

# Geographic patterns of genetic variation in native pecans

L. J. Grauke · Maria Azucena Mendoza-Herrera ·  
Allison J. Miller · Bruce. W. Wood

Received: 13 August 2010 / Revised: 23 February 2011 / Accepted: 1 March 2011 / Published online: 25 March 2011  
© Springer-Verlag 2011

**Abstract** A structured collection of 80 seedling pecan trees [*Carya illinoensis* (Wangenh.) K. Koch], representing 19 putatively native pecan populations across the species range, was evaluated at three plastid and 14 nuclear microsatellite (simple sequence repeat, SSR) loci. Data were analyzed using a priori population designations and also within a Bayesian framework, in which individuals were assigned to clusters regardless of population of origin. Population genetic analyses using a priori populations, clusters based on chloroplast microsatellite data (cpSSR), and clusters based on nuclear microsatellite data (nSSR) yielded consistent results. For all groupings, cpSSR variation exhibited more geographic structure than the nSSR data.

Furthermore, cpSSR microsatellite diversity decreased with increasing latitude, but this pattern was not observed with the nuclear data. Contrasting patterns in plastid and nuclear genetic diversity demonstrate unique aspects of postglacial recolonization reflected in the movement of seeds versus pollen. These data suggest that plastid SSRs are useful tools for identifying population structure in pecan and hold promise for ongoing efforts to identify and conserve representative germplasm in ex situ collections.

**Keywords** *Carya illinoensis* · Microsatellites · Plastid SSRs · Nuclear SSRs · Postglacial recolonization · Dispersal

---

Communicated by A. Dandekar

**Electronic supplementary material** The online version of this article (doi:10.1007/s11295-011-0384-4) contains supplementary material, which is available to authorized users.

---

L. J. Grauke (✉)  
USDA ARS Pecan Breeding and Genetics,  
10200 FM 50,  
Somerville, TX 77879, USA  
e-mail: lj.grauke@ars.usda.gov

M. A. Mendoza-Herrera  
Norman Borlaug Center for Southern Crop Improvement,  
Texas A&M University,  
College Station, TX 77843-2123, USA

A. J. Miller  
Department of Biology, Saint Louis University,  
3507 Laclede Ave,  
St. Louis, MO 63103, USA

B. W. Wood  
USDA ARS Southeastern Fruit and Tree Nut Research Lab,  
21 Dunbar Road,  
Byron, GA 31008, USA

## Introduction

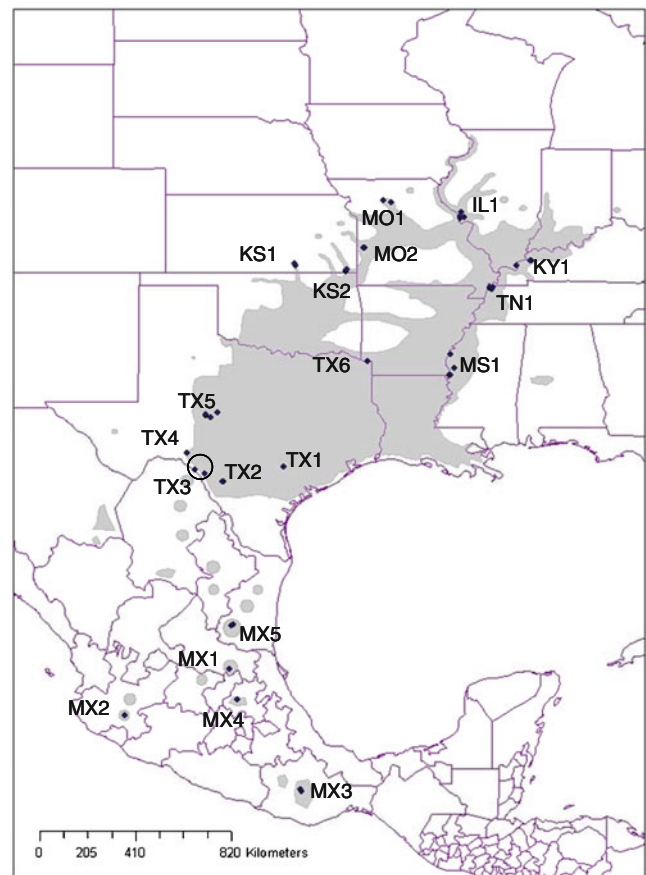
Genetic variation provides the foundation for species to respond to changes in the natural environment (Jump et al. 2009). It is well understood that intraspecific variation is distributed non-randomly within and among plant populations (Caicedo et al. 2004; Soltis et al. 2006), and that contemporary patterns of genetic variation among populations within species result from ongoing evolutionary processes as well as historical factors (Mitchell-Olds et al. 2007; Sloan et al. 2008). Understanding the processes shaping the geographic structure of genetic variation has become a central focus of evolutionary biology (Charlesworth et al. 2003; Mitchell-Olds et al. 2007). It is particularly relevant for the wild relatives of domesticated species, which constitute an important component of the crop gene pool (Harlan 1992).

