Tree Genetics & Genomes (2014) 10:541–554
DOI 10.1007/s11295-014-0702-8

ORIGINAL PAPER

Biotic and abiotic factors affecting the genetic structure and diversity of butternut in the southern Appalachian Mountains, USA

Amanda Parks · Michael Jenkins · Michael Ostry · Peng Zhao · Keith Woeste

Received: 28 June 2013 / Revised: 26 January 2014 / Accepted: 30 January 2014 / Published online: 16 February 2014

Abstract The abundance of butternut (*Juglans cinerea* L.) trees has severely declined rangewide over the past 50 years. An important factor in the decline is butternut canker, a disease caused by the fungus *Ophiognomonia clavigigenti-juglandacearum*, which has left the remaining butternuts isolated and sparsely distributed. To manage the remaining populations effectively, information regarding how butternut's population genetic structure is affected by environmental and historical factors is needed. In this study, we assessed genetic structure and diversity of 161 butternut trees from 19 adjacent watersheds in the southern portion of butternut's range using 12 microsatellite markers. We assessed the genetic diversity and genetic differentiation among trees grouped at various spatial scales. Our goal was to use historical abundance and land use data for these watersheds, which are now all a part of the Great Smoky Mountains National Park (GSMNP), to understand the ecological and evolutionary forces that challenge the conservation and management of butternut. In general, butternuts within the 19 neighboring watersheds were all part of one continuous population, with gene flow throughout. Significant genetic differentiation was detected between some groups of trees, but the differentiation was quite small and may not represent an ecologically significant distinction. The mean heterozygosity in all watersheds remained high, despite extensive mortality. Overall, genetic diversity and rare alleles were evenly distributed across all watersheds, with some variability in subpopulations containing butternut-Japanese walnut hybrids (*Juglans* x *bixbyi* or buarts). These results indicate that management of this species should focus on protection from future hybridization with Japanese walnut, promotion of regeneration, and persistence of all remaining butternut trees, which still retain high levels of genetic diversity.

Keywords Butternut canker disease · *Juglans cinerea* · Forest genetics · Conservation genetics · Hardwood · Hybridization · *Ophiognomonia clavigigenti-juglandacearum*

Communicated by G. G. Vendramin

Electronic supplementary material The online version of this article (doi:10.1007/s11295-014-0702-8) contains supplementary material, which is available to authorized users.

A. Parks · M. Jenkins
Department of Forestry & Natural Resources, Purdue University, 715 W. State St., West Lafayette, IN 47907, USA

M. Ostry
Northern Research Station, USDA Forest Service, 1561 Lindig Ave., St. Paul, MN 55108, USA

P. Zhao
Key Laboratory of Resource Biology and Biotechnology in Western China, Ministry of Education, College of Life Science, Northwest University, Xi'an, Shaanxi 710069, China

K. Woeste (✉)
USDA Forest Service, Northern Research Station, Hardwood Tree Improvement & Regeneration Center, Purdue University, 715 W. State St., West Lafayette, IN 47907, USA
e-mail: woeste@purdue.edu

Introduction

Effective management of threatened plant species requires an integrated understanding of their genetic structure and the historic and current ecological processes that resulted in the observed distribution of their genetic diversity (Sork and Waits 2010). In the absence of such an inclusive approach, it is difficult to decipher the effects of individual processes (Loveless and Hamrick 1984; Eckert et al. 2010) because different processes may result in similar patterns of genetic structure on environmentally similar sites (Kalisz et al. 2001). Conversely, dissimilar establishment dynamics can result in different degrees of genetic structure on similar sites (Knowles

et al. 1992; Epperson and Chung 2001). Over time, changes in demography and reproductive attributes such as successional stage and number of generations present interact with these ecological processes to affect genetic diversity (Loveless and Hamrick 1984; Epperson and Chung 2001; Jones et al. 2006). Despite their relevance, detailed, long-term site history data are not available for many sites and relatively few population genetic studies incorporate historic data (Jones et al. 2006).

The importance of habitat heterogeneity in the development of spatial genetic structure has been highlighted in previous studies (Loveless and Hamrick 1984; Kalisz et al. 2001; Kitamoto et al. 2005), as well as in the growing field of landscape genetics (Storfer et al. 2010; Sork and Waits 2010). Observations at a range of spatial scales may be required to understand reproductive connectivity (Loiselle et al. 1995; Vekemans and Hardy 2004; Kitamoto et al. 2005), and local differences in environmental conditions may lead to the emergence of different patterns of genetic structure within subpopulations (Epperson and Chung 2001; Kitamoto et al. 2005; Jones et al. 2006). Whatever the cause, restricted gene flow and subsequent loss of diversity implies decreased adaptability of populations (Sork and Smouse 2006; Holderegger and Wagner 2008; Holderegger et al. 2010).

Butternut (*Juglans cinerea* L.) is a short-lived forest tree species severely affected by butternut canker disease (Schultz 2003), which is caused by the fungal pathogen *Ophiognomonia clavigigenti-juglandacearum* [(Nair, Kostichka, & Kuntz) Broders & Boland] (=*Sirococcus clavigigenti-juglandacearum*). Because host genetic diversity can critically influence the resilience of populations exposed to disease (Altizer et al. 2003), maintenance of diversity must be a priority in butternut conservation efforts (Ostry and Woeste 2004). Recent studies have shown that despite rangewide declines in abundance, butternut's genetic diversity remains high (Ross-Davis et al. 2008b). As a wind-pollinated species, butternut has widespread pollen dispersal, but genetic isolation may occur as populations become increasingly fragmented by disease, poor regeneration, and habitat loss (Ross-Davis et al. 2008b; Hoban et al. 2012a). In order to maintain connectivity and diversity of the remnant populations of butternut, better resolution of relevant scales of gene flow in various environments will be necessary (Ross-Davis et al. 2008b). The genetic integrity of butternut is also threatened by hybridization with Japanese walnut (*Juglans ailantifolia* Carr.), a nonnative congener which has been planted throughout the range of butternut since the mid-19th century. The identification of hybrids is necessary for accurate interpretation of diversity estimates for butternut, and protection from future hybridization may be a necessary part of conservation efforts (Hoban et al. 2009).

This study takes a broad view of processes that potentially affect the genetic health of butternut populations: current and historic demographics, small- and large-scale processes, and natural and anthropogenic disturbances. We examined the population genetics of butternut in 19 neighboring watersheds of the southern Appalachian Mountains that are all contained within the 210,000-ha Great Smoky Mountains National Park (GSMNP). GSMNP is recognized internationally as a center of biological diversity in North America and was designated an International Biosphere Reserve in 1976. GSMNP serves as a bellwether and a baseline for management of state and federal lands and biological communities in the eastern USA (Jenkins and White 2002; Jenkins 2007). The relatively large census of butternut in the Park and the extensive records related to vegetation types and the history of land use in GSMNP represent a rare opportunity to understand the forces shaping the reproductive dynamics of the species (Hoban et al. 2012a). By sampling a large site with complex topography and vegetation types, as well as long-term site history information, we sought to understand the complex interactions occurring within butternut populations and to translate results into meaningful management recommendations. Our research questions were as follows: (1) Is there genetic structure within and among butternut subpopulations in the sampled watersheds? If so, is butternut's genetic structure affected by population density, age structure, disease severity/mortality, topography, or hybridization? (2) What is the relative similarity of individuals at different spatial scales? (3) What is the occurrence and distribution of rare alleles within the sampled sites and do subpopulations affected by hybridization exhibit distinctive patterns of genetic diversity?

Methods

Study site

Great Smoky Mountains National Park encompasses over 210,000 ha in the southern Appalachian Mountains. Topography in the park is rugged and variable with elevations ranging from 267 to over 2,000 m (Jenkins 2007). Interacting environmental gradients result in a high level of biological diversity (Jenkins 2007), and GSMNP serves a vital role in biological conservation within the southern USA. Although now protected, the Park was once extensively settled with farms and homesteads distributed across major watersheds (Pyle 1985). Varying degrees of logging occurred across nearly 75 % of the Park prior to its establishment in 1934 (Pyle 1988), but forest communities in GSMNP have received minimal human disturbance over the past 75 years. The exact date of the arrival of butternut canker to GSMNP is unknown. Forest surveys conducted in the 1960s upon the initial reports of the disease revealed extensive mortality in the southeastern USA (Ostry and Woeste 2004), indicating the disease had been active in the region for several years. Within our sample areas (Fig. 1a), butternut trees remain in isolated clusters of

individuals distributed across low to mid-elevation sites, often within riparian corridors. Most sampled watersheds were separated by high mountain ridges.

