

1.1 DNA Extraction and PCR Protocol

DNA extraction was via a modified CTAB procedure (Murray and Thompson 1980, Doyle and Doyle 1987). One-half gram of fresh leaf tissue was placed in each tube of a 96-well Costar plate and frozen at -80°C until the time of extraction. Tissue was then disrupted by dipping the plate in liquid nitrogen, and shaking frozen tissue with small steel balls in a Genogrinder 2000 for 4 minutes. I then added 540 µL of CTAB buffer, and ground the mixture for 4 minutes more. The plate was incubated at 65°C for 50 minutes, inverting periodically. After centrifuging, 430 µL of clear liquid was transferred to a new plate, and 400 µL of a 40:1 chloroform:isoamylalcohol mixture was added. This mixture was mixed thoroughly in the Genogrinder (slow speed, 3 minutes), and centrifuged at 4000 rpm for 10 minutes. The upper aqueous layer (280 µL) was then transferred to a new plate, 500 µL of 100% ethanol was added, and the plate was placed in a -20°C freezer for at least 30 minutes to precipitate the DNA. The tubes were then centrifuged at 4000 rpm for 10 minutes, and the liquid drained. The pellet was washed in 200 µL 70% ethanol, dried overnight (or 3 hours in warm drying oven), and dissolved in 60-80 µL of autoclaved distilled water. The DNA solution was stored in the refrigerator until needed for PCR.

PCR protocol:

Primers were developed by Aldrich et al. (2002, 2003a). Sequence, annealing temperature, size range, and observed and expected heterozygosities are shown in table 17.

100 µg DNA was combined with 3.2 µL Promega 5x flexi-buffer, 1.2 µL 25 mM MgCl, 1.2 µL 0.25 mM dNTP, 0.6 µL each 5µM forward and reverse primers, 0.1 µL Promega GoTaq. Total reaction volume 15 µL.

Temperature sequence:

94°, 2 min;[94°, 30 sec; T_a, 1 min 30 sec; 72°, 45 sec] x 33; 72°, 8 min; 4°, forever.

Table 17: Microsatellite loci used

Locus	**Primer sequence**	**T_a**	**Size range**	**#Alleles**	**H_e**	**H_o**
GA-1F07	*CCGGTCAAAGAAGTTATCAGA GGGTGGATTGGGTTTCTACCTA	58	157-191	17	0.872	0.705
GA-0E09	*TGCCATCCCTATACACAACCA CCTCCATCACAAAGTTGCC	53	179-265	33	0.922	0.702
GA-0C11	*ATACCCAGCTCCCATGACCA TCCCCAAATTCAGGTAGTGT	53	173-240	32	0.942	0.825
GA-1J11	*AGTTTGGGTCAAATACCTCC AGATAATCCTATGATTGGTCGAG	50	196-240	20	0.886	0.769
GA-0C19	*TTAGCTTTTACGCAGTGTCG CGGCTTCGGTTTCGTC	50	202-242	18	0.878	0.703
GA-1F02	*CCAATCCACCCTTCCAAGTTCC TGGTTGTTTTGCTTTATTCAGCC	50	299-357	29	0.945	0.786

Repeated genotyping enabled me to estimate the probability of mistyping and allelic dropout per allele for each locus (table 18).

Table 18: Genotyping error rate estimation

Locus	# regenotyped	# mistypes	# dropout	Mistyping rate (e1)	Dropout rate (e2)
F07	140	9	7	0.03	0.02
E09	139	5	12	0.02	0.04
C11	135	9	22	0.03	0.08
J11	168	15	14	0.04	0.04
C19	130	47	13	0.18	0.05
F02	163	14	20	0.04	0.06
Total	***875***	***99***	***88***	***0.06***	***0.05***

1.2. Fecundities

For trees within the original mapped area, average fecundities and their standard deviations were calculated using a model developed by Clark et al.(2004, In press). This model incorporates seedtrap and diameter-growth data to estimate the probability of maturity, and annual fecundity given maturity, for each tree. At each step of the Gibbs sampler, a fecundity value is drawn from the posterior distribution defined by these estimates.

Trees outside the original mapped area do not have individual yearly measurements, so fecundity is calculated base on diameter and species-level parameters derived from the same fecundity model, where the probability of maturity θ and fecundity in year *t* (F_t) given maturity are equal to:

$$\theta = \text{inverse logit } (\alpha_0 + \alpha_1(D) + \alpha_2(C))$$

$$\ln(F_t) = \beta_0 + \ln(D_t)\ \beta_1 + \ln((D_t)^2)\ \beta_2 + \ln(C_t)\ \beta_3 + \ln(I_{t-1})\ \beta_4 + \varepsilon.$$

where D is diameter-at-breast-height, C is exposed canopy area, I is diameter increment, and ε is an error term incorporating variability at the individual-, year-, and site-level.

Canopy area and increment estimates are not available for trees outside the original mapped area (border trees). I fit a linear regression relating canopy area to diameter for trees within the original mapped area. For Duke Forest, the relationship for trees 10-80 cm in diameter was: $C = 7.45 + 0.343(D) + \varepsilon$

where $\varepsilon \sim N(0, 3.69)$.

In 500 simulations, a canopy area for each border tree was drawn from a normal distribution with mean $\mu=7.45+0.343(D)$ and standard deviation = 3.69. This canopy area was used to compute a probability of maturity, and a fecundity for those trees estimated to be mature. From these 500 simulations, I calculated a probability of maturity for each border tree and a mean and standard deviation for each tree given maturity. At each Gibbs step, a maturity status is assigned to each border tree. A fecundity is drawn from a truncated normal distribution for each border tree currently estimated to be mature, while immature trees are assigned a fecundity of zero. Fecundity is capped at 3,000 seeds/year for border trees, as trees within the mapped stand never exceeded this value. A lower limit of 0 is also assigned, as fecundities cannot be negative.

Lacking more detailed information, pollen production c_i was assumed to be proportional to seed production f_i for each individual, because large trees that produce many seeds might also be expected to produce large quantities of pollen. A thorough search of the literature found no studies relating seed and pollen production for individual trees, though there does appear to be a positive relationship at the stand level (Saito et al. 1987).

1.3. Out of plot dispersal

In this model, I assume that the focal plot is part of a continuous forest, and that the average density and fecundity of trees outside the plot is equal to the density and fecundity inside. Of course, as most plant populations exhibit clumping, this is not strictly true at small scales. However, in a continuous population, and in the absence of information about the species' distribution outside the plot, I consider it a reasonable assumption that, on average, the rest of the population should resemble the portion inside the plot in density, fecundity, and allele frequencies. We can therefore substitute area at a given pre-defined distance for the specific position and fecundity of out-of-plot trees via numerical integration.

The censused plot is rectangular. Let $de_{i't}$, de_{it}, and de_{kt} be the distance of in-plot adults i' and i, and seedling k, respectively, from the midpoint of plot edge t; let l_t be the length of edge t; let ρ be the density of trees; let η_p be the average pollen production per tree; and let η_s be the average seed production per tree.

The probability of a given in-plot tree i being pollinated by an out-of-plot tree is approximately equal to the expected amount of pollen expected to reach the tree from outside the plot divided by the total expected amount of pollen. The expected amount of pollen reaching tree i from a given in-plot father i' is: $E(pollen) = p(d_{ii'} | u_p)c_{i'}$

Consider a sequence of potential out-of-plot parent-distances $dm_1 \dots dm_n$, where $dm_v - dm_{v-1} = 10$ meters, and n is defined so that $dm_n = 200$ m. These potential parent distances define a series of rectangular rings going outward from the mapped plot. We will approximate the input of pollen and seed from outside the plot by assuming that all the seed or pollen produced within each quarter of each ring v originates from a tree located dm_v meters from the midpoint of each side of the plot (figure 27).

The expected amount of pollen produced inside each quarter-ring (such as the lightly shaded area in figure 27) is calculated and multiplied by the probability of pollen transport over the distance from the midpoint to tree A. Similarly, we can calculate the amount of seed produced within each quarter-ring expected to reach seedling plot B. Summing over all such quantities gives the total expected amount of out-of-plot seed or pollen received. The probability (before taking into account genotypes) of a seed having an out-of-plot father or a seedling an out-of-plot mother is thus the expected amount of out-of-plot pollen or seed received divided by the total expected pollen or seed.

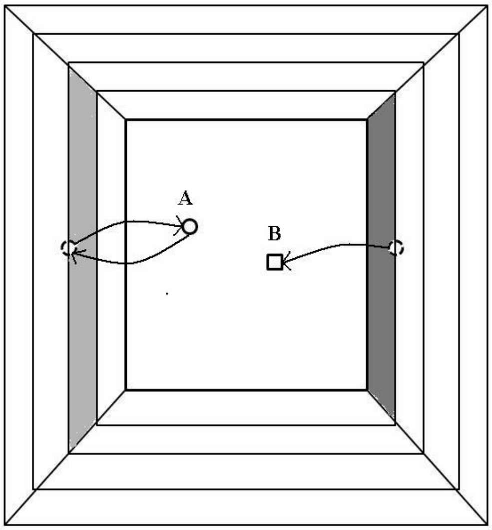

Figure 27: Calculating seed and pollen dispersal from outside to inside trees, pollen dispersal from inside to outside trees.

