Tree Genetics & Genomes (2016) 12:43
DOI 10.1007/s11295-016-1002-2

ORIGINAL ARTICLE

Population genetic diversity of the rare hardwood butternut (*Juglans cinerea*) in the northeastern USA

A. Boraks[1] · **K. D. Broders**[1,2]

Received: 1 July 2015 / Revised: 1 March 2016 / Accepted: 4 April 2016

Abstract Populations of butternut tree (*Juglans cinerea*) have undergone range-wide extirpation. A fungal pathogen, *Ophiognomonia clavigignenti-juglandacearum*, of unknown origin has been recognized as the causal factor. This population collapse has allowed for observations of a broadleaf hardwood in rapid decline. This study made use of six neutral microsatellite markers to describe the present genetic diversity of butternut in the northeastern USA. Our results indicated weak population differentiation ($F_{ST} = 0.084$), further supported by an absence of regional genetic structure. Despite reports of high mortality rates, genetic analysis revealed no sign of a recent bottleneck. Population statistics and Bayesian analysis indicated significant historical gene flow among butternut populations of the northeast. Attention should be given to genetic differences between upland and riparian habitat as riparian populations appear to contain greater allele diversity.

Keywords *Juglans cinerea* · Butternut · Genetic · Riparian · Microsatellite

Introduction

The butternut tree (*Juglans cinerea*), also known as white walnut, is native to eastern North America. Butternut plays a minor role in the ecology of North America (Schultz 2003) and is generally isolated to clusters of trees. Butternut is a shade intolerant, pioneer species that is often found in riparian habitats. Occasionally, butternuts are also found in well drained upland habitats among species of *Acer*, *Quercus*, *Betula*, and *Tilia* (Rink 1990). Butternut has a historical and cultural context as a food source and fabric dye (Ostry and Pijut 2000). Ecologically, butternut is most noteworthy as a cold-hardy, mast producing hardwood.

Butternut is currently under threat of extinction due to an introduced fungal pathogen (Broders et al. 2015). First reported as a pathogen on butternut in 1967 (Nair et al. 1979), *Ophiognomonia clavigignenti-juglandacearum* (*Oc-j*) (Broders and Boland 2011) previously known as *Sirococcus clavigignenti-juglandacearum* Nair, Kostichka, and Kuntz, now threatens butternut throughout the species' range (Ostry and Woeste 2004). *Oc-j* is a clonally propagating ascomycete, likely of Asian origin where it may exist as an endophyte on native species (Broders et al. 2015). The high rate of butternut mortality (Cummings-Carlson and Guthmiller 1993; Parks et al. 2013) provided opportunity to observe population decline and its effects on genetic structure.

Mature forest stands and disease are not the only pressures on butternut populations. *Juglans ailantifolia* (Carr.) was introduced to North America by horticulturalists in 1860 (Crane et al. 1937). A congener of butternut, *J. ailantifolia* can freely hybridize with *J. cinerea*. Hybrids are known as *J. bixbyi* Rehd. and pose a threat of genetic invasion. Historically, genetic invasion has not been considered a significant issue for two reasons. First, *J. ailantifolia* remained proximal to human establishments. Second, populations of *J. cinerea* greatly outnumbered those of *J. ailantifolia*. Butternut had a domestic advantage in both number and range distribution. Risk of genetic invasion has become more of a recent concern following investigations into the magnitude of hybridization in North

Communicated by A. Kremer

✉ K. D. Broders
kirk.broders@colostate.edu

[1] Department of Biological Sciences, University of New Hampshire, Durham, NH 03824, USA

[2] Department of Bioagricultural Sciences and Pest Management, Colorado State University, Ft. Collins, CO 80521, USA

Published online: 15 April 2016

America (Hoban et al. 2009). While the potential for infection is a documented possibility, *J. ailantifolia* is typically at low risk of mortality caused by *Oc-j* (Orchard et al. 1982; personal observation). The same can be said for *J. bixbyi*. Butternut hybrids show remarkable vigor in the face of disease pressure from *Oc-j* (McKenna et al. 2013). This apparent hybrid vigor may be the tipping point that allows genetic invasion to become a concern for butternut conservation. Comprehensive information on the ecological status of butternut has been published by Schultz (2003) and Michler et al. (2006).

Molecular markers are routinely used for the genetic analysis of hardwood tree populations (Yao et al. 2011; Jump and Peñuelas 2007). The variability of SSRs makes microsatellites an ideal candidate for the detection of regional genetic variability. A set of microsatellites were previously developed for a variety of applications in *Juglans* (Ross-Davis and Woeste 2007). The present study uses nuclear DNA microsatellite markers to determine relationships of butternut populations in the northeastern USA.

Regional genetic studies among butternut's northeastern range are limited. Past studies of northeastern populations concluded that there are low levels of genetic structure (Ross-Davis et al. 2008; Morin et al. 2000; Hoban et al. 2010). Recent studies from butternut's southern range indicate high levels of genetic diversity (Parks et al. 2014). Previous studies have not included butternut populations from New York, which represented a significant gap in butternut research. The objectives of this study were to determine the genetic structure of butternut populations in the northeastern USA and evaluate the correspondence of genetic data to several measures of tree vigor. Specifically, we address genetic divergence among sample locations and genetic difference between growth habitats and elucidate spatial genetic structure.

While butternut may only play a minor role in the ecology of eastern North American forests, the effect of an exotic disease on forest trees is important, given the influx of invasive pathogens into North America. This study presents a snapshot of the regional genetic structure as a tool to guide the recovery strategy for butternut (Poisson and Ursic 2013) and potentially other native hardwood species at risk.