Foliage and tree core sampling

We used a monitoring database compiled and maintained by the National Park Service (NPS) from contemporary and historic observations and surveys to locate butternut trees across GSMNP (Glenn Taylor, NPS, personal communication). Several additional groups of trees were located based upon personal communication with NPS employees. We collected leaf samples for genetic analysis from 161 individuals (including 14 seedlings and saplings) in 19 watersheds (Fig. 1). The sampled trees represented, to the best of our knowledge, a complete census of the living butternut within each sampled watershed. In a prior assessment of these trees, four individuals in the Hazel Creek watershed and one in the Abrams Creek watershed were identified as butternut-Japanese walnut hybrids using nuclear, chloroplast, and mitochondrial DNA markers (Zhao and Woeste 2010; Parks et al. 2013). We also collected tree cores from a subset of trees across the range of our sampling areas (see Parks et al. 2013 for details). The age of the sampled trees was used to assess genetic differentiation among age cohorts.

Genotyping

DNA was extracted using a chloroform-phenol extraction method described by Zhao and Woeste (2010). DNA concentration was measured using a NanoDrop-8000 spectrophotometer (Thermo Scientific, Wilmington, DE). Each individual was initially genotyped at 15 microsatellite loci (Hoban et al. 2008; Ross-Davis and Woeste 2008; Table 1). PCR amplifications were performed as described by Victory et al. (2006), Hoban et al. (2008), and Ross-Davis and Woeste (2008), and analyzed using an ABI 377 (Applied Biosystems, Foster City, CA) sequencer. We assessed the results using GeneMapper software (v.3.7; Applied Biosystems) to manually call allele sizes. Negative controls (nanopure water added in lieu of DNA template) were run with each marker and two positive controls (replicates of the same individual) were performed for each PCR reaction for scoring verification. To estimate our genotyping error rate, a subset of samples was reamplified and reanalyzed (15 samples were evaluated at least twice at all loci and five individuals were genotyped in triplicate). Other samples were replicated as necessary to obtain clear results. Estimates of genotyping error rates were determined based on the number of genotype calls that were not consistent between replicates. Consensus genotypes for these loci were determined by a final round of reruns for these samples/loci. All samples were amplified separately at each marker to reduce error and interactions encountered during multiplexed reactions.

Genetic analysis

Before performing genetic structure and diversity analyses, we used the software Microchecker (van Oosterhout et al. 2004) to check for the presence of null alleles. The program uses an algorithm to test each locus for an excess of homozygotes. Heterozygote deficiency distributed across loci occurs with admixtures (van Oosterhout et al. 2004) but can be indicative of null alleles or genotyping error when found only at individual loci. Loci that showed a significant deviation from expected heterozygosity, presumably due to null alleles, were flagged by the program and removed from subsequent analyses. For genetic structure assessment, we defined populations in two ways: by watershed (using designations described by Parker and Pipes 1990) and by north versus south of a high ridgeline that divides the Park into roughly equal parts and that may represent a gene flow barrier (corresponds to the Tennessee/North Carolina boundary) (Fig. 1a). These geographic units delineated subpopulations within the sampled butternuts as a whole, but they were also assessed as populations to evaluate hierarchical population structure. Watersheds containing only one individual (Hesse Creek and Straight Fork; Fig. 1b) were lumped with the nearest watershed (Abrams Creek and Oconaluftee River Lower, respectively), since population comparisons cannot be performed for a single individual. We calculated multilocus estimates of subpopulation differentiation using *F*-statistics [F_{IS} (individuals relative to their respective subpopulation), F_{IT} (individuals relative to the total sample), and F_{ST} (divergence of the subpopulation relative to the total sample; Weir and Cockerham 1984) and pairwise F_{ST} (relative genetic divergence between pairs of subpopulations)] and differentiation between all subpopulations using FSTAT (v.2.9.3.2; Goudet 1995). Relative pairwise F_{ST} values were averaged for each watershed to represent the differentiation of each from all other subpopulations. We estimated a second measure of genetic differentiation (D_{EST}), as calculated by SMOGD (v.1.2.5; Crawford 2010), a software that employs the methods described by Jost (2008). We used HP-RARE (v.1.1; Kalinowski 2005) to calculate allelic richness and private allelic richness with rarefaction to account for differences in sample size within watersheds and within north versus south subpopulations. For the watershed-scale assessment, the smallest sample with complete genotypes contained two individuals; since this is quite low, we omitted watersheds with less than four individuals (three watersheds) and allelic richness within watersheds was adjusted to eight genes (four individuals). The north-versus-south (of the main ridgeline) comparisons were adjusted to the number of individuals with complete genotypes at all loci in the smallest sample: 98

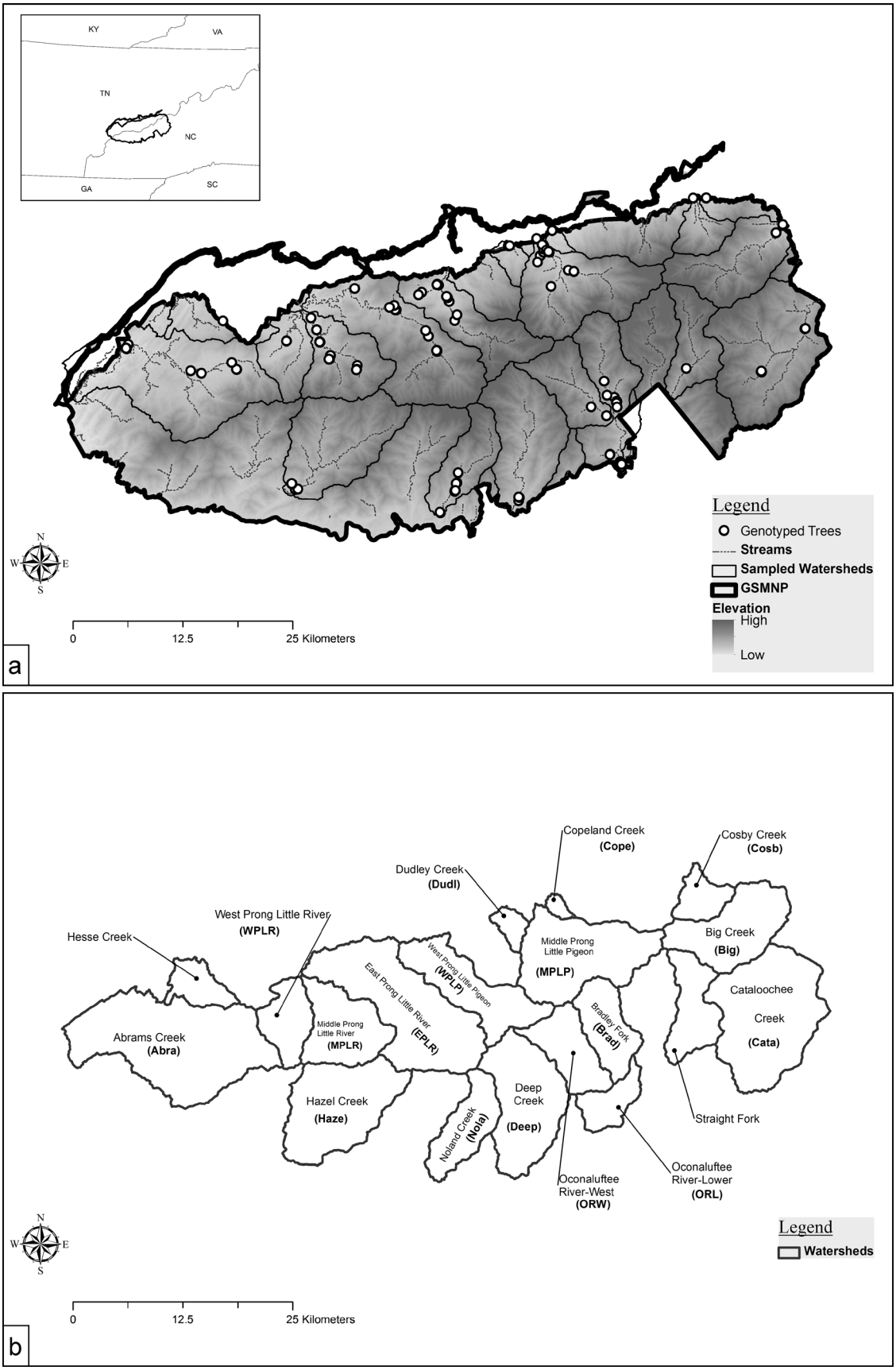

Fig. 1 **a** Locations of genotyped butternut trees within GSMNP; **b** Names and abbreviations of sampled watersheds

Table 1 Microsatellite loci used for genotyping (Victory et al. 2006; Hoban et al. 2008; Ross-Davis and Woeste 2008)