For pollen, as indicated in figure 27, this procedure can be reversed to find the probability that an in-plot tree pollinates an out-of-plot tree. This becomes important when we consider the probability that a given seedling has an out-of-plot mother but an in-plot father, which is discussed further below.

In order to assess the probability of a seed coming from outside the plot having an in-plot vs. out-of-plot father, we must calculate the amount of pollen expected to arrive at an out-of-plot tree from *another* out-of-plot tree. This will of course depend on the distance of the hypothetical out-of-plot mother tree from the edge of the plot; as this distance increases, the fraction of in-plot pollen drops to zero, but the probability that seed from this mother will disperse into the plot also approaches zero.

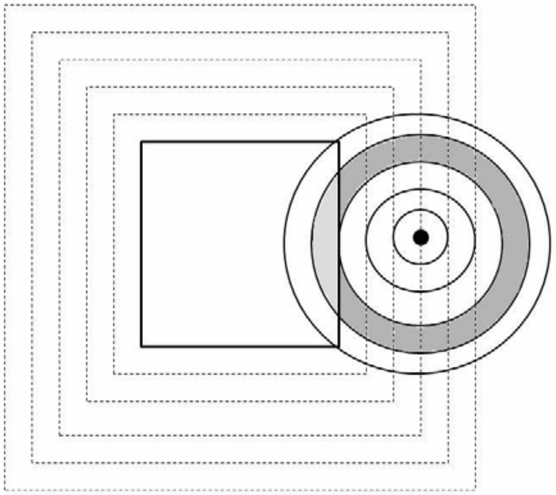

Figure 28: Calculating pollen dispersal between out of plot trees

Consider, a sequence of potential out-of-plot father-to-mother distances $df_1 \ldots df_n$, where $df_q - df_{q-1} = 10$ meters, $df_1 = 10$ m, and $df_n = 200$ m. These distances define a series of circular rings around a mother tree located at any given distance from the plot dm_v (Figure 28). We wish to determine the amount of area in each of these rings that falls outside vs. inside the plot. This determines the fraction of pollen received from each distance class expected to come from outside trees. Mother trees further outside the plot are expected to receive a higher proportion of out-of plot pollen.

Therefore, for a seedling, the probability of having both an out-of-plot mother and an out-of-plot father is equal to the probability of an out-of-plot tree at a given distance being pollinated by another out-of-plot tree, times the probability of seed traveling that distance to the seedling plot, summed over all out-of-plot mother distance classes.

1.4. Priors

Given the width of mature oak canopies, even seeds that are not dispersed by animals will on average fall some short distance away from the parent tree. Dispersal by rodents is thought to be on the order of tens of meters, while jays can move seeds several km (table 19).

At this site, it is unlikely that a majority of acorns are moved by jays, since these birds are observed very rarely, while grey squirrels are frequently observed foraging and caching acorns. **Therefore the prior for u_s was assigned a mean of 253**, corresponding to an average dispersal distance of 25 m, **and a standard deviation of 1000.** The

Gaussian jump distribution was truncated at u_s=10 and u_s=10,000, since u cannot be negative and it is unlikely that the mean dispersal distance for seed is less than 5 m or longer than 157 m; keep in mind that the fat-tail of a 2D-t kernel permits considerably more seed to be dispersed beyond the mean than an exponential or Gaussian kernel, so this limit does not rule out much longer distance dispersal events.

Table 19: Seed dispersal estimates from the literature

Authors	Species	Mean distance	Methods	Notes
Darley-Hill & Johnson 1981	*Q. palustris*, *Q. phellos*, *Q. velutina*	**1.1 km** (range 100 m – 1.9 km)	Direct observation of jays	Caching pref in Blacksburg, VI. 54% of crop cached, often in open suburban sites.
Dow & Ashley 1996	*Q.macrocarpa* (large-seeded oak, in former savanna)	unambiguous matches: **22.8 m** (SD 24.8) all: **26 m** (SD33.2)	Microsatellite analysis of saplings	Clusters of half-sibs around putative mother tree. Only half dispersed beyond crown. 16% went > 90 m, 14% with no in-plot parent.
Li & Zhang 2003	*Q. liaotungensis*	**6.82 m** in 1999, **9.75 m** 2000. Max 23 m.	Seed-tagging	China. Mostly rodents other than squirrels.
Moore et al 2007	*Q. rubra*, *Q. alba, Q. palustris*	For *Q. rubra*, **2.32 – 20.4 m**, depending on year, site, and acorn size	Seed tagging	Investigating effects of crop size on removal by mammals

Wind-blown pollen can travel very long distances. Long-distance pollen movement appears to have maintained genetic connectivity between scattered oak stands in savannah landscapes (Craft and Ashley 2007), and some studies estimate that >50% of pollinations result from pollen traveling >100m (table 20). On the other hand, oak pollen quickly degrades when exposed to UV light (Schueler et al. 2005), and paternity analyses of oaks in some closed-canopy forests have found that the effective pollination neighborhood can be as small as 41.9 m, due to large amounts of pollen contributed by nearby trees. High vegetation density, precipitation, and humidity, can also reduce pollen transport (Ducousso et al. 1993).

Table 20: Pollen dispersal estimates from the literature

Authors	Species	Mean distance	Methods	Notes
Dow & Ashley 1996	*Q. macrocarpa*	**In-plot: 76.9 m** (SD 45)	Microsatellite analysis of saplings	Former savanna. **>50% saplings had parent outside plot (200x250 m)**
Fernandez et al 2006	*Q. velutina*	**24.1 m. Effective pollination radius 41.9 m**	Paternity analysis (allozyme), 8 mother trees	Closed forest. Complete outcrossing.
Nakanishi et al 2004	*Q. salicina*	**In-plot: 66.7 m**	Paternity analysis (microsat), 8 mother trees	Japan, evergreen broadleaf forest. **52% of paternity from outside plot (4 ha)**
Streiff et al 1999	*Q. robur, Q. petraea*	**In-plot: 22.1 m** (SD 32.2) **to 64.6 m** (SD 38.4) **Total: >200m?**	Paternity analysis (microsat),13 mother trees	Europe. Average of **65-69%** in family **had fathers outside plot (5.76 ha)**

The prior for u_p was assigned a mean of 2000, corresponding to an average dispersal distance of 70.2 m, **and a standard deviation of 1500**. The Gaussian jump distribution was truncated at u_p=10 and u_p=15,000, since I considered it unlikely that the mean dispersal distance for pollen is less than 5 m or longer than 192 m.

1.5 Simulation

Simulated data was generated in the following manner. In order to investigate the effect of plot size on dispersal estimation I generated simulated plots of 3 sizes: small (200m x 200 m), medium (300 m x 300 m), and large (390 m x 430 m, the size of the real plot). Each of these simulated plots contained 150 seedling census plots 2 m^2 in size randomly placed in a central 200 m x 200 m area or (in the case of the large plots) the true number and arrangement of seedling plots (figure 29). Trees were generated in a 1,200 m x 1,200 m area, based on the true in-plot density of adults. Some of these individuals fell inside the "mapped" area, and so were considered in-plot trees (which will be assigned an observed genotype), and some outside. "True" dispersal parameters u_{st} and u_{pt} are chosen so as to yield reasonable mean dispersal distances for seed and pollen. Offspring are generated based on the dispersal parameters and adult fecundities (drawn from the true distribution of fecundities).

The expected number of seedlings in a given plot *h* originating from mother tree *i* is: $\lambda_{hi} = \Pr(d_{hi} | u_{st}) f_i S$,where *S* is a seed-to-seedling survival probability chosen so that the density of simulated seedlings is similar to the observed density (in this case, S = 0.1). The number of seedlings in each plot originating from a particular mother is generated by drawing from a Poisson distribution with mean $\lambda_{hi.}$ In this way, seedling locations are generated and maternal parentage determined.