Materials and methods

Sampling

The northeastern USA represents the northern distal portion of the butternut's range and includes a vast landscape of favorable growth locations. Leaf or bark tissue samples of *J. cinerea* ($n = 258$) were collected from 16 locations in Maine, New Hampshire, Vermont, and New York. Each sample cluster was defined as a geographically separated aggregate of butternut trees (Fig. 1). Our field collection was based on a convenience of sample method whereby we relied on private and public landowners for information to locate trees. Efforts were made to evenly distribute locations in order to minimize statistical error associated with sample clumping. Sites were not exhaustively sampled. A minimum of 10 trees per location was considered as sufficiently large enough to sample. Both live and dead trees were present at sampled locations. In three instances (sites 8, 14, 15) fewer than 10 live butternut trees remained, these instances resulted in reduced sample numbers. Dead trees were not sampled. Typically, trees within a sample location grew within an area of 1 km^2. Pairwise Euclidean distances between sample locations averaged 343 km (Std. Dev. 228), and locations varied in growth habitat from dynamic flood basins to more static upland forest. Trees established within 100 m of a water basin were given the designation *riparian*; all other trees were considered *upland*. Habitat classification, diameter at breast height (DBH), bark phenotype, vigor, and crown class were recorded for each individual. Vigor, bark type, and crown class were visually assessed and recorded on an ordinal scale.

Hybrid analysis

To ensure sampled trees were butternuts, a hybrid diagnostic test developed by McCleary et al. (2009) was employed. In summary, cleaved amplified polymorphic sequences (CAPS markers) were used to detect hybrids. The results of this assay revealed true *J. cinerea* chloroplast DNA as a single amplicon of 332 bp. In comparison, trees that contained *J. alantifolia* chloroplast DNA (hybrids) had cleaved amplicons of 235 and 97 bp in lengths. In practice, this hybrid test detects roughly 90 % of butternut hybrids (Hoban et al 2012). Samples that contained *J. alantifolia* DNA and samples that could not be resolved in this hybrid test were removed from the study.

Microsatellite selection and amplification

Eight sets of dinucleotide primers, previously developed for *J. cinerea* (Ross-Davis and Woeste 2007), were used to genotype all sampled individuals (Table 1). Two of these loci (WGA148 and WGA221) showed evidence of null alleles or had insufficient amplification for a majority of the samples and were not included in the study. Microsatellite amplifications were performed as 15-uL reactions in 96-well PCR plates and contained positive and negative controls. PCR products were visualized in stained agarose gel to verify amplification.

Forward primers were synthesized with either HEX or FAM fluorescent labels to allow for pool-plexing (pooled as follows: WGA147HEX; 221HEX; 204FAM; 256FAM and WGA004HEX; 082HEX; 090FAM; 148FAM). Verified amplicons were pool-plexed, diluted at a ratio of 1:10, and

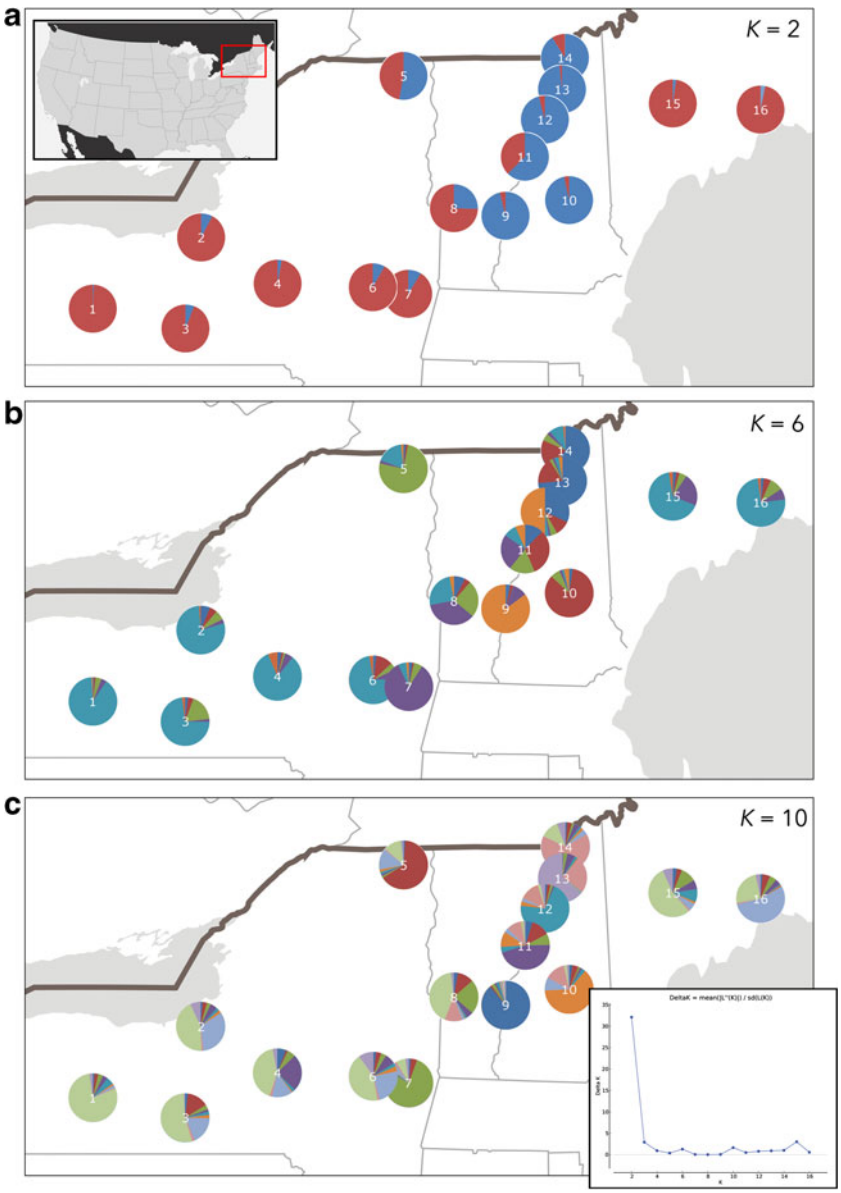

Fig. 1 **a**, **b**, **c** Sample locations overlaid with Bayesian probability of population assignment ($K = 2$, 6, 10). *Bottom-right inset* shows ΔK as calculated by STRUCTURE HARVESTER