Locus	Total observed alleles	Expected allelic range (bp)	Observed allelic range (bp)	HWE Test Results	
				Expected heterozygosity	Observed heterozygosity
AAG0001*	14	148–172[a]	155–181	0.785*	0.430*
JCINA5	13	195–221	193–219	0.858	0.870
JCINB114*	20	249–289	248–284	0.884*	0.748*
JCINB121	14	171–201	163–193	0.884	0.845
JCINB147	9	327–345	323–341	0.827	0.758
JCINB159	22	103–162	105–173	0.877	0.841
JCINB262	18	316–358	315–355	0.875	0.845
WAG004	19	225–273	224–274	0.857	0.789
WAG082	16	150–182	152–186	0.881	0.849
WAG090	9	126–144	124–142	0.524	0.522
WAG142	11	161–199	168–198	0.758	0.739
WAG148	15	232–282	234–278	0.750	0.675
WAG204*	20	168–200	168–194	0.850*	0.741*
WAG221	9	221–247	221–245	0.709	0.706
WAG256	17	205–241	205–253	0.883	0.826

Tests for Hardy-Weinberg equilibrium performed using Microchecker (van Oosterhout et al. 2004); those loci marked with * had significant (α=0.05) deviations from expected values and were removed from the dataset prior to population analyses

[a] Allele range for *J. nigra*; marker optimized for *J. nigra* (Victory et al. 2006) but used in this analysis to determine its utility in *J. cinerea* as well

alleles (39 individuals). Observed and expected heterozygosity were calculated for watershed and north/south population delineations to assess gene diversity within subpopulations and overall. Hybrid individuals were retained in the dataset for determining differentiation at the population level and to determine if subpopulations with hybrids have unique genetic diversity.

We used Bayesian cluster analysis (Pritchard et al. 2000) to assess genetic structure in the sampled area. First, the program STRUCTURE (v.2.3.3, 2010) was used to estimate the true number of distinct populations within our sample. We ran each scenario with burn-in of 10,000 iterations and 50,000 Markov chain Monte Carlo (MCMC) iterations, with number of populations (K) defined as one to twenty with 20 replicates for each value of K. We chose these values of K to evaluate the possibility of watersheds (n=19) as a population definition. We ran each scenario using both the uncorrelated allele frequency model, which assumes allele frequencies in each population are independent, and the correlated allele frequency model, which takes into account allele frequency similarity in populations that recently diverged (Pritchard et al. 2000; Falush et al. 2003). The true number of populations (K) was inferred using Evanno's method, which uses the ad hoc statistic Δk, which is based upon the rate of change of the log probability of data between different values of K (Evanno et al. 2005). We performed two identical analyses for each model: one with and without a dummy (out-group) population created to force a clear division in clusters. The dummy population was based on a dataset of genotypes from a distant butternut population. This method was used to evaluate the strength of clustering without the dummy population. To further assess genetic structure, we used Geneland (Guillot et al. 2005b), which also uses Bayesian clustering based on multilocus genotypes to estimate the number of populations represented in a sample. Geneland, however, takes spatial location into account when assessing genetic structure and can be used to determine spatial relationships and landscape features that serve as gene flow barriers (Guillot et al. 2005a). Using the Dirichlet (uncorrelated) model, which performs well for high and low levels of differentiation (Guillot et al. 2005a), we performed five independent runs with possible values of K set 1 to 20, using 100,000 MCMC iterations and 1,000 thinning. To ensure estimates of K were not affected by the error associated with collection of spatial locations, five runs were performed with coordinate uncertainty set at both 0 and 10 m, which represents the approximate accuracy of the GPS unit that was used to record the coordinates of all individuals. Finally, population structure was assessed using analysis of molecular variance (AMOVA), which incorporates a hierarchical model to partition genetic variation among subpopulations at various hierarchical levels (i.e., watersheds, north/south halves, entire Park) (Peakall and Smouse 2012). AMOVA was performed using GenAlEx 6.5 (Peakall and Smouse 2012) using the allelic distance matrix to calculate variance distribution and AMOVA-based F-statistics. The p-values were based on 999 standard permutations.

SPAGeDi (v.1.3; Hardy and Vekemans 2002, 2007) was used to assess spatial and genetic pairwise relationships at the subpopulation and individual levels. Using ArcMap (v. 9.3.1; ESRI 2009), we calculated the average location for all

individuals within each subpopulation. At the population level, measures of pairwise genetic distances were $F_{ST}/1-F_{ST}$ (relative genetic differentiation), G_{ST} (coefficient of gene differentiation; Pons and Petit 1995), and D_s (Nei's 1978 standard genetic distance). Pairwise matrices of genetic and spatial distances between watersheds were obtained and used in subsequent analyses (data not shown). For comparisons of genetic differentiation of individuals, we calculated pairwise kinship coefficients between all individuals in our sample (Loiselle et al. 1995). Kinship coefficients between individuals are based on the probability of sampling two alleles that are identical by descent (from the same ancestral source). Since pedigrees are not available for wild populations, direct measures of kinship cannot be calculated because of homoplasy. The estimator of kinship used here was based on the correlation coefficient between allelic states (Loiselle et al. 1995). An average kinship value was calculated for each watershed by averaging the kinship coefficients between all pairwise combinations of individuals within that watershed. To evaluate the relationship between density of individuals and the average pairwise kinship within a watershed, average kinship values were plotted against the natural logarithm of the average pairwise spatial distances between trees within a watershed. Since Oconaluftee River (West) contained only two individuals, it was omitted from this analysis. Simple linear regression analysis was used to quantify the relationship between density and kinship within watersheds.

To evaluate isolation by distance at the watershed scale, we compared spatial and genetic distance matrices (from SPAGeDi) using the program PASSaGE (v.2; Rosenberg and Anderson 2011) to perform Mantel tests. For the population-level analysis, we compared a pairwise matrix of natural logarithm of spatial distances to three matrices of pairwise genetic distances (independently): D_s (Nei's standard genetic distance, Nei 1978), G_{ST} (Pons and Petit 1995), and $F_{ST}/1-F_{ST}$ (Weir and Cockerham 1984). Correlations between the spatial and genetic matrices were obtained with a corresponding two-tailed p value.

To assess genetic similarity based on nonspatial factors, trees were grouped into categories based on habitat and health. Although we used neutral markers that are not explicitly linked with adaptive traits, we sought to assess the effects of nonrandom mating and long-term selection resulting from differences in disease response, habitat, and age of individuals. These factors can lead to assortative mating, since healthy trees may produce more reproductive structures and more fit offspring than diseased trees, and because tree age may affect survival after infection. Three health measures for each individual (vigor class, percent of basal area girdled by cankers, number of cankers below DBH) were recorded during field sampling (Parks et al. 2013). These were combined to give an overall health rating for each tree. The trees were grouped into 10 categories based on this score, and genetic differentiation between these groups was assessed using FSTAT (v.2.9.3.2; Goudet 1995) to calculate F-statistics and pairwise F_{ST} values between groups.

We also evaluated genetic differentiation between the two dominant age classes within the trees that had core samples (n=60), since divergence between age cohorts has been demonstrated in other studies of butternut (Hoban et al. 2012a). Cohort genetic differentiation in tree populations is presumably due to long-term selection pressures on the older cohort, resulting in unequal genetic contribution of the older cohort and relative lack of selective thinning in the new generation (Hamrick 1982). The "old" cohort (n=20) was defined as all trees that established prior to 1955 since this represents the primary break between the two recruitment peaks that were seen across all watersheds (Parks et al. 2013). All trees that established after 1955 were classified as "young" (n=40). We calculated global F-statistics (Weir and Cockerham 1984), heterozygosity (gene diversity), D_{EST}, gene differentiation (corrected for differences in sample size), and allelic richness with rarefaction to a common population size of 18 individuals using the same techniques as described above for spatial structure evaluation.

Results

Three of the 15 microsatellite loci (AAG001, B114, and WAG204) were identified as likely to contain null alleles (Table 1), so these were removed from the dataset, leaving a total of 12 loci that were used in the final population analyses. Based on the replicated samples, the genotyping error rate, averaged across all loci was 2.2 % when problematic loci were included (AAG001, B114, WAG204) and dropped to 1.1 % when these three loci were removed. An error was defined as a replication for which the allele calls were different than the consensus genotype (includes both false alleles and allelic dropout). The numbers of alleles per locus were consistent with reasonable values based upon the locus diversity reported in previous studies using these markers (Victory et al. 2006; Hoban et al. 2008; Ross-Davis et al. 2008b; Ross-Davis and Woeste 2008). Pairwise comparisons of F_{ST} revealed that some watersheds differed significantly (α=0.05) from others (Supplemental Table 1), but overall, watershed boundaries did not represent units of population structure. D_{EST} values showed low levels of population differentiation overall (data not shown). Trees in the northern versus southern side of the Park showed low but significant differentiation (F_{ST}=0.012; p=0.05). The D_{EST} values for northern versus southern trees were quite low (D_{EST}=0.031), indicating that although a difference was detectable, the magnitude of this separation was quite small. Pairwise comparisons of all watershed subpopulations showed that Copeland Creek, West Prong Little River, and Cosby Creek were the most genetically distinct

subpopulations, with average relative pairwise F_{ST} values of 0.091, 0.068, and 0.049, respectively (Fig. 2).