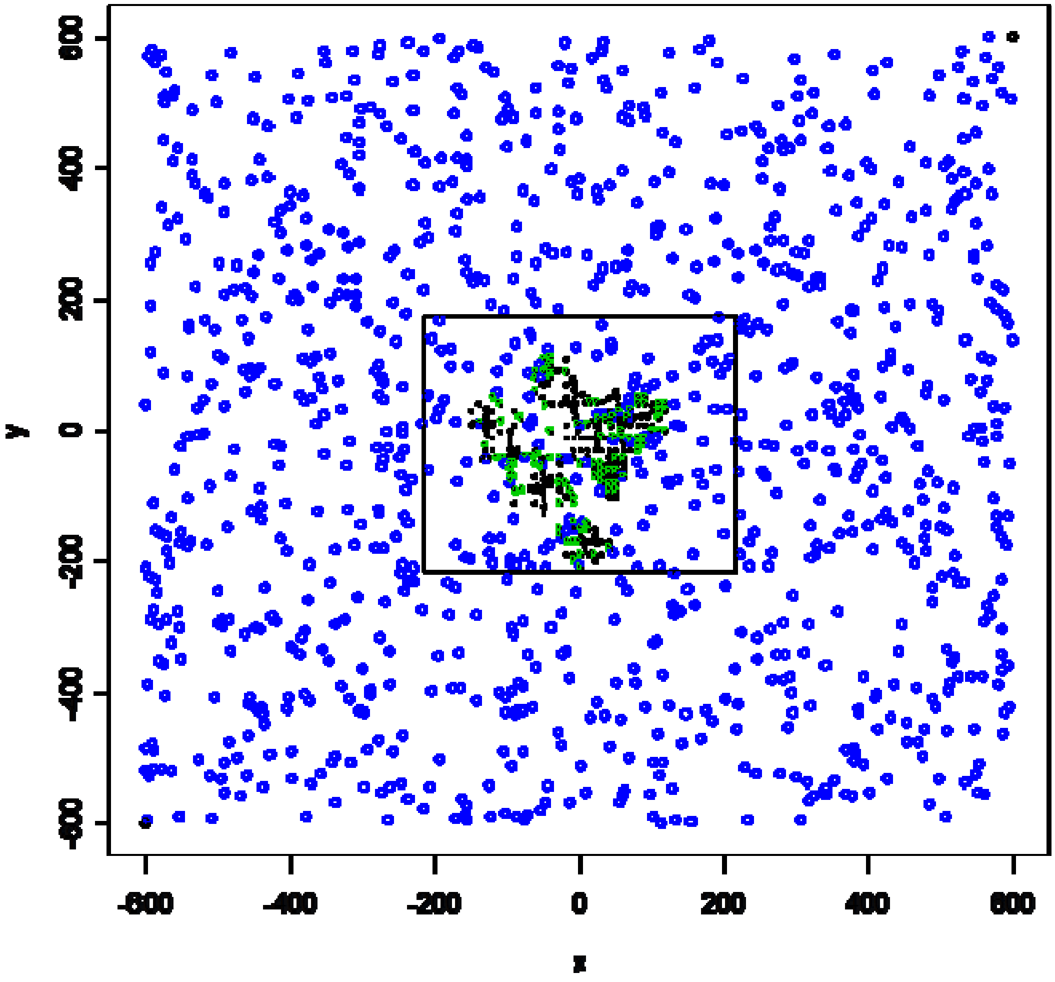

Figure 29: Map of simulated trees and seedlings.

Large square – mapped stand; blue circles- adult trees; small black squares – census plots; green dots - seedlings

Similarly, the expected amount of pollen arriving at mother *i* from father *i'* is:

$$\lambda_{ii'} = \Pr(d_{ii'} | u_{pt})c_{i'}$$

The probability that a given seed originating from mother *i* has father *i′* is therefore equal to the proportion of the total pollen received that comes from *i′*:

$$\theta_{ii'} = \frac{\lambda_{ii'}}{\sum_{i'=1}^{tadult} \lambda_{ii'}}$$

Each seedling with mother *i* is assigned a father using one draw from a multinomial distribution with the vector of probabilities θ_i.

I then ran two sets of simulations: one in which the true parentage of seedlings was known, and one in which it had to be estimated from genotype data containing errors, as in the real analysis. In the second case, "true" genotypes for all adults are drawn from the true population allele frequencies, while "true" offspring genotypes are generated following the principles of Mendelian inheritance. At each locus, one allele is chosen from each parent. Each of the two maternal or paternal alleles has an equal probability of ending up in the offspring. "Observed" genotypes are then generated using true genotyping error rates. A locus is ungenotyped with probability 0.04, in which case the allele values at both positions are replaced by zeros. If a locus is genotyped once, with a probability of 0.4 it is genotyped a second time. Genotyping error is simulated according to the probabilities listed in table 1.

For each plot size, with and without parentage uncertainty, I generated three datasets using three sets of "true" parameters ($u_s = 50$, $u_p = 150$; $u_s = 150$, $u_p = 500$; and $u_s = 300$, $u_p = 1000$), for a total of 54 simulations. Initial values for dispersal parameters were drawn from a uniform distribution ranging from +/- 30 around the true u_s and +/- 50 around the true u_p. The chain was run for 1500 steps (parentage known) or 2000 steps (parentage unknown), with a burn-in of 500 steps. Samples from the posterior distributions of u_s, u_p, and P were evaluated. Estimated pedigrees and dispersal parameters were compared to the "true" values

Table 21: Results of simulation, mean over 3 replicates

Mapped area	# census plots	Parents known?	% within CI, all	SE	% within CI, u_s	ratio estimate to true	True parents	Inverted parents
200 x 200	150	Yes	66.7	116.9	88.9	2.3	NA	NA
300 x 300	150	Yes	55.5	67.5	44.4	1.2	NA	NA
390 x 430	150	Yes	83.3	59.7	100	1.07	NA	NA
200 x 200	150	No	83.3	169.3	88.9	1.74	85.6	9.95
300 x 300	150	No	66.7	111.1	88.9	1.5	86.7	11.3
390 x 430	150	No	77.8	105.9	100	1.2	85.8	11.9
390 x 430	273	No	83.3	57.1	100	1.04	84.5	12.5

Larger plot sizes yielded more accurate dispersal estimates than small plots (table 21). For instance, when parentage was known true parameter values fell within the 95% CI of the estimate in 83.3% of cases in the large-plot simulations but just 66.7% of small-plot simulations. Estimates were more accurate for seed dispersal – true parameter values always fell within 95 % CI in large plots - than for pollen dispersal parameters. Whether a parameter converged to the "correct" value depended on the simulated data. Because dispersal is random and the number of samples limited, there may be slightly more long-distance events or more short-distance events captured than the original dispersal kernel would indicate. For any given simulated dataset, parameters always converged to the same value. As might be expected confidence intervals were narrower in large plots and for larger seedling sample sizes, and larger when parentage was unknown. In small plots, dispersal parameters were more often over-estimated than under-estimated, but there was no positive or negative bias in dispersal estimates in simulated plots of the same dimensions and sampling intensity as the real Duke Forest plot.

Estimates of parentage were highly accurate. In 97% of cases (range 83.8-100%), the true parent pair was the pair most frequently identified by the Gibbs sampler. In 86 % of cases, the true mother and father were identified, while in a further 11 % of cases the true mother was identified as the father and vice versa. As might be expected, the frequency of correct parent pairs was somewhat higher and the frequency of inversions lower for simulations in which true dispersal parameters fell within the 95% CI of estimated dispersal parameters. However, even for less-accurate dispersal estimates, correct parent pairs were identified most of the time.

Appendix 2

2.1 Sensitivity of migration rate to changes in life history

Dispersal distance

SLIP uses a 2D-t kernel, a fat-tailed distribution, to describe seed dispersal probabilities, where the probability of traveling a distance d is equal to:

$$\frac{1}{\pi u\left(1+\frac{d^{2}}{u}\right)^{2}}$$

The mean dispersal distance for a 2D-t kernel is $\frac{\pi\sqrt{u}}{2}$, where u is the parameter describing the shape of the distribution. For *Q. rubra*, the u estimated from seed trap data is 34.9, corresponding to a mean dispersal distance of 9.27 meters. A 50% increase in mean dispersal distance, to 13.9 m, corresponds to u = 78.36, while a 50% reduction corresponds to u = 4.64 (parameter sets Disp+, Disp -).

Fecundity

In SLIP, the natural log of fecundity is equal to

$$\alpha_0+\alpha_1\ln(DBH)+\alpha_2\ln(DBH^2)+\alpha_3\ln(ECA)+\alpha_4\ln(inc)+C$$

where the α's are fitted parameters, and *DBH* is diameter-at-breast-height (the second, squared term accounting for senescence of very large trees), *ECA* exposed canopy area (indicating light resources), and *inc* the growth increment (width of annual growth ring). If baseline fecundity is f_0, and we wish a new fecundity f_n to be equal to bf_0, then we can see that

$$\ln(f_n) = \ln(bf_0) = \ln(b) + \ln(f_0)$$

Therefore, if we desire f_n to be 50% larger than f_0 , then we must set $\alpha_n = \alpha_0 + \ln(1.5)$ (parameter set Fec+). To decrease fecundity by 50%, I set $\alpha_n = \alpha_0 + \ln(0.5)$ (parameter set Fec-).

Size at maturity

The probability θ of a tree being sexually mature is $\frac{e^F}{1+e^F}$, where $F = \beta_0 + \beta_1(DBH) + \beta_2(ECA)$, and the β's are fitted parameters. Thus, if D is the diameter at which the probability of being mature is 50%, then

$D = \frac{\beta_2 \overline{ECA} - \beta_0}{\beta_1}$ where $\overline{ECA}$ is the average exposed canopy area. If we wish to increase or decrease D by a certain percentage, this could be done by changing β_1, β_2 and β_0, or all three together. In one set of simulations (Mat1+, Mat1-), β_1 was altered, while in another set (Mat2+,Mat2-) we changed β_2 and β_0. These two approaches lead to a difference in the shape of the relationship between diameter and probability of being mature. The three lines in figure 30 depict the probability of being mature for trees from 1 to 100 cm dbh given a canopy exposure of 20 m^2. The approach to 100% probability of being mature is different for Mat1+ (dashed line) and Mat2+ (dotted). Note, however, that the 50% probability of being mature occurs at the same dbh, which is 50% larger than the dbh corresponding to 50% probability for the baseline parameter set (solid line).