Table 1 Descriptive statistics for microsatellites

Locus	N_A[a]	Size range (bp)	F_{IT}[b]	F_{ST}[c]	F_{IS}[d]
WGA90	8	126–142	−0.002	0.069	−0.076
WGA4	24	225–273	0.169	0.073	0.104
WGA82	15	153–181	0.178	0.075	0.113
WGA256	18	206–242	0.199	0.071	0.137
WGA204	13	172–196	0.086	0.098	−0.013
WGA147	9	174–200	0.358	0.068	0.311
Mean	14.5		0.168	0.077	0.098

[a] Number of alleles within a locus

[b] Variance of alleles relative to total

[c] Variance among loci relative to total

[d] Variance within locus

genotyped on an ABI 3130 Genetic Analyzer. Raw microsatellite data was reviewed and sorted with GeneMapper v. 4.0 software (Applied Biosystems Inc., Foster City, CA).

Data analysis

Samples with more than 15 % of their allelic data unresolved or samples identified as hybrid trees were omitted from the data set. The final data set contained 206 individuals in 16 populations (Table 2). Many of the analyses discussed below allow for incomplete data sets. However, missing data points can be particularly problematic for pairwise distance-based analyses such as AMOVA and Mantel tests. Rarefaction was not used to standardize sample sizes. Several programs can interpolate missing microsatellite data by inserting average genetic distances for each population level pairwise contrast. Where

Table 2 Sampling locations and genetic characteristics

Site	Habitat	N	Latitude	Longitude	H_O	H_E	uH_E	F_{ST}	F_{IS}	Het	R_A	N_A	N_{PA}	P_{PA}
1	R/U	13	42.515	−77.890	0.649	0.740	0.779	0.073	0.157	–	6.167	42	3	0.071
2	U	16	43.222	−76.608	0.570	0.718	0.750	0.062	0.253	Deficient	6.500	45	–	–
3	U	11	42.534	−76.696	0.736	0.706	0.757	0.061	0.042	–	6.00	40	1	0.025
4	U	11	42.903	−75.633	0.690	0.756	0.804	0.076	0.130	–	6.167	43	2	0.047
5	U	13	44.923	−73.779	0.619	0.608	0.647	0.121	0.081	–	4.667	32	–	–
6	U	11	42.719	−74.110	0.593	0.681	0.725	0.064	0.183	Deficient	5.500	38	–	–
7	R	10	42.67	−73.650	0.583	0.585	0.645	0.130	0.201	–	3.833	26	1	0.038
8	U	2	43.506	−72.929	0.750	0.500	0.667	0.128	0.542	–	[a]	17	–	–
9	U	14	43.519	−72.296	0.764	0.691	0.735	0.121	0.183	–	5.333	40	3	0.075
10	R/U	20	43.657	−71.503	0.771	0.739	0.761	0.077	0.062	–	6.833	49	2	0.041
11	R	22	44.046	−72.064	0.595	0.691	0.731	0.078	0.211	Deficient	6.833	53	–	–
12	R	25	44.475	−71.622	0.672	0.687	0.717	0.093	0.091	Deficient	6.833	49	3	0.061
13	R	20	44.654	−71.565	0.604	0.744	0.769	0.091	0.239	Deficient	6.833	48	1	0.021
14	R	6	44.756	−71.620	0.608	0.608	0.678	0.150	0.138	–	3.833	29	–	–
15	U	5	44.510	−70.519	0.850	0.612	0.705	0.105	–0.299	Excess	3.333	22	–	–
16	U	7	44.424	−69.006	0.667	0.659	0.723	0.092	0.098	–	4.667	31	1	0.032
	Total	206		Mean	0.659	0.691	0.733	0.092	0.158		5.94	42	1.29	0.03
				±Std. Err.				0.012	0.019					

Sample size (N) and habitat are labeled as riparian (R) or upland (U), and locations are in degrees latitude north and longitude west. Sample size (N), observed heterozygosity (H_O), expected heterozygosity (H_E), unbiased expected heterozygosity (uH_E), and allele richness (R_A) are averaged across all loci. N_A indicates number of alleles, N_{PA} the number of private alleles, and P_{PA} the proportion of alleles that were private for a given site. Significance of heterozygosity (Het) deficiency or excess is indicated at $P < 0.05$. Total values are weighted means according to sample size. Variance among subpopulation relative to the total (F_{ST}) and inbreeding coefficient (F_{IS}) are averages

[a] Population is of insufficient size to determine R_A

interpolation could not be applied, sample units were removed from the dataset.

Populations were tested for conformity to Hardy-Weinberg equilibrium, heterozygote deficiency and excess, and significance were estimated by Markov chain Monte Carlo method with 1000 randomizations in GENEPOP v4.2 (Rousset 2008). Tests for possible null alleles were performed with MICROCHECKER v2.2.3 (Van Oosterhout et al. 2004) with 1000 Monte Carlo simulations and a confidence interval of 95 % adjusted by Bonferroni correction. Private alleles (those that occurred in only a single location), alleles per loci (N_A), and allelic richness (R_A) were determined with GenALEX v6.5 (Peakall and Smouse 2006) (Table 2).