Bayesian clustering analysis indicated a strong tendency for all trees to group into a single population. In the STRUCTURE analyses, when a dummy population was included, there was a clear Δk maximum at k=2. When the out-group was removed, there was no single, definitive peak. Both the correlated and uncorrelated models gave similar results, with a clear peak at two when the out-group was included, and more variable, lower-magnitude peaks at several points for the GSMNP subpopulations alone. Geneland analyses resulted in estimates of k=1 for all 10 independent runs (5 with coordinate uncertainty=0; 5 with coordinate uncertainty set to 10 m). Typically, a second set of runs are performed with a fixed value of K, which was determined from the initial analyses, to determine population assignment of each individual to its appropriate cluster (Guillot et al. 2005a). However, since all individuals were assigned to a single population, this step was unnecessary in our analyses. AMOVA revealed low levels of variance at higher hierarchical levels, with 93 % of all variation within individuals (Table 2).

Pairwise multilocus kinship coefficients averaged within each watershed ranged from 0.001 to 0.165 (Fig. 3a). Although a range of values were obtained, the average kinship within watersheds was quite low. Pairwise kinship values between individuals within the Park as a whole ranged from −0.2325 to 0.7109, with an average of 0.0004 (data not shown). Regression analysis revealed a significant correlation between the density of individuals and the average kinship within a watershed (Fig. 3b). The Mantel tests comparing spatial and genetic distances between watersheds were not significant for any of the three measures of genetic distance (D_S: r=0.131, p value=0.262; $F_{ST}/1-F_{ST}$: r=0.103 p value=0.384; G_{ST}: r=0.061 p value=0.618).

Mean gene diversity within subpopulations overall was H_E=0.778 and 0.805 within watersheds and north/south subpopulations, respectively. Observed heterozygosity (H_O) ranged from 0.768 to 0.780 in the north/south subpopulations with a mean of 0.774. Within watersheds, H_O ranged from 0.646 to 0.917 (Table 3) across loci with a mean corrected for differences in sample sizes of 0.783 (Nei 1978).

Watershed	Average Pairwise F_{ST}	D_S
Abrams Creek	0.0376	0.1577
Big Creek	0.0302	0.0999
Bradley Fork	0.0199	0.0749
Cataloochee Creek	0.0416	0.1434
Copeland Creek	0.0908	0.3179
Cosby Creek	0.0486	0.2082
Deep Creek	0.0443	0.1846
Dudley Creek	0.0369	0.1402
East Prong Little River	0.0318	0.1297
Hazel Creek	0.0451	0.2519
Middle Prong Little Pigeon	0.0386	0.1826
Middle Prong Little River	0.0366	0.1587
Noland Creek	0.0428	0.1976
Oconaluftee River (Lower)	0.0404	0.1628
Oconaluftee River (West)	0.0177	0.0579
West Prong Little Pigeon	0.0220	0.0972
West Prong Little River	0.0678	0.2446

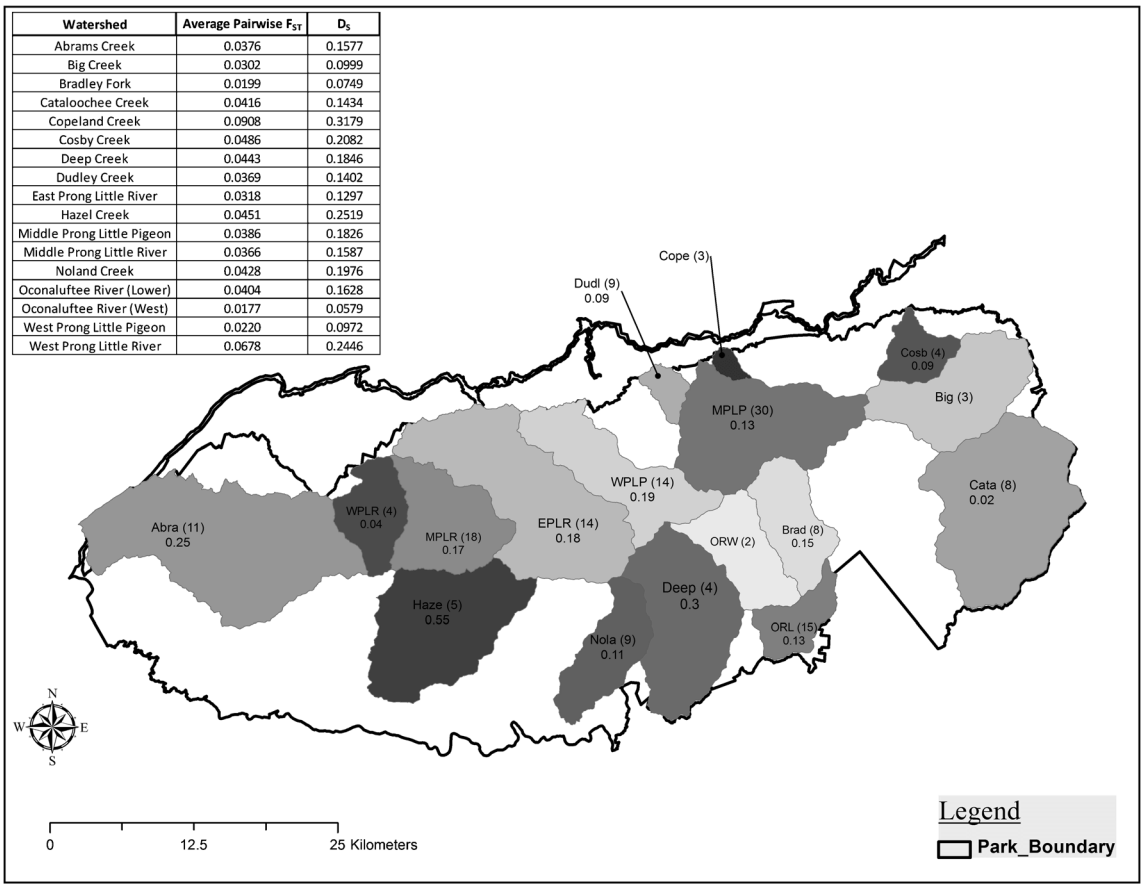

Fig. 2 Genetic differentiation among all sampled watersheds; *color* of watershed represents the average pairwise D_S to all other watersheds; watersheds with darker colors have the greatest genetic distance from all other watersheds; sample size (n) is given in parentheses and private allelic richness (with rarefaction to common sample size of 4) is below watershed abbreviation; Big Creek, Copeland Creek, and Oconaluftee River (West) had less than four individuals and were excluded from calculations of allelic richness; values of average pairwise D_S and F_{ST} of all watersheds are given in the *upper left corner of the map*

Table 2 Analysis of molecular variance (AMOVA) showing distribution of genetic variation at different hierarchical divisions of the GSMNP butternut population

	d.f.	Sum of squares	Estimated variance	Percentage of variation	*F*-statistics		
					Statistic	Value	*p* value
Among regions (north/south)	1	13.241	0.033	1	F_{RT}[a]	0.007	0.001
Among subpopulations (watersheds)/regions	15	116.560	0.163	3	F_{SR}[b]	0.033	0.001
Among individuals/subpopulations	144	697.695	0.135	3	F_{ST}[c]	0.040	0.001
Within Individuals	161	736.500	4.575	93	F_{IS}[d]	0.029	0.004
					F_{IT}[e]	0.067	0.001

[a] F_{RT}: variance among regions relative to the total variance

[b] F_{SR}: variance among subpopulations within regions

[c] F_{ST}: variance among subpopulations relative to the total variance

[d] F_{IS}: inbreeding coefficient of individuals relative to subpopulation

[e] F_{IT}: variance in the total population

Rare and private alleles appear to be relatively evenly distributed among subpopulations (Fig. 2; Table 3). All watersheds had private alleles, but the private allelic richness values were low and similar across all watersheds. Allelic richness for the north and south halves of the Park was 11.2 and 11.1, respectively. The north half of the Park had a private