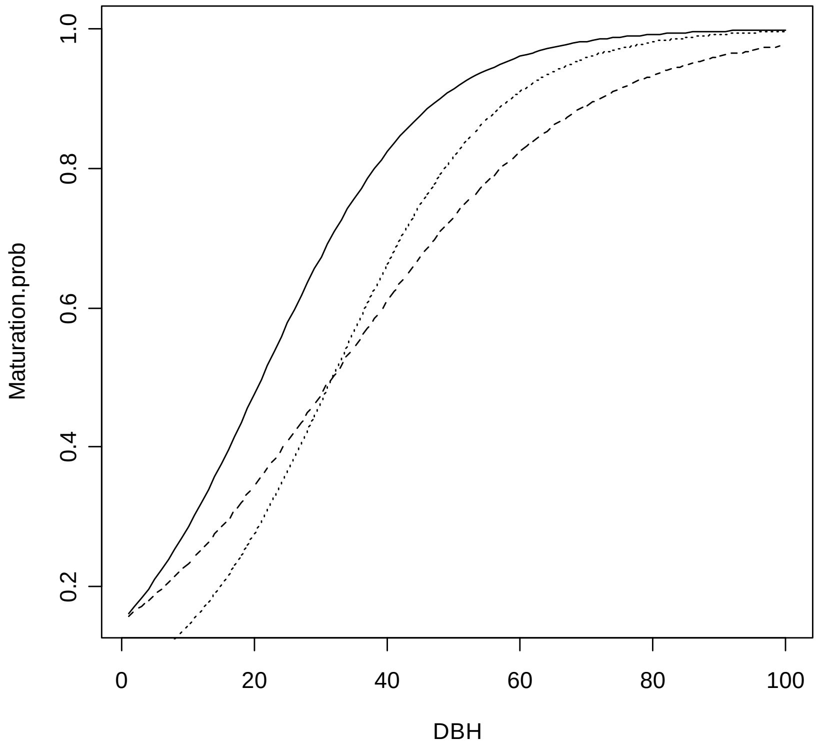

Figure 30: Maturation probability at ECA = 20 m^2.
Solid line – baseline parameters. Dotted line – Mat 2+. Dashed line – Mat 1+.

2.2 Initial conditions

The initial size and locations of the migrant trees had a strong effect on population trajectory. Table 22 details these initial conditions.

Table 22: Initial conditions

	Initial pop 1		Initial pop 2		Initial pop 3	
Tree	**Location (x,y)**	**DBH**	**Location (x,y)**	**DBH**	**Location (x,y)**	**DBH**
1	(19.05, 4.18)	43.27	(5.14, 17.52)	42.03	(16.53, 2.4)	39.8
2	(18.9, 2.91)	30.63	(5.24, 19.08)	40.55	(9.41, 14.86)	32.86
3	(1.12, 3.25)	35.92	(0.59, 2.78)	26.15	(10.71, 7.2)	22.3
4	(16.3,10.77)	33.48	(11.5, 0.58)	44.53	(7.72, 18.81)	27.12
5	(3.1, 19.01)	41.01	(4.47, 5.23)	33.28	(11.87, 1.34)	29.91
6	(13.5, 6.18)	32.08	(17.22, 7.37)	20.62	(3.5, 10.85)	35
7	(6.34, 16.34)	37.82	(15.35, 16.21)	40.98	(19.41, 16.19)	38.55
8	(6.75, 0.67)	37.33	(2.59, 0.73)	8.14	(11.34, 19.86)	26.96
9	(19.07, 13.88)	31.08	(7.03, 4.37)	3.99	(2.53, 5.96)	27.88
10	(10.16, 19.82)	39.54	(16.22, 15.01)	25.04	(15.7, 18.86)	20.77

The “competition” simulations were based on a simulated multi-species community derived from the community present in the Duke Forest stand. Demographic parameters have been estimated for all 29 species. Trees were allowed to fill in a 512 m X 512 m landscape to form a continuous forest. Over time, species are lost if they do not reproduce before the parent trees die (figure 21). At the end of 220 years, these simulated forests contained between 13 and 20 species, in addition to the newly-introduced migrants.

After an additional 150 more years, with or without disturbance, 4-8 resident species remained. These always included red maple (a species which has greatly increased in abundance in eastern deciduous forests over the past century), tulip poplar and loblolly pine (both good colonizers of gaps), and willow oak.

Table 23: Competing species present for all 3 sets of initial conditions at year 1, and after 150 years in disturbed (d) and undisturbed (ud) conditions.

All species listed present in Duke Forest stand. Background forest 1 = A, forest 2 = B, forest 3 = C.

	IC.1			**IC.2**			**IC.3**		
Species	**1**	**150ud**	**150d**	**1**	**150ud**	**150d**	**1**	**150ud**	**150d**
Acer barbatum									
A. rubrum	ABC	ABC	ABC	ABC	ABC	ABC	ABC	ABC	ABC
A. spp				ABC					
Carpinus car.									
Carya glabra				ABC					
C. tomentosa				ABC					
C. spp.				ABC					
Cercis canad.	ABC			ABC					
Cornus florida									
Fraxinus am.	ABC			ABC			ABC		A
Liquidambar styraciflua				ABC					
Liriodendron tulipifera	ABC	ABC	ABC	ABC	ABC	ABC	ABC	ABC	ABC
Nyssa sylv.									
Oxydendron arboreum	ABC			ABC			ABC		
Pinus echinata	ABC	AC	C	ABC	AB		ABC	A	
P. taeda	ABC	ABC	ABC	ABC	ABC	ABC	ABC	ABC	ABC
P. virginiana	ABC	BC	AB	ABC	C	AC	ABC		
Quercus alba	ABC	ABC	AB	ABC	B	B	ABC	ABC	ABC
Q. falcata									
Q.marylandica	ABC								
Q. phellos	ABC	ABC	ABC	ABC	ABC	ABC	ABC	ABC	ABC
Q. rubra	ABC		C	ABC			ABC	ABC	A
Q. stellata				ABC					
Q. spp.									
Q. velutina									
Robinia pseud.									
Ulmus alata	ABC			ABC	A	BC	ABC	A	B
U. americana	ABC			ABC		AC	ABC	C	
U. spp.	ABC			ABC			ABC	B	

Without disturbance, white oak survived for 150 years in 7 out of 9 simulations, shortleaf pine in 5, pitch pine in 3, winged and American elm in 2. With disturbance, white oak survived for 150 years in 6 out of 9 simulations, pitch pine in 4, winged elm in 3, shortleaf pine, American elm, and American ash in 1. *Quercus rubra* with baseline parameters (ie short dispersal), although initially abundant, survived in just 2 out of 9 simulations with disturbance, and 3 out of 9 without.

Literature Cited

Abrams, M. D. 1992. Fire and the development of oak forests. Bioscience **42**:346-353.

Agarwal, A. A., P. M. Kotanen, C. E. Mitchell, A. G. Power, W. Godsoe, and J. Klironomos. 2005. Enemy release? An experiment with congeneric plant pairs and diverse above- and belowground enemies. Ecology **86**:2979-2989.

Aldrich, P. R., P. G.R., J. Romero-Severson, and C. H. Michler. 2005a. Confirmation of Oak Recruitment Failure in Indiana Old-Growth Forest: 75 Years of Data. Forest Science **51**:406-416.

Aldrich, P. R., J. C. Glaubitz, G. R. Parker, O. E. Rhodes, and C. H. Michler. 2005b. Genetic structure inside a declining red oak community in old-growth forest. Journal of Heredity **96**:627-634.

Aldrich, P. R., J. L. Hamrick, P. Chavarriaga, and G. Kochert. 1998. Microsatellite analysis of demographic genetic structure in fragmented populations of the tropical tree *Symphonia globulifera*. Molecular Ecology **7**:933-944.

Aldrich, P. R., M. Jagtap, C. H. Michler, and R. 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., C. H. Michler, W. Sun, and R. Romero-Severson. 2002. Microsatellite markers for northern red oak (Fagaceae: *Quercus rubra)*. Molecular Ecology Notes **2**:472-474.

Aldrich, P. R., G. R. Parker, C. H. 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.

Allendorf, F. W., R. F. Leary, P. Spruell, and J. K. Wenburg. 2001. The problem with hybrids: setting conservation guidelines. TREE **16**:613-622.

Anderson. 1948. Hybridization of the habitat. Evolution **2**:1-19.

Anekonda, T. S., M. C. Lomas, W. T. Adams, K. L. Kavanagh, and S. N. Aitken. 2002. Genetic variation in drought hardiness of coastal Douglas-fir seedlings from British Columbia. Canadian Journal of Forest Research **32**:1701-1716.

Asuka, Y., N. Tomaru, Y. Munehara, N. Tani, Y. Tsumura, and S. Yamamoto. 2005. Half-sib family structure of *Fagus crenata* saplings in an old-growth beech-dwarf bamboo forest. Molecular Ecology **14**:2565-2575.

Asuka, Y., N. Tomaru, N. Nisimura, Y. Tsumura, and S. Yamamoto. 2004. Heterogeneous genetic structure in a *Fagus crenata* population in an old-growth beech forest revealed by microsatellite markers. Molecular Ecology **13**:1241-1250.

Bacilieri, R., A. Ducousso, R. J. Petit, and A. Kremer. 1996. Mating system and asymmetric hybridization in a mixed stand of European oaks. Evolution **50**:900-908.