Genetic structure was measured by unbiased $_{ST}$ measurements of Wright's F-statistics (expected heterozygosity, observed heterozygosity, F_{ST}, F_{IS}) (Wright 1931) and Shannon's diversity index (Shannon and Weaver 1949). Associated P values were calculated in FSTAT v2.9.3 (Goudet 2001) jackknifing over all loci. Values that differed significantly from zero are used to reject the null hypothesis of panmixia (Balloux and Lugon-Moulin 2002). Genetic bottleneck was tested with Bottleneck V.1.2.02 (Cornuet and Luikart 1997). Samples were pooled into a single regional population and estimated over 1000 replications.

Spatial genetic structure of the sampled populations was evaluated with STRUCTURE v2.3.3 (Pritchard et al. 2000; Pritchard et al. 2003). Genetic clusters were tested by specifying the number of potential populations (K), ranging from the null hypothesis of panmixia ($K=1$) through to the maximum number of sample locations ($K=16$). Twenty iterations of each K were performed in STRUCTURE. This process was run twice with variable burn-in values of 100,000 and 300,000 and a Markov chain Monte Carlo (MCMC) set to 100,000. Optimal K-value was determined by the delta-K likelihood evaluations from Evanno et al. (2005) and implemented in the program STRUCTURE HARVESTER (Earl 2012).

GenAlEx was used to calculate Nei's unbiased genetic distance (Nei 1978) to perform a principal coordinate analysis (PCoA). The procedure in GenAlEx is based on algorithms published by Orlóci (1978). Results from STRUCTURE ($K=2$) were overlaid to test the hypotheses of relative panmixia.

Isolation by distance was analyzed by a comparison of geographic and genetic distance. A paired Mantel test (Smouse and Long 1992; Smouse et al. 1986) was implemented in GenAlEx to correspond pairwise orthodromic distance with pairwise R_{ST} and $_{ST}$ (Weir and Cockerham 1984). Mantel tests of 999 permutations contrasted R_{ST} and $_{ST}$ to geographic distance and a third matrix of randomized data to

test significance. Small populations may lead to statistical error. To adjust for this, populations with fewer than ten samples were removed and the Mantel test was repeated.

Additional multivariate analyses were used to test for a relationship between genetic structure and habitat. An unconstrained approach employed the use of an F_{ST} distance matrix using the Bray-Curtis coefficient to conduct a non-metric multidimensional scaling (NMDS) (Fig. 2). A Sorensen distance matrix was run through 500 iterations in PC-ORD. Rather than clustering data on similarity, NMDS ranks data based on dissimilarity. Habitat was used as the overlay for the NMDS. A similar approach was used for diameter at breast height (DBH), bark phenotype, vigor, and crown class.

Results

Eight loci were chosen for the analysis of 258 butternut trees from 16 sites across the northeastern range of butternut. Two of these loci (WGA148 and WGA221) showed evidence of null alleles or had insufficient amplification for a majority of the samples. These loci, along with individuals of insufficient data coverage ($N=21$), were removed from the study and were not included in further analyses. Of the remaining 237 samples, 31 trees (13 %) were verified to contain *J. ailantifolia* ctDNA and these individuals were removed from the dataset. Two locations (site 8 and 16) contained hybrid butternut. The study was therefore based on 206 individuals at six loci from 16 locations (Table 2).

The number of alleles per locus ranged from 8 (WGA90) to 24 (WGA4) with the greatest number of alleles found at locations with large butternut populations (Table 2). Locations with the lowest allele richness occurred in Maine or at sites where the total number of trees sampled was low. Heterozygosity (H_O) was highest towards the northeastern portion of butternut's range at site 15 ($H_O=0.850$) and lowest towards the south-southwest portion of our study in New York (site 2 = 0.570, site 7 = 0.583, site 6 = 0.593) (Table 2). Half of the locations displayed private alleles. The largest number of private alleles (N_{PA}) occurred in riparian sites 1, 9, and 12. The degree of relatedness within a population (F_{IS}) was relatively low with a global mean of 0.144, as was the relatedness among locations ($F_{ST}=0.084$). R_{ST} and F_{ST} pairwise matrices were not significantly different as indicated by a paired Mantel test ($P=0.016$; $R^2=0.103$). As a consequence, F_{ST} values were used to facilitate future comparison of our results with other butternut studies.

Genetic differences among populations appeared to be independent of spatial isolation or geographic distance as indicated by a Mantel regression ($P=0.230$, $R^2=0.0156$) of genetic distance ($_{ST}/1-{}_{ST}$) and geographical distance (km) (Fig. 3). After the removal of small populations (<10