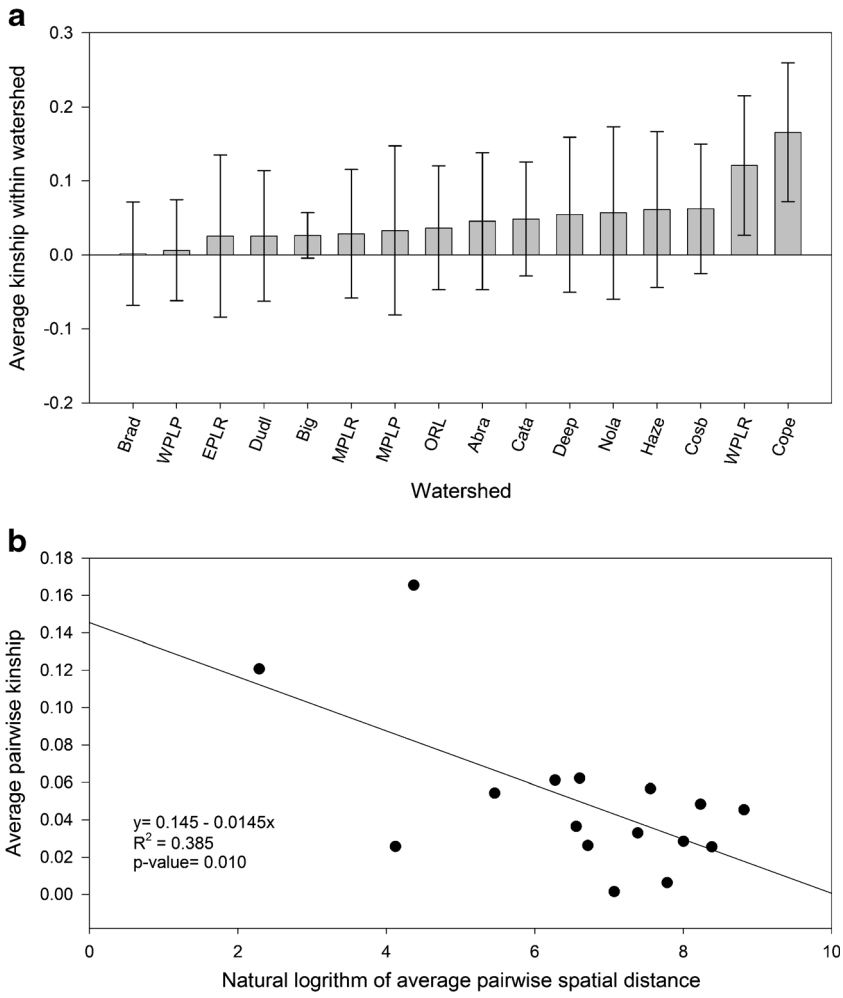

Fig. 3 **a** Distribution of average kinship within all watersheds, with standard deviation of within each watershed represented by *error bars*; see Fig. 1b for watershed abbreviations; **b** Natural logarithm of average pairwise geographic distance versus average pairwise kinship of all individuals within a watershed; *P* value from ANOVA

Table 3 Subpopulation differentiation at watershed scale within GSMNP

	F_{IS}	H_O	Allelic Richness	Private Allelic Richness
Abrams Creek	−0.015	0.801	4.52	0.25
Big Creek	−0.123	0.917	[a]	[a]
Bradley Fork	−0.017	0.820	4.49	0.15
Cataloochee Creek	−0.052	0.750	4.16	0.02
Copeland Creek	−0.033	0.681	[a]	[a]
Cosby Creek	−0.051	0.854	4.5	0.09
Deep Creek	−0.04	0.792	4.42	0.3
Dudley Creek	−0.07	0.863	4.33	0.09
East Prong Little River	0.051	0.714	4.35	0.18
Hazel Creek	0.157	0.646	4.6	0.55
Middle Prong Little Pigeon	0.077	0.748	4.44	0.13
Middle Prong Little River	0.007	0.786	4.5	0.17
Noland Creek	0.076	0.731	4.26	0.11
Oconaluftee River (Lower)	−0.051	0.828	4.32	0.13
Oconaluftee River (West)	0.122	0.792	[a]	[a]
West Prong Little Pigeon	0.044	0.771	4.54	0.19
West Prong Little River	−0.234	0.917	3.92	0.04

Mean Heterozygosity Across 12 Loci	
H_O	0.783
H_S (H_E)	0.778
H_T	0.812

Mean Differentiation±SE	
G_{ST}	0.0433±0.0079

Overall *F*-statistics		
	Mean ± SE	95 % Confidence Interval
F_{IT}	0.05±0.01	[0.031, 0.068]
F_{ST}	0.038±0.005	[0.028, 0.047]
F_{IS}	0.012±0.01	[−0.005, 0.032]

For the overall *F*-statistics and G_{ST}, standard error obtained by jackknifing over all loci; confidence intervals determined by bootstrapping over loci; for allelic richness, subpopulations were adjusted to a common sample size of n=4 (based on smallest sample with complete genotypes at all loci) using rarefaction (Kalinowski 2005)

[a] Richness values not given for watersheds with fewer than four individuals

allelic richness (averaged across loci) of 2.14, compared to 2.09 for the south, which indicates that about 20 % of the alleles at each locus were unique to individuals on one side of the main ridgeline. The Hazel Creek watershed, from which 80 % of the sampled trees were hybrids, contained the highest private allelic richness (0.55 averaged over all loci). Interestingly, one allele (allele 163 at locus B121) was found in all hybrid individuals (n=5) but in none of the butternuts.

Assessment of genetic similarity between trees grouped by health ratings revealed no significant pairwise F_{ST} between any of the health classes (F_{ST} values between −0.0265 and 0.0431). When trees were grouped based on the overstory forest type of their location, pairwise genetic comparisons revealed a few pairs that were significantly differentiated, but generally forest type classification did not represent distinct genetic groups. The largest pairwise F_{ST} between all overstory classes was 0.0917, but this value was between the two classes with the least individuals (n=3 & 5) and was not significant.

The young cohort of trees (established after 1955) showed very low and nonsignificant levels of differentiation from the old cohort (F_{ST}=0.0033, p=0.15). D_{EST} for these two classes was 0.001 (Table 4). Allelic richness was slightly higher in the older age class, but gene diversity was somewhat higher in the younger cohort (Table 4). Overall, there was little differentiation between the two age cohorts.

Discussion

Life history traits of a species and ecological characteristics of its habitat affect the degree and distribution of genetic diversity in populations by affecting dispersal, mating, and regeneration (Loveless and Hamrick 1984; Kalisz et al. 2001; Vekemans and Hardy 2004). The threats to butternut include disease, isolation, habitat fragmentation and loss, hybridization with nonnative congeners, and lack of regeneration due to alteration of disturbance regimes (Woeste et al. 2009; Hoban et al. 2012a; Parks et al. 2013). In order for managers to respond appropriately, it is necessary to identify the factors that amplify the effects of these threats.

Although recent studies have shown extensive gene flow for butternut at spatial scales greater than our study area (Ross-Davis et al. 2008b; Hoban et al. 2010), we hypothesized that the complex topography of GSMNP results in barriers to gene flow, as suggested by Ross-Davis et al. (2008b). This hypothesis was not confirmed; the topography of GSMNP was not a barrier to gene flow for butternut in the past. We observed some genetic variation among subpopulations, but the presence of strong hierarchical structure was not detected. AMOVA analysis indicates that little of the genetic variance in this population is due to divergence across the regions and subpopulations that we defined. Given the lack of genetic structure observed in *Juglans* species across huge geographic regions (Victory et al. 2006, 2008; Ross-Davis et al. 2008b),

Table 4 Genetic differentiation between old (established prior to 1955) and young (established during or after 1955) cohorts for a subsample of 60 cored individuals; for the overall *F*-statistics and G_{ST}, standard error obtained by jackknifing over all loci; confidence intervals determined by bootstrapping over loci; allelic richness for the young and old age cohorts, with rarefaction to a common population size of $n=18$ and averaged across all loci; gene diversity for both cohorts averaged across all loci

Genetic differentiation			
G_{ST}[a]	D_{EST}[b]		D_S[c]
0.0047±0.0045	0.0012		0.0191
Overall *F*-statistics			
F_{IT}[d]	F_{ST}[e]		F_{IS}[f]
0.076±0.019	0.003±0.005		0.073±0.019
[0.041, 0.111]	[−0.005, 0.012]		[0.036, 0.11]
	Allelic richness	Private allelic richness	Gene diversity
Old	8.73	1.77	0.7997
Young	8.41	1.46	0.8007

[a] Alternative estimator of F_{ST} (Pons and Petit 1995)

[b] Estimator of genetic differentiation based on Jost 2008; obtained using SMOGD software