Bacles, C. F. E., A. J. Lowe, and R. A. Ennos. 2006. Effective seed dispersal across a fragmented landscape. Science **311**:628.

Beckage, B. and J. S. Clark. 2003. Seedling survival and growth of three forest tree species: the role of spatial heterogeneity. Ecology **84**:1849-1861.

Bonin, A., E. Bellemain, P. Bronken Eidesen, F. Pompanon, C. Brochmann, and P. Taberlet. 2004. How to track and assess genotyping errors in population genetics studies. Molecular Ecology **13**:3261-3273.

Buckland, S. T., D. R. Anderson, K. P. Burnham, and J. L. Laake. 1993. Distance sampling: estimating abundance of biological populations. Chapman and Hall Inc., New York.

Buckley, D. S., T. L. Sharik, and J. G. Isebrands. 1998. Regeneration of northern red oak: positive and negative effects of competitor removal. Ecology **79**:65-78.

Burczyk, J., S. P. DiFazio, and W. T. Adams. 2004. Gene flow in forest trees: How far do genes really travel? Forest Genetics **11**:1-14.

Burgarella, C., Z. Lorenzo, R. Jabbour-Zahab, R. Lumaret, E. Guichoux, R. J. Petit, A. Soto, and L. Gil. 2009. Detection of hybrids in nature: application to oaks (*Quercus suber* and *Q. ilex*). Heredity **102**:442-452.

Burger, W. C. 1975. The species concept in *Quercus*. Taxon **24**:45-50.

Cadenasso, M. L. and S. T. A. Pickett. 2001. Effect of edge structures on the flux of species into forest interiors. Conservation Biology **15**:91-97.

Clark, C. J., J. R. Poulsen, B. M. Bolker, E. F. Connor, and V. T. Parker. 2005. Comparative seed shadows of bird-, monkey-, and wind-dispersed trees. Ecology **86**:2684-2694.

Clark, J. S. 1998. Why trees migrate so fast: confronting theory with dispersal biology and the paleorecord. the American Naturalist **152**:204-224.

Clark, J. S. Submitted. Climate change vulnerability of competing species. PNAS.

Clark, J. S., D. Bell, C. Chu, B. Courbaud, M. Dietze, M. H. Hersh, J. Hille Ris Lambers, I. Ibanez, S. LaDeau, S. M. McMahon, J. Metcalf, J. Mohan, E. V. Moran, L. Pangle, S. Pearson, C. Salk, Z. Shen, D. Valle, and P. Wyckoff. In press. High dimensional coexistence based on individual variation: a synthesis of evidence. Ecological Monographs.

Clark, J. S., M. Dietze, S. Chakraborty, P. K. Agarwal, I. Ibanez, S. LaDeau, and M. Wolosin. 2007. Resolving the biodiversity paradox. Ecology Letters **10**:647-662.

Clark, J. S., S. LaDeau, and I. Ibanez. 2004. Fecundity of trees and the colonization competition hypothesis. Ecological Monographs **74**:415-442.

Clark, J. S., M. Lewis, and L. Horvath. 2001. Invasion by extremes: population spread with variation in dispersal and reproduction. the American Naturalist **157**:537-554.

Clark, J. S., J. Mohan, M. Dietze, and I. Ibanez. 2003. Coexistence: how to identify trophic trade-offs. Ecology **84**:17-31.

Clark, J. S., M. Silman, R. Kern, E. Macklin, and J. H. R. Lambers. 1999. Seed dispersal near and far: patterns across temperate and tropical forests. Ecology **80**:1475-1494.

Clayton, N. S., J. M. Dally, and N. J. Emery. 2007. Social cognition by food-caching corvids. The western scrub-jay as a natural psychologist. Phil. Trans. Royal. Soc. B. **362**:507-522.

Clinton, B. D., L. R. Boring, and W. T. Swank. 1993. Canopy gap characteristics and drought influences in oak forests of the Coweeta basin. Ecology **74**:1551-1558.

Clutton-Brock, T. H. 1988. Reproductive success. Studies of individual variation in contrasting breeding systems. University of Chicago Press, Chicago.

Collingham, Y. C. and B. Huntley. 2000. Impacts of habitat fragmentation and patch size upon migration rates. Ecological Applications **10**:131-144.

Connell, J. H. 1978. Diversity in tropical rain forests and coral reefs. Science **199**:1302-1310.

Cottam, W. P., J. M. Tucker, and F. S. Santamour. 1982. Oak hybridization at the University of Utah. State Arboretum of Utah, Salt Lake City.

Coyne, J. A. and H. A. Orr. 2004. Speciation. Sinauer Associates, Inc., Sunderland, MA.

Craft, K. J. and M. V. Ashley. 2007. Landscape genetic structure of bur oak (*Quercus macrocarpa*) savannas in Illinois. Forest Ecology and Management **239**:13-20.

Craft, K. J., M. V. Ashley, and W. D. Koenig. 2002. Limited hybridization between *Quercus lobata* and *Quercus douglasii* (Fagaceae) in a mixed stand in central coastal California. American Journal of Botany **89**:1792-1798.

Crow, T. R. 1988. Reproductive mode and mechanisms for self-replacement of Northern Red Oak (*Quercus rubra*) - a review. Forest Science **34**:19-40.

Curtu, A. L., O. Gailing, and R. Finkeldey. 2007. Evidence for hybridization and introgression within a species-rich oak (*Quercus spp*.) community. BMC Evolutionary Biology **7**.

Darley-Hill, S. and W. C. Johnson. 1981. Acorn dispersal by the blue jay (*Cyanocitta cristata*). Oecologia **50**:231-232.

Darwin, C. 1859. The origin of species by means of natural selection.

Davis, M. B. 1969. Climatic changes in southern connecticut recorded by pollen deposition at Rogers lake. Ecology **50**:409-422.

Davis, M. B. and R. G. Shaw. 2001. Range shifts and adaptive responses to quaternary climate change. Science **292**.

DeGange, A. R., J. W. Fitzpatrick, J. N. Layne, and G. E. Woolfenden. 1989. Acorn harvesting by Florida scrub jays. Ecology **70**:348-356.

Dewoody, J., J. Nason, and V. D. Hipkins. 2006. Mitigating scoring errors in microsatellite data from wild populations. Molecular Ecology Notes **6**:951-957.

Dietze, M. 2006. Regeneration dynamics in large forest gaps. Duke University, Durham, NC.

Dodd, M. E. and J. Silvertown. 2000. Size-specific fecundity and the influence of lifetime size variation upon effective population size in *Abies balsamea*. Heredity **85**:604-609.

Dodd, R. S. and Z. Afzal-Rafii. 2004. Selection and dispersal in a multispecies oak hybrid zone. Evolution **58**:261-269.

Dow, B. D. and M. V. Ashley. 1996. Microsatellite analysis of seed dispersal and parentage of saplings in bur oak, *Quercus macrocarpa*. Molecular Ecology **5**:615-627.

Doyle, J. J. and J. L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin **19**:11-15.

Ducousso, A., H. Michaud, and R. Lumaret. 1993. Reproduction and gene flow in the genus *Quercus* L. Ann Sci For **50**:91s-106s.

Dullinger, S., T. Dirnbock, and G. Grabherr. 2004. Modelling climate change-driven treeline shifts: relative effects of temperature increase, dispersal, and invasibility. Journal of Ecology **92**:241-252.

Dyer, J. M. 1995. Assessment of climatic warming using a model of forest species migration. Ecological Modeling **79**:199-219.

Eguiarte, L. E., A. Burquez, J. Rodriguez, M. Martinez-Ramos, J. Sarukhan, and D. Pinero. 1993. Direct and indirect estimates of neighborhood and effective population size in a tropical palm, *Astrocaryum mexicanum*. Evolution **47**:75-87.

Elena-Rossello, J. A., I. R. Lumaretz, E. Cabrera, and H. Michaud. 1992. Evidence for hybridization between sympatric holm-oak and cork-oak in Spain based on diagnostic enzyme markers. Vegetatio **99-100**:115-118.

Elliott, K. J., R. L. Hendrick, A. E. Major, J. M. Vose, and W. T. Swank. 1999. Vegetation dynamics after a prescribed fire in the southern Appalachians. Forest Ecology and Management **114**:199-213.

Emigh, T. H. and E. Pollak. 1979. Fixation probabilities and effective population numbers in diploid populations with overlapping generations. Theoretical Population Biology **15**:86-107.

Fernandez, J. F., J. Idol, and V. L. Sork. 2006. Mating patterns of black oak *Quercus velutina* (Fagaceae) in a Missouri oak-hickory forest. Journal of Heredity **97**:451-455.

Fernandez, J. F. and V. L. Sork. 2007. Genetic variation in fragmented forest stands of the andean oak *Quercus humboldtii* Bonpl. (Fagaceae). Biotropica **39**:72-78.

Fuchs, E. J., J. A. Lobo, and M. Quesada. 2003. Effect of forest fragmentation and flowering phenology on the reproductive success and mating patterns of the tropical dry forest tree *Pachira quinata*. Conservation Biology **17**:149-157.

Garant, D., J. J. Dodson, and L. Bernatchez. 2001. A genetic evaluation of mating system and determinants of individual reproductive success in Atlantic Salmon (*salmo salar* L.). Journal of Heredity **92**:137-145.