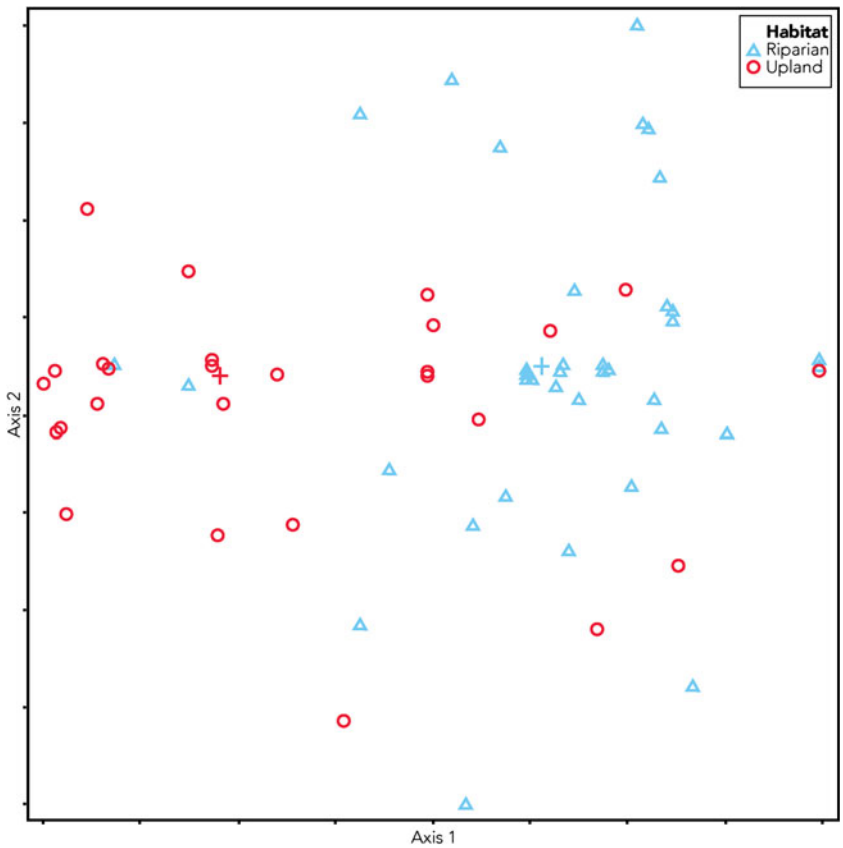

Fig. 2 NMDS ordination based on genetic pairwise distance matrix, codeded by habitat. The (+) represent the centroid for each respective variable

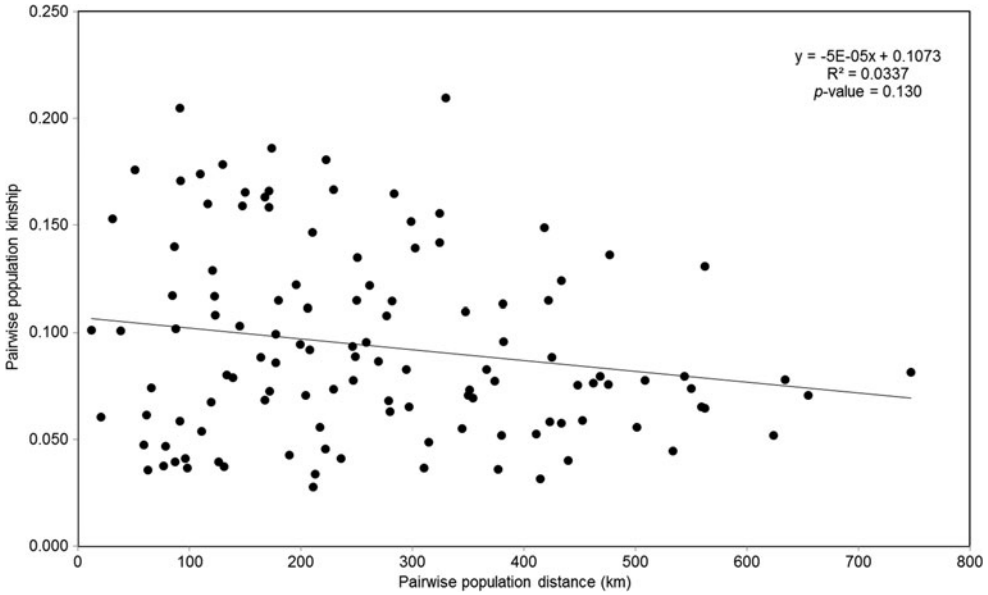

Fig. 3 Isolation by distance as tested by Mantel regression comparing matrices of pairwise population geographic distance and pairwise linearized $_{ST}$ distance (*y-axis*). No correlation of geographic and genetic distance detected ($P = 0.130$)

samples) from the dataset, IBD remained non-significant ($P = 0.250$, $R^2 = 0.0056$). The Bayesian analysis in STRUCTURE identified $K = 2$ population clusters justified by the likelihood method (Evanno et al. 2005). To a lesser degree, the likelihood method also identified $K = 6$ and $K = 10$ potential population clusters. The population differences identified by STRUCTURE ($K = 2, 6$) are noteworthy as those found within and apart from the Connecticut River valley (Fig. 1). A principal coordinate analysis (PCoA) revealed patterns of genetic structure that supported evidence found in the Bayesian analyses (Fig. 4). Trees sampled around the Connecticut River valley and sites designated as riparian in habitat appeared to cluster separately from those deemed upland habitats. Both upland and riparian trees contained similar levels of observed heterozygosity (H_O). However, riparian sample sites significantly differed from upland sites in allelic richness (R_A) and total number of alleles (N_A) (Table 3).

The stability of the NMDS ordination was assessed by the relative stress within a scree plot. A 2-D solution provided sufficiently low stress in accordance with Clark's rule of thumb (Clarke and Ainsworth, 1993). The difference in stress between a 2-D (7.671) and a 3-D (4.196) solution did not warrant further investigation into a more complex ordination. Genetic clustering was associated with habitat in the NMDS ordination (Fig. 2). NMDS ordination revealed no clear