[c] Nei's standard genetic distance (1978) between the two age cohorts

[d] F_{IT} averaged across all loci and individuals

[e] Pairwise F_{ST} between trees in the younger and older cohorts ± standard error, calculated by jackknifing over loci. 95 % confidence interval given in brackets, obtained by bootstrapping over loci

[f] F_{IS} averaged across all loci and individuals

these results are not extremely surprising, but the result is important because little investigation into the effect of topography on the genetic structure of temperate hardwoods has been completed. Victory et al. (2006) found that hydrologic units were not effective definitions of subpopulation structure in black walnut. They suggested that although the heavy seeds of *Juglans* species typically have limited dispersal distances, pollen flow eliminates genetic structure resulting from seed dispersal (Victory et al. 2006). Butternuts in some GSMNP watersheds differed from others, and F_{ST} values for subpopulations on opposite sides of a high (1,500–2,000 m) mountain ridge were significantly different from zero, but the magnitude of these differences was small, implying that all sampled individuals established as part of a continuous population with gene flow throughout. High levels of gene flow among populations, as well as the persistence of a small number of relatively healthy trees, can buffer the immediate effects of population declines in species with wide dispersal mechanisms (Hoban et al. 2010). Unrestricted gene flow is the best explanation for low levels of genetic structure observed across large spatial scales in *Juglans nigra*, a close relative of butternut (Victory et al. 2006). However, since nearly all the butternut trees in GSMNP are at least 30 years old (Parks et al. 2013) and widespread mortality has occurred in the years since these individuals established, the observed genetic structure may represent historic levels of connectivity (Holderegger et al. 2010; Segelbacher et al. 2010). Contemporary individuals may be more isolated than they were previously, and genetic structure and diversity assessments in the next generations may reveal marked changes in gene flow and genetic structure.

Although the number of butternuts in GSMNP has declined sharply due to disease and poor regeneration in the past 30 years (Parks et al. 2013), our results, along with previous studies, indicate that a substantial amount of diversity still remains for the species as a whole (Ross-Davis et al. 2008b; Hoban et al. 2010). Genetic diversity remains high across the 19 sampled watersheds ($H_E=0.81$; averaged over all loci). Diversity was also high within watersheds and within the north and south halves of the park ($H_E=0.778$ & 0.0805, respectively). These values are similar to levels observed in other studies of butternut populations (Ross-Davis et al. 2008b; Hoban et al. 2012a) and a related species (Victory et al. 2006). Allelic richness in GSMNP butternut was as high as or higher than others have observed in assessments over wider geographic areas (Ross-Davis et al. 2008b; Hoban et al. 2010). The average number of alleles per locus (A) observed within watersheds (A=4.2), for the north and south sections of the Park (A=11.1) and overall (A=14.3), along with our heterozygosity estimates, were consistent with other microsatellite studies for butternut (Ross-Davis et al. 2008b; Hoban et al. 2010, 2012a).

Private alleles were found within all watersheds in the Park, but were generally evenly distributed across the Park. No single area contained an abundance of rare alleles, with the exception of the Hazel Creek watershed, which was mostly comprised of hybrids. The loss of rare alleles may be a more sensitive measure of genetic consequences of disease-related

mortality (Kalinowski 2004; Ross-Davis et al. 2008b; Hoban et al. 2012a). Maintenance of rare alleles may be important for retaining the full range of genetic diversity of a species and identification of these alleles may be useful to elucidate areas or subpopulations that require special management considerations (Petit et al. 1998; Kalinowski 2004). We did not observe extensive declines in rare alleles between older and younger butternuts in our sample, as reported by Hoban et al. (2012a), although we found very few seedling and sapling butternuts in GSMNP. Scale-dependent processes and variability in site characteristics such as dispersal mechanisms, competition, and disturbance regimes, may explain variability in allele loss among sites. As butternut populations become more isolated and the survival of butternut becomes more dependent on management and human intervention, fine-scale processes will have a greater influence on the diversity and health of the species as a whole. A highly fragmented distribution of individuals and populations can decrease population diversity and increase reproductive isolation, making populations vulnerable to local stochastic events (Ellstrand and Elam 1993). Because the factors that affect survival and regeneration are varied and site-specific, an array of management techniques will be needed to mitigate the effects of disease and changing disturbance regimes.

Global F-statistics and pairwise F_{ST} did not indicate that trees categorized by their health ratings were genetically differentiated from trees in different health categories. Distribution of genotypes can be altered due to disease pressure (Altizer et al. 2003), leading to differentiation of subpopulations based on disease response (i.e., relative health and reproduction). Although disease severity and tree health were variable within our sample, it is likely that the trees in our survey were a subsample of the entire, pre-disease population. If trees were not all equally susceptible, then it is possible that the remaining trees have a higher genetic covariance than was present in the pre-disease population, and therefore show less genetic structure. Although microsatellite markers are neutral measures of diversity and do not confer adaptive advantages, they may be in linkage disequilibrium with genomic blocks affecting disease response. As functional genes related to disease resistance are identified, markers that are specific to desirable traits can be utilized. Until then, neutral markers can be used to estimate the amount of diversity (and therefore potential for beneficial genotypes) found within populations. More in-depth and controlled experiments will be necessary to determine the complex relationships between genetic traits and the range of disease tolerance that is seen among individual butternuts.

A small number of butternut-Japanese walnut hybrids were identified previously within the sampled watersheds using molecular markers (Parks et al. 2013), but most areas of GSMNP were unaffected by hybridization. Since hybrids are associated with fragmented, human-planted landscapes and even small forest buffers can be barriers to invasion (Hoban et al. 2009, 2012b), it appears that the continuous, forested landscape of the Park was a substantial barrier to entry of *J. ailantifolia* genes from outside the Park. Since hybrids occurred in only a small area in the western part of the Park, pollen and seeds from the few hybrids that remained appeared to have been overwhelmed by the butternut propagules available, especially with increasing distance from the nonnative individuals. Additionally, the absence of human settlement in GSMNP over the past 75 years has prevented additional plantings of nonnative species, and the scarcity of regeneration over the last 30 years may have prevented more recent invasions.

The Hazel Creek subpopulation, in which four of the five individuals were confirmed to be hybrids (Parks et al. 2013), was significantly different from only two other watersheds, based on pairwise F_{ST}, indicating the overall genotypes of the hybrids were not significantly different from most butternut within the Park (Supplemental Table 1). Since the hybrids are of unknown pedigree and may be the result of multiple backcrosses (especially since no Japanese walnut trees are documented in the Park), it was not unusual that they were similar genetically to nearby butternuts. Gugerli et al. (2007) found hybrids of oak species did not cluster distinctly from the parental species when assessed using multilocus microsatellite genotypes. The Hazel Creek subpopulation also had the highest private allelic richness, but overall, it was not differentiated from butternuts in other watersheds. Although the goal of this investigation was not to determine the genetic differentiation of the hybrid individuals, the presence of unique alleles in and from hybrids can be a metric of genetic invasion and could be useful in future investigations of genetic introgression. Hybridization with nonnative congeners can be a latent hazard for butternut populations, since hybrids can be quite similar morphologically to butternuts and difficult for managers to identify (Ross-Davis et al. 2008a; Zhao and Woeste 2010). The genetic consequences of hybridization are irreversible and can quickly overtake a rare plant population (Levin et al. 1996). Whenever feasible, a crucial measure in rare plant conservation is to maintain spatial separation of endangered populations from cross-compatible species, especially nonnatives (Levin et al. 1996). Because genetically diverse, nonhybrid butternut populations are increasingly rare (Hoban et al. 2010), efforts to prevent future hybridization, such as removal of Japanese walnut or hybrids, are crucial to protecting the "purity" of the GSMNP population.

Average kinship within watersheds showed variability, but the values were quite low in general (Fig. 3a). By the definition of kinship coefficient, full siblings (individuals sharing a both parents) are expected to have kinship coefficients of 0.25 (Hardy and Vekemans 2007). An unusually high value (0.7109) was recorded between two individuals that were within 5 m of one another and part of a tightly clustered group

of eight individuals that likely descended from the same parents. Since the average kinship within watersheds was near zero, we can assume that few of the trees were siblings or descended from closely related sources. Because there is a paucity of young trees in the GSMNP population, most of the remaining individuals have been subject to many years of selection and competition. Any pattern of spatial clustering of individuals from the same family tends to decrease with time since prolonged competition often leaves only a few of the siblings that may have established simultaneously (Epperson 1992).