Garcia, D. and G. Houle. 2005. Fine-scale spatial patterns of recruitment in red oak (*Quercus rubra*): What matters most, abiotic or biotic factors? Ecoscience **12**:223-235.

Germein, J. B. and R. Kesseli. 2006. Genetic diversity and population structure of *Quercus rubra* (Fagaceae) in old-growth and secondary forests in southern New England. Rhodora **108**:1-18.

Gillespie, J. H. 2004. Population genetics; a concise guide. 2nd edition. Johns Hopkins University Press.

Godoy, J. A. and P. Jordano. 2001. Seed dispersal by animals: exact identification of source trees with endocarp DNA microsatellites. Molecular Ecology **10**:2275-2283.

Gomez, J. M. 2003. Spatial patterns in long-distance dispersal of *Quercus ilex* acorns by jays in a heterogeneous landscape. Ecography **26**:573-584.

Gonzalez-Martinez, S. C., J. Burczyk, R. Nathan, N. Nanos, L. Gil, and R. Alia. 2006. Effective gene dispersal and female reproductive success in Mediterranean maritime pine (*Pinus pinaster* Aiton). Molecular Ecology **15**:4577-4588.

Goto, S., K. Shimatani, H. Yoshimaru, and Y. Takahashi. 2006. Fat-tailed gene flow in the dioecious canopy tree species *Fraxinus mandshurica* var. *japonica* revealed by microsatellites. Molecular Ecology **15**:2985-2996.

Govindarajan, S., M. Dietze, P. K. Agarwal, and J. S. Clark. 2004. A Scalable Simulator for Forest Dynamics. Pages 106-115 *in* Proceedings of the twentieth annual symposium on compuational geometry, Brooklyn NY.

Govindarajan, S., M. Dietze, P. K. Agarwal, and J. S. Clark. 2007. A scalable algorithm for dispersing population. J. Intel. Inform. Syst. **29**:39-61.

Greenberg, C. H. and W. H. McNab. 1998. Forest disturbance in hurricane-related downbursts in the Appalachian mountains of North Carolina. Forest Ecology and Management **104**:179-191.

Gribko, L. S., T. M. Schuler, and W. M. Ford. 2002. Biotic and abiotic mechanisms in the establishment of northern red oak seedlings: a review. Department of Agriculture, Forest Service, Northeastern Research Station, Newtown Square, PA.

Grivet, D., P. E. Smouse, and V. L. Sork. 2005. A novel approach to an old problem: tracking dispersed seeds. Molecular Ecology **14**:3585-3595.

Guttman, S. I. and L. A. Weigt. 1989. Electophoretic evidence of relationships among *Quercus* (oaks) of eastern North America. Canadian Journal of Botany **67**:339-351.

Hadfield, J. D., D. S. Richardson, and T. Burke. 2006. Towards unbiased parentage assignment: combining genetic, behavioral and spatial data in a Bayesian framework. Molecular Ecology **15**:3715-3730.

Hamrick, J. L. 2004. Response of forest trees to global environmental changes. Forest Ecology and Management **197**:323-335.

Hannah, L., G. Midgley, G. Hughes, and B. Bomhard. 2005. The view from the cape: extinction risk, protected areas, and climate change. Bioscience **55**:231-242.

Hardesty, B. D., S. P. Hubbell, and E. Bermingham. 2006. Genetic evidence of frequent long-distance recruitment in a vertebrate-dispersed tree. Ecology Letters **9**:516-525.

Hein, E. W. 1997. Demonstration of line transect methodologies to estimate urban gray squirrel density. Environmental Management **21**:943-947.

Herrera, C. M. and R. Jovani. 2010. Lognormal distribution of individual lifetime fecundity: insights from a 23-year study. Ecology **9**:422-430.

Heywood, J. S. 1986. The effect of plant size variation on genetic drift in populations of annuals. the American Naturalist **127**:851-861.

Higgins, S. I., J. S. Clark, R. Nathan, T. Hovestadt, F. Schurr, J. M. Fragoso, M. R. Aguiar, E. Ribbens, and S. Lavorel. 2003. Forecasting plant migration rates: managing uncertainty for risk assessment. Journal of Ecology **91**:341-347.

Higgins, S. I. and D. M. Richardson. 1999. Predicting plant migration rates in a changing world: the role of long-distance dispersal. the American Naturalist **153**:464-474.

Hill, W. G. 1972. Effective size of populations with overlapping generations. Theoretical Population Biology **3**:278-289.

Holladay, C. A., C. Kwit, and B. Collins. 2005. Woody regeneration in and around aging southern bottomland hardwood forest gaps: effects of herbivory and gap size. Forest Ecology and Management **223**:218-225.

Holt, R. D. 1990. The microevolutionary consequences of climate change. TREE **5**:311-315.

Hubbell, S. P. 1979. Tree dispersion, abundance, and diversity in a tropical dry forest. Science **203**:1299-1309.

Huntley, B. 1991. How plants respond to climate change: migration rates, individualism, and the consequences for plant communities. Annals of Botany **67**:15-22.

Ibanez, I., J. S. Clark, and M. Dietze. 2008. Evaluating the sources of potential migrant species: implications under climate change. Ecological Applications **18**:1664-1678.

Ibanez, I., J. S. Clark, S. LaDeau, and J. H. R. Lambers. 2007. Exploiting temporal variability to understand tree recruitment response to climate change. Ecological Monographs **77**:163-177.

IPCC. 2007. Climate change 2007: The physical science basis. Summary for policy makers.

Iverson, L. R. and A. M. Prasad. 2001. Potential changes in tree species richness and forest community types following climate change. Ecosystems **4**:186-199.

Iverson, L. R., M. W. Schwartz, and A. M. Prasad. 2004. Potential colonization of newly available tree-species habitat under climate change: an analysis for five eastern US species. Landscape Ecology **19**:787-799.

Janzen, D. 1975. A note on optimal mate selection by plants. the American Naturalist:365-371.

Janzen, D. H. 1970. Herbivores and the number of tree species in tropical forests. the American Naturalist **104**:501-526.

Jensen, R. J. 1977. A preliminary numerical analysis of the red oak complex in Michigan and Wisconsin. Taxon **26**:399-407.

Jogesh, T., D. Carpenter, and N. Cappuccino. 2008. Herbivory on invasive exotic plants and their non-invasive relatives. Biological Invasions **10**:797-804.

Johnson, A. S., P. E. Hale, W. M. Ford, J. M. Wentworh, J. R. French, O. F. Anderson, and G. B. Pullen. 1995. White-tailed deer foraging in relation to successional stage, overstory type and management of southern Appalachian forests. American Midland Naturalist **133**:18-35.

Johnson, M. T. J. and J. R. Stinchcombe. 2007. An emerging synthesis between community ecology and evolutionary biology. TREE **22**.

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. I. 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.

Jones, F. A., J. L. Hamrick, C. J. Peterson, and E. R. Squiers. 2006. Inferring colonization history from analyses of spatial genetic structure within populations of *Pinus strobus* and *Quercus rubra*. Molecular Ecology **15**:851-861.

Kashani, N. and R. S. Dodd. 2002. Genetic differentiation of two California red oak species, *Quercus parvula* var. *shreveii* and *Q. wislizeni*, based on AFLP genetic markers.*in* U. F. Service, editor.

Kawecki, T. J. 2008. Adaptation to marginal habitats. Annual Review of Ecology, Evolution, and Systematics **39**:321-342.

Keever, C. 1953. Present composition of some stands of the former oak-chestnut forest in the southern Blue Ridge Mountains. Ecology **34**:44-54.

Kimbrell, T. and R. D. Holt. 2007. Canalization breakdown and evolution in a source-sink system. the American Naturalist **169**:370-382.

Kirkpatrick, M. and N. H. Barton. 1997. Evolution of a species' range. the American Naturalist **150**:1-23.

Klinkhamer, G. L. and T. J. de Jong. 1993. Attractiveness to pollinators: a plant's dilemma. Oikos **66**:180-184.

Knapp, E. E., M. A. Goedde, and K. J. Rice. 2001. Pollen-limited reproduction in blue oak: implications for wind pollination in fragmented populations. Oecologia **128**:48-55.

LaDeau, S. and J. S. Clark. 2001. Rising CO_2 levels and the fecundity of forest trees. Science **292**:95-98.

Larsen, J. A. 1953. A study of an invasion by red maple of an oak woods in southern Wisconsin. American Midland Naturalist **49**:908-914.

Larson, K. C. and T. G. Whitham. 1997. Competition between gall aphids and natural plant sinks: plant architecture affects resistance to galling. Oecologia **109**:575-582.

Lenormand, T. 2002. Gene flow and the limits to natural selection. TREE **17**:183-189.

Levin, D. A. 1981. Dispersal versus gene flow in plants. Annals of the Missouri Botanical Garden **68**:233-253.

Li, H. J. and Z.-B. Zhang. 2003. Effect of rodents on acorn dispersal and survival of the Liaodong oak (*Quercus liaotungensis* Koidz.). Forest Ecology and Management **176**:387-396.

Little, E. L. 1980. National Audubon Society Field Guide to Trees, eastern region. Alfred A. Knopf, Inc., New York.