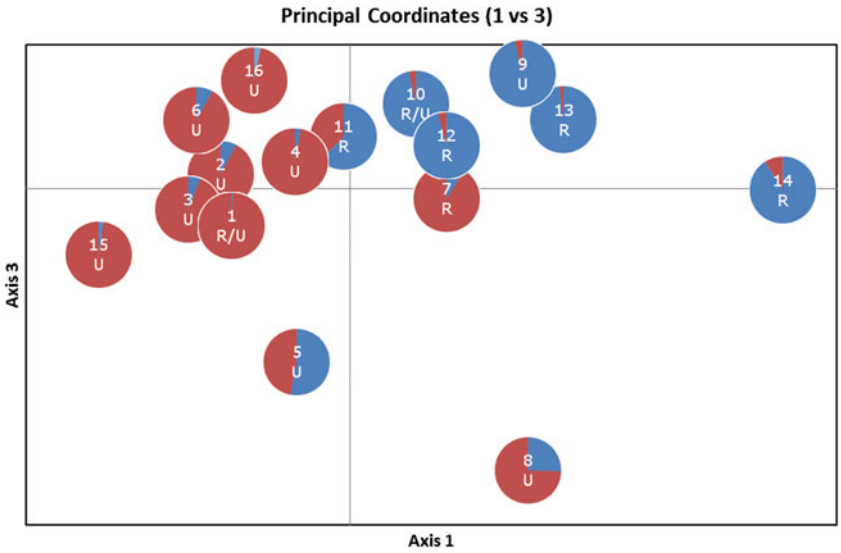

Fig. 4 Principal coordinate analysis of sample locations using alleles as distance measurements. Population assignment for $K = 2$ calculated by STRUCTURE is overlaid for comparison of Bayesian and FST statistics. Riparian sample sites are marked with "R" and upland sampling sites are marked with "U". Sample sites that contained trees in both upland and riparian habitat are labeled "R/U"

Table 3 Comparison of genetic diversity between upland and riparian butternut

Habitat	H_O	R_A	N_A
Riparian	0.649	*6.34*	*45.7*
Upland	0.685	*5.72*	*39.9*

Weighted mean heterozygosity (H_O), allele richness (R_A), and total allele number (N_A) for upland and riparian butternut trees. Italics indicate a significant difference ($P < 0.05$) between riparian and upland sites

pattern of association between crown class, DBH, or canker number and pairwise F_{ST}.

Discussion

Past studies have found relatively low genetic differentiation in butternut populations (Hoban et al. 2010; Morin et al. 2000; Parks et al. 2014; Ross-Davis et al. 2008), which agrees with the results of this study. In the present study, all sampled sites had similar observed heterozygosity (range = 0.57–0.85, mean = 0.659), lower than previously reported heterozygosity values of similar locations (*pop. 17–20* range = 0.789–0.842; Hoban et al. 2010) (*Ontario pop. Ho* = 0.83; Ross-Davis et al. 2008). Differences can be attributed to a number of factors although the genetic markers used (Jump and Peñuelas 2007), variation in sample size, and a lack of experimental randomization are likely candidates. Additionally, it should be noted that the CAPS method used to detect hybridity is roughly 90 % accurate. This method identifies only hybrids of *J. ailantifolia* maternal lineage (Hoban et al. 2009). It is quite likely that parts of our analyses contain a small representation of *J. ailantifolia* loci.

The sampling scheme and moderate sample size in this present study ($n = 206$) allows for decent estimation of population differentiation for butternut in the northeast. Experimental randomization did not occur during genotyping, which may have skewed estimations of relatedness. Steps taken ameliorated skewed F_{ST} included the genotyping of multiple populations and multiple loci in a single PCR run and standardization of genotyping results using a positive control. Results from our unbiased expected heterozygosity (Table 2) are similar to the expected heterozygosity calculated for butternut in Great Smoky Mountain National Park (GSMNP) (Parks et al. 2014). The H_E in GSMNP and this study may be similar due to the fact that both studies sampled from butternut's range periphery.

Despite the differences in observed heterozygosity, data presented here supports reports of reduced heterozygosity along the margin of butternut's range. Butternut trees in the northeastern peripheral are less diverse than those found more south in the GSMNP based on H_O (Parks et al. 2014). Our methodology does not allow us to speculate whether this supports the post Pleistocene-era range shift hypothesis presented by Hoban et al. (2010) or is an example of the abundant center hypothesis. Five locations were significantly deficient in heterozygotes. Three of these five locations were along the Connecticut River valley. It is possible that site 11, site 12, and site 13 were a single outcrossed population. Under this scenario, three locations (Connecticut River valley, site 2, and site 6) remain deficient in heterozygotes.

Heterozygosity is a poor gauge of whether butternut has experienced a recent bottleneck. It is of importance to conservation scientists as a measure by which a population can recover from a bottleneck event (Allendorf 1986). For butternut, this is of particular importance in light of its recent severe population decline. A genetic bottleneck was not apparent in this dataset. Sign, Standardized Differences, and Wilcoxon tests revealed no evidence of a bottleneck. Additionally, NMDS overlaid with DBH revealed no apparent pattern. The lack of evidence to support a genetic bottleneck may be the result of a lag between occurrence and appearance. Genetic diversity may influence the resilience of populations faced with pathogenic epidemics (Altizer et al. 2003). Estimates of heterozygosity presented here, along with those previously reported, may be used in the future by conservation scientists as a baseline of genetic pressure caused by disease, isolation, and habitat loss.

The deficiency in butternut heterozygosity cannot be explained by population sub-structure. Under Hardy-Weinberg assumptions, a population divided into isolated subpopulations has less heterozygosity than had the population been undivided. This is not the case with butternut as demonstrated by a lack of isolation by distance (Fig. 3). Some sites that were geographically close were distant genetically (e.g., site 13 and site 14), and some distant sites appeared genetically similar (e.g., site 16 and site 3). This result was verified by a number of analytical approaches and can be potentially explained by the dispersal method of butternut (Sork and Smouse 2006). Butternut is a wind pollinated species. Other studies of wind pollinated forest trees have demonstrated the ability of pollen to travel distances over 19 km (Ward et al. 2005). The maximum range of viable butternut pollen has not yet been reported although Pollegioni et al. (2009) report common walnut pollen traveling distances of a mile. Long distance dispersal has likely influenced our ability to detect regional genetic structure. This is consistent with previous studies of butternut (Hoban et al. 2010; Parks et al. 2014) and *J. nigra* (Victory 2006) each of which demonstrates weak isolation by distance. To date, butternut had demonstrated sufficient pollen movement for the maintenance of gene flow across the northeast. Again, this study represents only a snapshot of the population genetics for established trees. With

widespread mortality and the potential for population fragmentation, it should be expected that the gene flow of successive butternut generations will be affected.