The spatial arrangement of butternut trees varied widely from watershed to watershed. Some subpopulations consisted of dense clusters; others contained solitary individuals distributed linearly along streams. The average pairwise distance between trees within a watershed was negatively correlated with the average pairwise kinship within that watershed, indicating that average kinship was higher in watersheds with more densely clustered individuals. Although no pattern of isolation by distance was found by comparing genetic and spatial distances at the watershed scale, these results indicate that some degree of isolation by distance may be occurring at highly localized scales within watersheds, as has been observed in other butternut populations (Hoban et al. 2012a) and within other heavy-seeded species (Dutech et al. 2005). The variability of kinship among watersheds we observed may represent differences in dispersal and disturbance processes. Clusters of related trees probably reflect limited dispersal of seeds, since they are often dispersed by gravity or rodents with small home ranges (Hoban et al. 2012a). Squirrels disperse butternut via their caching behavior (Rink 1990), and seeds also can be dispersed by flowing water, but since light is often the limiting factor in butternut regeneration, the pattern of distribution we observed is likely due to the presence of suitable light conditions along streams and in natural disturbance gaps. For example, local floods or stream-bank shifts may open a large area that allows several unrelated individuals to establish. Whereas in other locations, tree fall gaps may be utilized by sporadic individuals, and propagules may come from a single nearby source. Humans contributed to the current distribution as well; the vast majority of butternut trees were found in areas previously occupied by small farms and settlements (Parks et al. 2013).

Significant genetic differentiation between the young and old cohort was not found. Genetic differentiation between age cohorts has been noted in other studies (Hoban et al. 2012a) and is attributed to the effects of selection on the older generation (Hamrick 1982). Nevertheless, the two age classes we defined for the GSMNP population remained genetically similar. Average gene diversity was slightly higher in the younger cohort but allelic richness was lower. Although there were two distinct peaks in recruitment, we observed few gaps in the distribution of tree ages, indicating continuous recruitment that resulted in indistinct cohort structure. Since historic records indicate disparity in the date of land abandonment for different areas of the Park, the peak establishment periods among watersheds was variable. Alternatively, the recent lack of regeneration may have resulted in the absence of a "true" younger cohort and all the sampled trees may represent a continuum of older cohorts across watersheds. Again, this result probably indicates that older trees did not harbor unique diversity that warrants special attention for conservation.

Management implications and future research

Despite the topographic, historical, and ecological variability within the Park, we observed relatively high levels of diversity, allelic richness, and a general absence of genetically distinct subpopulations. Mortality estimates for GSMNP over the past couple of decades are high and may not even include the initial wave of mortality prior to monitoring efforts (Parks et al. 2013). A dramatic lack of regeneration has been noted for this population (Parks et al. 2013), likely due to lack of suitable conditions and poor survival of young trees due to disease. Since the effects of gene flow restrictions can take several generations to be manifest in populations (Segelbacher et al. 2010), the threat from loss of adaptive diversity is secondary to the rapid declines caused by disease and lack of regeneration, and dramatic measures may be necessary to address these factors. The development of disease resistant breeding stock, combined with active management that mimics disturbance that allows butternut regeneration, are the most critical steps to ensure survival of the species. Some individuals in natural populations of butternut have survived many years of disease pressure when other nearby members of the population succumbed (Ostry and Woeste 2004; Michler et al. 2006; Parks et al. 2013). Whether these surviving individuals are expressing a type of resistance that can be used in breeding remains to be determined. Some hypotheses attribute the natural resistance found in some butternut populations to extensive hybridization with the Japanese walnut, which is reported to be more resistant to butternut canker than butternut (Michler et al. 2006; McCleary et al. 2009). However, if the goals of the breeding program are to maintain the purity of the butternut population and utilize natural resistance, then nonnative genes may not be desirable. Protection from additional hybridization, by removal of Japanese walnut and hybrids will ensure that future populations are protected from genetic invasion. Introgression can overwhelm rare plant populations rapidly and lead to extinction of the unique gene pool of the native species (Levin et al. 1996; Rhymer and Simberloff 1996).

Genetic structure and diversity analyses are useful for inferring population dynamics and connectivity, but assessment of functionally relevant diversity is also needed,

especially in rare species where variable fitness exists (Holderegger et al. 2010; Ouborg et al. 2010). Functional gene analysis will likely be necessary as progress is made toward exploitation of natural disease resistance. The diversity remaining in the GSMNP population makes it a desirable source of genotypes for breeding efforts and development of disease resistant breeding selections for restoration, since genetic diversity is a goal of conservation efforts in addition to disease resistance. Some have argued that environmental factors lead to species extinction before genetic problems (Ouborg et al. 2010). If this is the case, maintaining butternut on the landscape by promoting regeneration and selecting resistant genotypes for breeding programs is potentially the most vital goal of conservation. The effects of disease and poor regeneration will lead to population reductions much more rapidly than loss of genetic diversity.

Acknowledgments We would like to thank the employees and volunteers of Great Smoky Mountains National Park for providing invaluable support for this project, especially Tom Remaley, Janet Rock, Keith Langdon, Kristine Johnson, Glenn Taylor, Emily Darling, Emily Guss, and Paul Super for contributing their time and resources to facilitate our work in GSMNP. Field assistance was provided by H. Seiner, B. Greene, V. Meyer, B. Moehl, J. Albritton, M. Cooke, S. Simpson, J. Byrne, M. Duffy, M. Clebsch, C. Pan, M. Williams, K. Hansen, C. deGuia, L. Combs, C. Black, D. Love, D. Clabo, and K. Mann. William Beatty provided useful advice for statistical analyses. This work was supported by the Hardwood Tree Improvement and Regeneration Center (United States Department of Agriculture Forest Service Northern Research Station) and the Department of Forestry and Natural Resources at Purdue University. Mention of a trademark, proprietary product, or vendor does not constitute a guarantee or warranty of the product by the US Department of Agriculture and does not imply its approval to the exclusion of other products or vendors that also may be suitable.

Data archiving statement Microsatellite genotypes at all loci for all samples including hybrids, and the watershed from which the samples were obtained, are archived at TreeGenes as TGDR007. The specific locations of sampled trees are available for research purposes from Great Smoky Mountains National Park; contact Tom Remaley, Supervisory Ecologist/Inventory & Monitoring Coordinator, TOM_REMALEY@NPS.GOV.

References

Altizer S, Harvell D, Friedle E (2003) Rapid evolutionary dynamics and disease threats to biodiversity. Trends Ecol Evol 18:589–596

Crawford NG (2010) SMOGD: software for the measurement of genetic diversity. Mol Ecol Resour 10:556–557

Dutech C, Sork VL, Irwin AJ, Smouse PE, Davis FW (2005) Gene flow and fine-scale genetic structure in a wind-pollinated tree species, Quercus lobata (Fagaceaee). Am J Bot 92(2):252–261

Eckert AJ, Eckert ML, Hall BD (2010) Effects of historical demography and ecological context on spatial patterns of genetic diversity within foxtail pine (*Pinus balfouriana*; Pinaceae) stands located in the Klamath Mountains, California. Am J Bot 97:650–659

Ellstrand NC, Elam DR (1993) Population genetic consequences of small population size: implications for plant conservation. Annu Rev Ecol Syst 24:217–242

Epperson BK (1992) Spatial structure of genetic variation within populations of forest trees. New Forest 6:257–278

Epperson BK, Chung MG (2001) Spatial genetic structure of allozyme polymorphisms within populations of *Pinus strobus* (Pinaceae). Am J Bot 88:1006–1010

ESRI (Environmental Systems Resource Institute) (2009) ArcMap 9.3.1. ESRI, Redlands, California

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure: extensions to linked loci and correlated allele frequencies. Genetics 164:1567–1587

Goudet J (1995) FSTAT: a computer program to calculate *F* statistics, version 2.9.3. J Hered 86:485–486

Gugerli F, Walser JC, Dounavi K, Holderegger R, Finkeldey R (2007) Coincidence of small-scale spatial discontinuities in leaf morphology and nuclear microsatellite variation of *Quercus petraea* and *Q. robur* in a mixed forest. Ann Bot 99:713–722

Guillot G, Estoup A, Mortier F, Cosson JF (2005a) A spatial statistical model for landscape genetics. Genetics 170:1261–1280

Guillot G, Mortier F, Estoup A (2005b) Geneland: a computer package for landscape genetics. Mol Ecol Resour 5:712–715

Hamrick JL (1982) Plant population genetics and evolution. Am J Bot 69: 1685–1693

Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Mol Ecol Notes 2:618–620

Hardy OJ, Vekemans X (2007) SPAGeDi 1.2—a program for spatial pattern analysis of genetic diversity. User's manual. Available from http://www.ulb.ac.be/sciences/ecoevol/spagedi.html

Hoban SM, Anderson R, McCleary T, Schlarbaum S, Romero-Severson J (2008) Thirteen nuclear microsatellite loci for butternut (*Juglans cinerea* L.). Mol Ecol Resour 8:643–646

Hoban SM, McCleary TS, Schlarbaum SE, Romero-Severson J (2009) Geographically extensive hybridization between the forest trees American butternut and Japanese walnut. Biol Lett 5:324–327