Loarie, S. R., P. B. Duffy, H. Hamilton, G. P. Asner, C. B. Field, and D. D. Ackerly. 2009. The velocity of climate change. Nature **462**:1052-1055.

Lopez, S., F. Rousset, F. H. Shaw, R. G. Shaw, and O. Ronce. 2007. Migration load in plants: role of pollen and seed dispersal in heterogeneous landscapes. Journal of Evolutionary Biology **21**:294-309.

Lorenzo, Z., C. Burgarella, U. Lopez de Heredia, R. Lumaret, R. J. Petit, A. Soto, and L. Gil. 2009. Relevance of genetics for conservation policies: the case of Minorcan cork oaks. Annals of Botany **104**:1069-1076.

Lorimer, C. G. and A. S. White. 2003. Scale and frequency of natural disturbances in the northeastern US: implications for early-successional forest habitats and regional age distributions. Forest Ecology and Management **185**:41-64.

Loveless, M. D. and J. L. Hamrick. 1984. Ecological determinants of genetic structure in plant populations. Annual Review of Ecology and Systematics **15**:65-95.

Malanson, G. P. and D. M. Cairns. 1997. Effects of dispersal, population delays, and forest fragmentation on tree migration rates. Plant Ecology **131**:67-79.

Malcolm. 2002. Estimated migration rates under scenarios of global climate change. Journal of Biogeography **29**:835-849.

Mank, J. E., J. E. Carlson, and M. C. Brittingham. 2004. A century of hybridization: Decreasing genetic distance between American black ducks and mallards. Conservation Genetics **5**:395-403.

Marshall, T. C., J. Slate, L. E. B. Kruuk, and J. M. Pemberton. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. Molecular Ecology **7**:639-655.

Martin, P. H. and P. L. Marks. 2006. Intact forests provide only weak resistance to a shade-tolerant invasive Norway maple (*Acer platanoides* L.). Journal of Ecology **94**:1070-1079.

Masek, J. G. 2001. Stability of boreal forest stands during recent climate change: evidence from Landsat satellite imagery. Journal of Biogeography **28**:967-976.

McDonald, R. I., R. K. Peet, and D. L. Urban. 2002. Environmental correlates of oak decline and red maple increase in the North Carolina Piedmont. Castanea **67**:84-95.

McKenney, D. W., J. H. Pedlar, K. Lawrence, K. Campbell, and M. F. Hutchinson. 2007. Potential impacts of climate change on the distribution of North American trees. Bioscience **57**:939-948.

McLachlan, J. S., J. S. Clark, and P. S. Manos. 2005. Molecular indicators of tree migration capacity under rapid climate change. Ecology **86**:2088-2098.

McMahon, S. M., M. Dietze, M. H. Hersh, E. V. Moran, and J. S. Clark. 2009. A predictive framework to understand forest responses to global change. Pages 221-236 *in* R. S. Ostfeld and W. H. Schlesinger, editors. The year in ecology and conservation biology, 2009. New York Academy of Sciences, New York.

McMahon, S. M., H. Yu, M. Dietze, P. K. Agarwal, and J. S. Clark. Submitted. Immigration cannot maintain forest biodiversity. Nature.

McShea, W. J., W. M. Healy, P. Devers, T. Fearer, F. H. Koch, D. Stauffer, and J. Waldon. 2006. Forestry matters: Decline of oaks will impact wildlife in hardwood forests. Journal of Wildlife Management **71**:1717-1728.

Meagher, T. R. and E. Thompson. 1987. Analysis of parentage for naturally established seedlings of *Chamaelirium luteum* (Liliaceae). Ecology **68**:803-812.

Mearns, L. O., F. Giorgi, L. McDaniel, and C. Shields. 2003. Climate scenarios for the southeastern U.S. based on GCM and regional model simulations. Climatic Change **60**:7-35.

Melbourne, B. A. and A. Hastings. 2009. Highly variable spread rates in replicated biological invasions: fundamental limits to predictability. Science **325**:1536-1539.

Mimura, M. and S. N. Aitken. 2010. Local adaptation at range peripheries in Sitka spruce. Journal of Evolutionary Biology.

Mimura, M. and S. N. Aitken. In press. Local adaptation at range peripheries in Sitka spruce.

Moore, J. E., A. B. McEuen, R. K. Swihart, T. A. Contreras, and M. A. Steele. 2007. Determinants of seed removal distance by scatter-hoarding rodents in deciduous forests. Ecology **88**:2529-2540.

Muir, G. and C. Schlotterer. 2005. Evidence for shared ancestral polymorphism rather than recurrent gene flow at microsatellite loci differentiating two hybridizing oaks (*Quercus* spp.). Molecular Ecology **14**:549-561.

Muller, C. H. 1952. Ecological control of hybridization in *Quercus*: a factor in the mechanism of evolution. Evolution **6**:147-161.

Murray, M. G. and W. F. Thompson. 1980. Rapid isolation of high molecular weight plant DNA. Nucleic Acids Research **8**:4321-4325.

Nakanishi, A., N. Tomaru, H. Yoshimaru, T. Kawahara, T. Manabe, and S. Yamamoto. 2004. Patterns of pollen flow and genetic differentiation among pollen pools in *Quercus salicina* in a warm temperate old-growth evergreen broad-leaved forest Silvae Genetica **53**:258-264.

Neilson, R. P., L. F. Pitelka, A. M. Solomon, R. Nathan, G. Midgley, J. M. Fragoso, H. Lischke, and K. Thompson. 2005. Forecasting regional to global plant migration in response to climate change. Bioscience **55**:749-759.

Nunney, L. and D. R. Elam. 1994. Estimating the effective population size of conserved populations. Conservation Biology **8**:175-184.

Oak, S. W., C. M. Huber, and R. M. Sheffield. 1986. Incidence and impact of oak decline in western Virginia, 1986. United States Department of Agriculture, Forest Service.

Oosting, H. J. 1942. An ecological analysis of the plant communities of Piedmont, North Carolina. American Midland Naturalist **28**:1-126.

Overpeck, J. T., P. J. Bartlein, and T. Webb. 1991. Potential magnitude of future vegetation change in eastern North America: Comparisons with the past. Science **254**:692-695.

Pairon, M., M. Jonard, and A.-L. Jacquemart. 2006. Modeling seed dispersal of black cherry, an invasive forest tree: how microsatellites may help? Canadian Journal of Forest Research **36**:1385-1394.

Palmer, E. J. 1942. The red oak complex in the United States. American Midland Naturalist:732-740.

Peakall, R. and P. E. Smouse. 2006. Genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes **6**:288-295.

Pearson, R. G. 2006. Climate change and the migration capacity of species. TREE **21**:111-113.

Pearson, R. G. and T. P. Dawson. 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? Global Ecology & Biogeography **12**:361-371.

Pedersen, B. S. and A. M. Wallis. 2004. Effects of white-tailed deer herbivory on forest gap dynamics in a wildlife preserve, Pennsylvania, USA. Natural Areas Journal **24**:82-94.

Pemberton, J. M. 2008. Wild pedigrees: the way forward. Proc. Roy. Soc. B.

Penaloza-Ramirez, J. M., A. Gonzalez-Rodriguez, L. Mendoza-Cuenca, H. Caron, A. Kremer, and K. Oyama. 2010. Interspecific gene flow in a multispecies oak hybrid zone in the Sierra Tarahumara of Mexico. Annals of Botany **105**:389-399.

Petit, R. J., R. Bialozyt, P. Garnier-Gere, and A. Hampe. 2004. Ecology and genetics of tree invasions: from recent introductions to Quaternary migrations. Forest Ecology and Management **197**:117-137.

Petit, R. J., C. Bodenes, A. Ducousso, G. Roussel, and A. Kremer. 2003. Hybridization as a mechanism of invasion in oaks. New Phytologist **161**:151-164.

Petit, R. J., S. Brewer, S. Bordacs, K. Burg, R. Cheddadi, E. Coart, J. Cottrell, U. M. Csaikl, B. van Dam, J. D. Deans, S. Espinel, S. Fineschi, R. Finkeldey, I. Glaz, P. G. Goicoechea, J. J.S., A. O. Konig, A. J. Lowe, S. F. Madsen, G. Matyas, R. C. Munro, F. Popescu, D. Slade, H. Tabbener, S. G. M. de Vries, B. Ziegenhagen, J. L. de

Beauliu, and A. Kremer. 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.

Pitelka, L. F. 1997. Plant migration and climate change. American Scientist **85**:464-474.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics **155**:945-959.

Purves, D. W., M. A. Zavala, K. Ogle, F. Prieto, and J. M. Rey Benayas. 2007. Environmental heterogeneity, bird-mediated directed dispersal, and oak woodland dynamics in mediterranean Spain. Ecological Monographs **77**:77-97.

Rehfeldt, G. E., C. C. Ying, D. L. Spittlehouse, and D. A. Hamilton. 1999. Genetic responses to climate change in Pinus contorta: niche breadth, climate change, and reforestation. Ecological Monographs **69**:375-407.

Reid, C. 1899. The origin of the British flora. Dulau, London.

Rizzo, D. M. and M. Garbelotto. 2003. Sudden oak death: endangering California and Oregon forest ecosystems. Frontiers in Ecology and the Environment **1**:197-204.