The Bayesian analysis performed by STRUCTURE ($K=2$ and $K=6$) indicated that trees within the Connecticut valley were genetically distinct from trees sampled in New York and Maine (Fig. 1). Habitat is a potential biological reason for this demarcation. While the majority of sites sampled in the Connecticut valley were riparian, samples from New York and Maine were generally from upland sites. The STRUCTURE analyses $K=10$ made biological interpretation difficult as populations displayed low probability of any particular assignment. Upon PCoA ordination, riparian sites in New York clustered more closely with those of the Connecticut River valley than with geographically closer upland sites in New York (Fig. 4). Further inspection, as delineated by the NMDS (Fig. 2), demonstrated that trees with a riparian classification clustered separately from those designated as upland. Finally, when levels of genetic diversity are addressed, riparian trees are slightly more diverse than upland trees (Table 3).

Factors that influence greater diversity among riparian populations likely relate to the levels of fragmentation and recruitment frequency. Riparian sites were generally found in a habitat associated with intermediate disturbance. These habitats were generally quite large and present the possibility of supporting large dynamic communities. In contrast, butternuts sampled among upland sites were generally established on reforested agricultural land or along borders of agricultural land. Most forests in the northeastern USA were logged, burned, and farmed prior to 1900 (Williams 1992). For this reason, many forested sites represent a snapshot of forest conditions from the last 100 years. Current hardwood forest management in the northeast favors disturbance suppression, a factor that limits the amount of natural upland butternut regeneration due to uninterrupted shade and disease pressure (Boraks and Broders 2014). Longer-lived tree species suppress the recruitment of butternut, and without disturbance, butternut has ebbed from the forest. The elevated levels of diversity seen in riparian habitats potentially resulted from intermediate habitat disturbance, as compared to more static upland sites.

In addition to disturbance frequency, gene flow may be more common in riparian zones. Butternut seed is heavy and requires a vector like water (hydrochory) for effective dispersal. Additionally, the complex canopy associated with disturbed riparian sites may facilitate open pollen movement. Hu et al. (2010) report a similar genetic differentiation between riparian and upland sites for the endangered *Fraxinus mandshurica*. As with *F. mandshurica*, butternut demonstrates a genetic demarcation between upland and riparian habitat despite low genetic structure.

The heterozygote deficiency observed in butternut of the northeast is the product of several factors. A combination of range periphery, poor regeneration, and mortality are likely important factors contributing to small populations and heterozygote deficiency (Hoban et al. 2012). The long-term recovery goal outlined in the recovery strategy for butternut in Ontario aims to achieve broadly distributed populations within its current range in Canada (Poisson and Ursic 2013). Evidence presented here demonstrates weak isolation between butternut clusters, indicating minimal population sub-structure. Based on current microsatellite data, current butternut of the northeastern USA may be considered as a single broadly distributed population. This scenario may quickly change based on the rate of mortality and potential population fragmentation.

Data presented here helps fulfill many of the short-term recovery objectives outlined by Poisson and Ursic (2013) Future butternut research should shift towards marker assisted breeding programs in an effort to identify loci associated with tolerance or resistance. Long-term butternut research should revisit the topic of population genetics and range contiguity to allow for a perspective of genetic change over time. Current management policy should consider and accommodate for the differences in allele variation between upland and riparian sites. Riparian habitats are capable of supporting greater butternut diversity and should be given special consideration when developing management plans.

Acknowledgments We would like to thank everyone who responded with information used to locate butternut and those who were gracious enough to assist in sampling, Dale Bergdahl, Kyle Lombard, John Werits, the New York Nut Growers Association, and Cornell Extension services. Alejandra Sanchez and Gloria Broders, thank you for the assistance with technical support and genotyping expertise. Ideas and edits were carefully pored over by Isabel Munck and Adrienne Kovack. We would like to express our gratitude to the anonymous reviewers who provide encouraging and constructive feedback. This research was funded in part by the National Geographic Society Committee for Research and Exploration and the University of New Hampshire Agriculture Experiment Station.

Data Archiving Statement Microsatellite genotypes for the butternut sampled have been archived in TreeGenes under Accession number TGDR056. Habitat, phenotypic, and geographic location data may be requested from the corresponding author (kirk.broders@colostate.edu)

References

Allendorf FW (1986) Genetic drift and the loss of alleles versus heterozygosity. Zoo Biol 5:181–190

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

Balloux F, Lugon-Moulin N (2002) The estimation of population differentiation with microsatellite markers. Mol Ecol 11:155–165

Boraks A, Broders K (2014) Butternut (*Juglans cinerea*) health, hybridization and recruitment in the northeastern United States. Can J Forest Res 44:1244–1252

Broders KD, Boland GJ (2011) Reclassification of the butternut canker fungus, *Sirococcus clavigignenti-juglandacearum*, into the genus *Ophiognomonia*. Fungal Biol 115:70–79

Broders K, Boraks A, Barbison L, Brown J, Boland GJ (2015) Recent insights into the pandemic disease butternut canker caused by the invasive pathogen *Ophiognomonia clavigignenti-juglandacearum*. Forest Pathol 45:1–8

Clarke K, Ainsworth M (1993) A method of linking multivariate community structure to environmental variables. Mar Ecol-Prog Ser 92: 205–219

Cornuet JM, Luikart G (1997) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics 144:2001–2014

Crane HL, Reed CA, Wood MN (1937) Nut breeding. In: USDA yearbook of agriculture: 75th Congress, 1st Session. House document 28. US Government Printing Office, Washington DC, pp 827–889

Cummings-Carlson J, Guthmiller M (1993) Incidence and severity of butternut canker in Wisconsin in 1976 and 1992. Phytopathology 83:1352