Hoban SM, Borkowski DS, Brosi SL, McCleary TS, Thompson LM, McLachlan JS, Pereira MA, Schlarbaum SE, Romero-Severson J (2010) Range-wide distribution of genetic diversity in the North American tree *Juglans cinerea*: a product of range shifts, not ecological marginality or recent population decline. Mol Ecol 19:4876–4891

Hoban SM, Schlarbaum SE, Brosi SL, Romero-Severson J (2012a) A rare case of natural regeneration in butternut, a threatened forest tree, is parent and space limited. Conserv Genet 13:1447–1457

Hoban SM, McCleary TS, Schlarbaum SE, Anagnostakis SL, Romero-Severson J (2012b) Human-impacted landscapes facilitate hybridization between a native and an introduced tree. Evol Appl 5(7):720–731

Holderegger R, Wagner HH (2008) Landscape genetics. Bioscience 58: 199–207

Holderegger R, Buehler D, Gugerli F, Manel S (2010) Landscape genetics of plants. Trends Plant Sci 15:675–683

Jenkins MA (2007) Vegetation communities of the Great Smoky Mountains National Park. Southeast Nat 6(Special Issue 1):35–56

Jenkins MA, White PS (2002) *Cornus florida* L. mortality and understory composition changes in western Great Smoky Mountains National Park. J Torrey Bot Soc 129:194–206

Jones FA, Hamrick JL, Peterson CJ, Squiers ER (2006) Inferring colonization history from analyses of spatial genetic structure within populations of *Pinus strobus* and *Quercus rubra*. Mol Ecol 15:851–861

Jost L (2008) G_{ST} and its relatives do not measure differentiation. Mol Ecol 17:4015–4026

Kalinowski ST (2004) Counting alleles with rarefaction: private alleles and hierarchical sampling designs. Conserv Genet 5:539–543

Kalinowski ST (2005) HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic diversity. Mol Ecol Notes 5:187–189

Kalisz S, Nason JD, Hanzawa FM, Tonsor SJ (2001) Spatial population genetic structure of *Trillium grandiflorum*: the roles of dispersal, mating, history and selection. Evolution 55:1560–1568

Kitamoto N, Honjo M, Ueno S, Takenaka A, Tsumura Y, Washitani I, Ohsawa R (2005) Spatial genetic structure among and within populations of *Primula sieboldii* growing beside separate streams. Mol Ecol 14:149–157

Knowles P, Perry DJ, Foster H (1992) Spatial genetic structure in two tamarack [*Larix laricina* (Du ROI) K. Koch] populations with differing establishment histories. Evolution 46:572–576

Levin DA, Franscisco-Oretega J, Jansen RK (1996) Hybridization and extinction of rare plant species. Conserv Biol 10:10–16

Loiselle BA, Sork VL, Nason J, Graham C (1995) Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). Am J Bot 82:1420–1425

Loveless MD, Hamrick JL (1984) Ecological determinants of genetic structure in plant populations. Annu Rev Ecol Syst 15:65–95

Mccleary TS, Robichaud RL, Nuanes S, Anagnostakis SL, Schlarbaum SE, Romero-Severson J (2009) Four cleaved amplified polymorphic sequence (CAPS) markers for the detection of the *Juglans ailantifolia* chloroplast in putatively native *J. cinerea* populations. Mol Ecol Resour 9:525–527

Michler CH, Pijut PM, Jacobs DF, Meilan R, Woeste KE, Ostry ME (2006) Improving disease resistance of butternut (*Juglans cinerea*), a threatened fine hardwood: a case for single-tree selection through genetic improvement and deployment. Tree Physiol 26:121–128

Nei M (1978) Estimation of average heterozygosity and genetic distance for a small number of individuals. Genetics 89:583–590

Ostry ME, Woeste K (2004) Spread of butternut canker in North America, host range, evidence of resistance within butternut populations and conservation genetics. In: Michler CH, Pijut PM, van Sambeek J, Coggeshall M, Seifert J, Woeste K, Overton R (eds) Black walnut in a new century: Proc. 6th Walnut Council Research Symposium.. USDA For Serv, North Central Res Stn, Gen Tech Rep NC-243, St. Paul, MN, pp 114–120

Ouborg NJ, Francesco A, Vergeer P (2010) An essay on the necessity and feasibility of conservation genomics. Conserv Genet 11:643–653

Parker CR, Pipes DW (1990) Watersheds of Great Smoky Mountains National Park: a geographical information system analysis. Uplands Field Research Laboratory, Great Smoky Mountains National Park

Parks AM, Jenkins MA, Woeste KE, Ostry ME (2013) Conservation status of a threatened tree species: establishing a baseline for restoration of *Juglans cinerea* L. in the southern Appalachian Mountains. Natural Areas Journal (in press)

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics 28:2537–2539

Petit RJ, Abdelhamid EM, Pons O (1998) Identifying populations for conservation on the basis of genetic markers. Conserv Biol 12:844–855

Pons O, Petit RJ (1995) Estimation, variance and optimal sampling of gene diversity. Theor Appl Genet 90:462–470

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Pyle C (1985) Vegetation disturbance history of Great Smoky Mountains National Park: an analysis of archival maps and records. USDI, NPS, Res/Resour Manage Rep SER-77, Southeast Regional Office, Atlanta, Georgia, 69 p

Pyle C (1988) The type and extent of anthropogenic vegetation disturbance in the great Smoky Mountains before National Park Service acquisition. Castanea 53:183–196

Rhymer JM, Simberloff D (1996) Extinction by hybridization and introgression. Annu Rev Ecol Syst 27:83–109

Rink G (1990) *Juglans cinerea* L., butternut. In: Burns RM, Honkala BH (eds) Silvics of North America. Vol. 2. Hardwoods. Tech. Coords. . USDA Forest Service Agricultural Handbook 654, Washington, DC, pp 386–390

Rosenberg MS, Anderson CD (2011) PASSaGE: pattern analysis, spatial statistics and geographic exegesis. Version 2. Methods Ecol Evol 2: 229–232

Ross-Davis A, Woeste K (2008) Microsatellite markers for *Juglans cinerea* L. and their utility in other Juglandaceae species. Conserv Genet 9:465–469

Ross-Davis A, Huang Z, McKenna J, Ostry M, Woeste K (2008a) Morphological and molecular methods to identify butternut (*Juglans cinerea*) and butternut hybrids: relevance to butternut conservation. Tree Physiol 28:1127–1133

Ross-Davis A, Ostry M, Woeste K (2008b) Genetic diversity of butternut (*Juglans cinerea*) and implications for conservation. Can J For Res 38:899–907

Schultz J (2003) Conservation assessment for butternut or white walnut (*Juglans cinerea* L.). USDA For Serv, Eastern Region. Milwaukee, WI, 76 p

Segelbacher G, Cushman SA, Epperson BK, Fortin M, Francois O, Hardy OJ, Holderegger R, Taberlet P, Waits LP, Manel S (2010) Applications of landscape genetics in conservation biology: concepts and challenges. Conserv Genet 11:375–385

Sork VL, Smouse PE (2006) Genetic analysis of landscape connectivity in tree populations. Landscape Ecol 21:821–836

Sork VL, Waits L (2010) Contributions of landscape genetics—approaches, insights, and future potential. Mol Ecol 19:3489–3495

Storfer A, Murphy MA, Spear SF, Holderegger R, Waits LP (2010) Landscape genetics: where are we now? Mol Ecol 19:3496–3514

van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes 4:535–538

Vekemans X, Hardy OJ (2004) New insights from fine-scale spatial genetic structure analyses in plant populations. Mol Ecol 13:921–935

Victory ER, Glaubitz JC, Rhodes OE Jr, Woeste K (2006) Genetic homogeneity in *Juglans nigra* (Juglandaceae) at nuclear microsatellites. Am J Bot 93(1):118–126

Victory ER, Glaubitz JC, Fike JA, Rhodes OE Jr., Woeste K (2008) Different histories but similar genetic diversity and structure for black walnut in Indiana and Missouri. In: Jacobs DF, Michler CH. Proceedings, 16th central hardwood forest conference. Gen. Tech. Rep. NRS-P-24. Newtown Square, PA: USDA For Serv, Northern Res Stn, pp 436-445

Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. Evolution 38:1358–1370

Woeste K, Farlee L, Ostry M, McKenna J, Weeks S (2009) A forest manager's guide to butternut. North J Appl For 26:9–14

Zhao P, Woeste KE (2010) DNA markers identify hybrids between butternut (*Juglans cinerea* L.) and Japanese walnut (*Juglans ailantifolia* Carr.). Tree Genet Genomes 7:511–533