Rooney, T. P. 2001. Deer impacts on forest ecosystems: a North American perspective. Forestry **74**:201-208.

Roth, J. K. and S. B. Vander Wall. 2005. Primary and secondary seed dispersal of bush chinquapin (Fagaceae) by scatterhoarding rodents. Ecology **86**:2428-2439.

Runkle, J. R. 1982. Patterns of disturbance in some old-growth mesic forests of eastern North America. Ecology **63**:1533-1546.

Rushton, B. S. 1993. Natural hybridization within the genus *Quercus* L. Ann Sci For **50**:73s-90s.

Saito, H., T. Nakaguchi, C. Kugo, and M. Takeoka. 1987. Dry matter production of each component of reproductive organs and the relation of the number of pollen grains and female flowers to seed crop in a Konara oak (*Quercus serrata*) mature stand. Scientific Reports of the Kyoto Prefectural University **39**:40-48.

Salvini, D., P. Bruschi, S. Fineschi, P. Grossoni, E. D. Kjaer, and G. G. Vendramin. 2009. Natural hybridisation between *Quercus petraea* (Matt.) Liebl. and *Quercus pubescens* Willd. within an Italian stand as revealed by microsatellite fingerprinting. Plant Biology **11**:758-765.

Sato, T., Y. Isagi, H. Sakio, K. Osumi, and S. Goto. 2006. Effect of gene flow on spatial genetic structure in the riparian canopy tree *Cercidiphyllum japonicum* revealed by microsatellite analysis. Heredity **96**:79-84.

Schnabel, A., J. Nason, and J. L. Hamrick. 1998. Understanding the population genetic structure of *Gleditsia triacanthos* L.: seed dispersal and variation in female reproductive success. Molecular Ecology **7**:819-832.

Schoen, D. J. and S. C. Stewart. 1986. Variation in male reproductive investment and male reproductive success in white spruce. Evolution **40**:1109-1120.

Schueler, S., K. H. Schlunzen, and F. Scholz. 2005. Viability and sunlight sensitivity of oak pollen and its implication for pollen-mediated gene flow. Trees **19**:154-161.

Schwartz, M. W. 1993. Modelling effects of habitat fragmentation on the ability of trees to respond to climatic warming. Biodiversity and Conservation **2**:51-61.

Schwarzmann, J. F. and H. D. Gerhold. 1991. Genetic structure and mating system of Northern Red Oak (*Quercus rubra* L.) in Pennsylvania. Forest Science **37**:1376-1389.

Selkoe, K. A. and R. J. Toonen. 2006. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. Ecology Letters **9**:615-629.

Sezen, U. U., R. L. Chazdon, and K. E. Holsinger. 2005. Genetic consequences of tropical second-growth forest regeneration. Science **307**:891.

Shure, D. J., D. L. Phillips, and P. E. Bostick. 2006. Gap size and succession in cutover southern Appalachian forests: an 18 year study of vegetation dynamics. Plant Ecology **185**:299-318.

Skelly, D. K., L. N. Joseph, H. P. Possingham, L. K. Freidenburg, T. J. Farrugia, M. T. Kinnison, and A. P. Hendry. 2007. Evolutionary responses to climate change. Conservation Biology **21**:1353-1355.

Slatkin, M. 1995. A measure of population subdivision based on microsatellite allele frequencies. Genetics **139**:1463.

Smouse, P. E. and R. Peakall. 1999. Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. Heredity **82**:561-573.

Sork, V. L. 1984. Examination of seed dispersal and survival in red oak, *Quercus rubra* (Fagaceae), using metal-tagged acorns. Ecology **65**:1020-1022.

Sork, V. L., F. W. Davis, P. E. Smouse, V. J. Apsit, R. J. Dyer, and J. F. Fernandez-M. 2002. Pollen movement in declining populations of California Valley oak, *Quercus lobata*: where have all the fathers gone? Molecular Ecology **11**:1657-1668.

Spetich, M. A. 2004. Upland oak ecology symposium:a synthesis.*in* Upland oak ecology symposium:history, current conditions, and sustainability. US Department of Agriculture, Southern Research Station, Asheville,NC.

Stapanian, M. A. and C. C. Smith. 1986. How fox squirrels influence the invasion of praries by nut-bearing trees. Journal of Mammalogy **67**:326-332.

Steele, M. A., P. D. Smallwood, W. B. Terzaghi, J. E. Carlson, T. Conreras, and A. McEuen. 2004. Oak dispersal syndromes: do red and white oaks exhibit different dispersal strategies? , Department of Agriculture, Forest Service, Southern Research Station, Asheville, NC.

Stephenson, S. L. 1986. Changes in a former chestnut-dominated forest after a half century of succession. American Midland Naturalist **116**:173-178.

Streiff, R., A. Ducousso, C. Lexer, H. Steinkellner, J. Gloessl, and A. Kremer. 1999. Pollen dispersal inferred from paternity analysis in a mixed oak stand of *Quercus robur* L. and *Q. petraea*(Matt.)Liebl. Molecular Ecology **8**:831-841.

Streiff, R., T. Labbe, R. Bacilieri, H. Steinkellner, J. Glossl, and A. Kremmer. 1998. Within-population genetic structure in *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl. assessed with isozymes and microsatellites. Molecular Ecology **7**:317-328.

Terborgh, J., G. Nunez-Iturri, N. C. A. Pitman, F. H. C. Valverde, P. Alvarez, V. Swamy, E. G. Pringle, and C. E. T. Paine. 2008. Tree recruitment in an empty forest. Ecology **89**:1757-1768.

Thomas, C. D., A. M. A. Franco, and J. K. Hill. 2006. Range retractions and extinction in the face of climate warming. TREE **21**.

Thomas, L., S. T. Buckland, E. A. Rexstad, J. L. Laake, S. L. Strindberg, S. L. Hedley, J. R. B. Bishop, T. A. Marques, and K. P. Burnham. 2010. Distance software: design and analysis of distance sampling surveys for estimating population size. Journal of Applied Ecology.

Thuiller, W., C. Albert, M. B. Araujo, P. M. Berry, M. Cabeza, A. Guisan, T. Hickler, G. F. Midgley, J. Paterson, F. M. Schurr, M. T. Sykes, and N. E. Zimmermann. 2008. Predicting global change impacts on plant species' distributions: future challenges. Perspectives in plant ecology, evolution, and systematics **9**:137-152.

Thuiller, W., S. Lavorel, M. B. Araujo, M. T. Sykes, and C. Prentice. 2005. Climate change threats to plant diversity in Europe. PNAS **102**:8245-8250.

Tomlinson, P. T., R. J. Jensen, and J. F. Hancock. 2000. Do whole tree silvic characters indicate hybridization in red oak (*Quercus* Section *Lobatae)?* American Midland Naturalist **143**:154-168.

Tovar-Sanchez, E. and K. Oyama. 2004. Natural hybridization and hybrid zones between *Quercus crassifolia* and *Quercus crassipes* (Fagaceae) in Mexico: morphological and molecular evidence. American Journal of Botany **91**:1352-1363.

Tramer, E. J. 1993. Breeding bird censuses at the Coweeta Hydrologic Laboratory: A comparison of 1967 and 1993.

Troupin, D., R. Nathan, and G. G. Vendramin. 2006. Analysis of spatial genetic structure in an expanding *Pinus halepensis* population reveals development of fine-scale genetic clustering over time. Molecular Ecology **15**:3617-3630.

Valbuena-Carabana, M., S. C. Gonzalez-Martinez, O. J. Hardy, and L. Gil. 2007. Fine-scale spatial genetic structure in mixed oak stands with different levels of hybridization. Molecular Ecology **16**:1207-1219.

Van der Veken, S., M. Hermy, M. Vellend, A. Knapen, and K. Verheyen. 2008. Garden plants get a head start on climate change. Frontiers in Ecology and the Environment **6**:212-216.

Vander Wall, S. B. 2001. The evolutionary ecology of nut dispersal. The Botanical Review **67**:74-117.

Walther, G. R., E. Post, P. Convey, A. Menzel, C. Parmesan, T. J. C. Beebee, J. M. Fromentin, O. Hoegh-Guldberg, and F. Bairlein. 2002. Ecological responses to recent climate change. Nature **416**:389-395.

Weber, E. F. 1997. The alien flora of Europe: a taxonomic and biogeographic review. Journal of Vegetation Science **8**:565-572.

Weiner, J. and O. T. Solbrig. 1984. The meaning and measure of size hierarchies in plant populations. Oecologia **61**:334-336.

Whittemore, A. T. and B. A. Schaal. 1991. Interspecific gene flow in sympatric oaks. Proc. Natl. Acad. Sci. USA **88**:2540-2544.

Willis, K. J. and K. J. Niklas. 2004. The role of Quaternary environmental change in plant macroevolution: the exception or the rule? Phil. Trans. Royal. Soc. B. **359**:159-172.

Wolfe, L. M. 2002. Why alien invaders succeed: support for the escape-from-enemy hypothesis. the American Naturalist **160**:705-711.

Biography

Emily Moran was born in 1983 in Bloomington, Indiana. She graduated from the University of Michigan in 2005 with highest honors in biology. She began graduate study at Duke University that same year, and is the recipient of a number of awards, including fellowships from the Center for Theoretical and Mathematical Sciences and the Association for Women in Science.