Earl DA (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359–361

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

Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9. 3). J Hered 86:485–486

Hoban SM, McCleary TS, Schlarbaum SE, Romero-Severson J (2009) Geographically extensive hybridization between the forest trees American butternut and Japanese walnut. Biol Letters rsbl-2009

Hoban SM, Borkowski DS, Brosi SL et al (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 (2012) A rare case of natural regeneration in butternut, a threatened forest tree, is parent and space limited. Conserv Genet 13:1447–1457

Hu LJ, Uchiyama K, Shen HL, Ide Y (2010) Multiple-scaled spatial genetic structures of *Fraxinus mandshurica* over a riparian–mountain landscape in Northeast China. Conserv Genet 11:77–87

Jump AS, Peñuelas J (2007) Extensive spatial genetic structure revealed by AFLP but not SSR molecular markers in the wind-pollinated tree, *Fagus sylvatica*. Mol Ecol 16:925–936

McCleary TS, Robichaud RL, Nuanes S et al (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

McKenna JR, Ostry ME, Woeste KE (2013) Resistance breeding to mitigate butternut canker disease. 24th USDA Interagency Research Forum on Invasive Species, January 8–11, 2013. Loews Annapolis Hotel, Annapolis

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

Morin R, Beaulieu J, Deslauriers M, Daoust G, Bousquet J (2000) Low genetic diversity at allozyme loci in *Juglans cinerea*. Can J Bot 78: 1238–1243

Nair V, Kostichka CJ, Kuntz J (1979) *Sirococcus clavigignenti-juglandacearum*: an undescribed species causing canker on butternut. Mycologia 71:641–646

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

Orchard LP, Kuntz JE, Kessler KJ Jr (1982) Reaction of Juglans species to butternut canker and implications for disease resistance. In Black Walnut for the Future, USDA Forest Service. North Central Forest Experiment Station. Gen. Tech. Rep. NC pp 27–31

Orlóci L (1978) Multivariate analysis in vegetation research. Junk, Den Haag, 451 pp

Ostry ME, Pijut PM (2000) Butternut: an underused resource in North America. HortTechnology 10:302–306

Ostry M, 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 JW, Coggeshall MV, Seifert J, Woeste K, Overton R, Ponder F Jr (eds) Black walnut in a new century, proceedings of the 6th Walnut Council research symposium; 2004 July 25–28; Lafayette, IN. Gen. Tech. Rep. NC-243. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station. 114–120. 188 pp

Parks A, Jenkins M, Woeste K, Ostry M (2013) Conservation status of a threatened tree species: establishing a baseline for restoration of *Juglans cinerea* L. in the southern Appalachian Mountains, USA. Nat Area J 33:413–426

Parks A, Jenkins M, Ostry M, Zhao P, Woeste K (2014) Biotic and abiotic factors affecting the genetic structure and diversity of butternut in the southern Appalachian Mountains, USA. Tree Genet Genomes 10:541–554

Peakall R, Smouse PE (2006) GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 6:288–295

Poisson G, Ursic M (2013) Recovery strategy for butternut (*Juglans cinerea*) in Ontario. Ontario Recovery Stategy Series. Prepared for the Ontario Ministry of Natural Resources, Peterborough, Ontario. Adoption of the Recovery Strategy for the Butternut (*Juglans cinerea*) in Canada (Environment Canada 2010)

Pollegioni P, Woeste K, Mugnozza GS, Malvolti ME (2009) Retrospective identification of hybridogenic walnut plants by SSR fingerprinting and parentage analysis. Mol Breeding 24:321–335

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

Pritchard J, Wen W, Falush D (2003) Documentation for structure software: version 2. http://pritchardlab.stanford.edu. Accessed 23 October 2015

Rink G (1990) *Juglans cinerea* L. butternut. In Silvics of North America. Vol. 2. Hardwoods. Tech. Coords. R.M. Burns and B.H. Honkala. USDA For. Serv. Agric. Handbook 654, Washington, DC, 386–390

Ross-Davis A, Woeste K (2007) Microsatellite markers for *Juglans cinerea* L. and their utility in other Juglandaceae species. Conserv Genet 9:465–469

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

Rousset F (2008) Genepop'007: a complete re-implementation of the genepop software for Windows and Linux. Mol Ecol Res 8:103–106

Schultz J (2003) Conservation assessment for butternut or white walnut (*Juglans cinerea* L.). USDA For. Serv., Eastern Region. http://www.fs.usda.gov/Internet/FSE_DOCUMENTS/fsm91_054130.pdf

Shannon C, Weaver W (1949) The mathematical theory of communication. University of Illinois Press, Urbana

Smouse PE, Long JC (1992) Matrix correlation analysis in anthropology and genetics. Am J Phys Anthropol 35:187–213

Smouse P, Long J, Sokal R (1986) Multiple regression and correlation extensions of the Mantel test of matrix correspondence. Syst Zool 35:627–632

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

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

Victory E (2006) Genetic homogeneity in *Juglans nigra* (Juglandaceae) at nuclear microsatellites. Am J Bot 93:118–126

Ward M, Dick CW, Gribel R, Lowe J (2005) To self, or not to self… a review of outcrossing and pollen-mediated gene flow in neotropical trees. Heredity 95:246–254

Weir B, Cockerham C (1984) Estimating f-statistics for the analysis of population structure. Evolution 38:1358–1370

Williams M (1992) Americans and their forests: a historical geography. Cambridge University Press, New York

Wright S (1931) Evolution in Mendelian populations. Genetics 16: 97–159

Yao X, Zhang J, Ye Q, Huang H (2011) Fine-scale spatial genetic structure and gene flow in a small, fragmented population of *Sinojackia rehderiana* (Styracaceae), an endangered tree species endemic to China. Plant Biol 13:401–410