Pecan [*Carya illinoensis* (Wangenh.) K. Koch] is the most economically important native nut tree in North America. A member of the hickory genus (*Carya* spp., Juglandaceae) pecans are monoecious, with male and

female flowers produced on the same tree but maturing at different times in a heterodichogamous blooming pattern that minimizes inbreeding (Thompson and Romberg 1985). Pollen is distributed by wind and may travel long distances. In contrast, fruits appear to be dispersed on a local scale. This reproductive system results in different levels of dispersal between male and female gametes. Over time, such a system has implications for population structure (Namkoong and Gregorius 1985). To date, no studies have compared geographic patterns of variation in pecan based on nuclear (pollen) and plastid (seed) transmission.

Native pecan populations are an important resource for the pecan industry. Pecan production is marketed in two major categories: “Native/Seedling” and “Improved”. “Native” pecans are non-grafted trees growing in natural, regenerating stands not established by humans. “Seedling” pecans are non-grafted trees grown from selected nuts and/or intentionally planted. “Improved” pecans are selected cultivars that are asexually propagated by budding or grafting onto native or seedling rootstocks. The USA and Mexico are world leaders in pecan production (Flores and Ford 2009), but pecans are also grown commercially in Argentina, Australia, Brazil, China, Egypt, Peru, and South Africa. Pecans were first successfully grafted in 1846 or 1847 on the Oak Alley Plantation in St. James Parish, Louisiana (Taylor 1905). Commercial nursery production of grafted pecan trees began in the late 1800s and has resulted in steadily increasing acreage and production from improved pecans. Pecan production from improved trees began to exceed native production in the 1960s, while native production began a steady decline in the 1970s. Range-wide collections from native pecan populations have been added to National Clonal Germplasm Repository collections in order to preserve broad genetic diversity of a diminishing resource.

Native pecan populations occupy the westernmost range of any *Carya* species (Thompson and Grauke 1991) and occur in regenerating stands from Iowa south to the Gulf Coast of Louisiana, Texas, and into Mexico (Fig. 1). Some populations considered to be native may have been initially dispersed into their present locations by prehistoric people. Pecan was used by the people living at Modoc Rock Shelter in Randolph County, Illinois, near the northern limit of the native range, over 10,000 years ago (Styles et al. 1983). Pecans were recovered from strata over 8,000 years old from Baker's Cave in Val Verde County (Dering 1977; Hester 1981), the westernmost of the contiguous native stands in Texas. Braun (1961) considered the pecans in Butler County, OH, USA to be native. They are closely associated with Adena culture mounds, artifacts of a people who lived in that area from 2500 BP to 400 AD (Fagan 1992). Among the easternmost populations of putatively native pecans is the disjunct population in west central



**Fig. 1** Map of the distribution of native pecan trees (shaded area), showing the locations of 19 populations as designated in Table 1

Alabama on the Black Warrior River (Harper 1928). That population is near Moundville, an important site of the Mississippian culture of west central Alabama where archeological excavations confirm pecan was present as early as the Moundville I phase (1050–1250 AD) (Welch and Scarry 1995). The distribution of pecan in Mexico (Manning 1949, Manning 1962; Miranda and Sharp 1950) is not qualified by archeological excavations, but is associated in many locations with other plant species that imply the remnants of an ancient flora (Stone 1962; Graham 1999). Ultimately, long-established populations imply adaptation to local conditions that may be genetically unique and valuable.

In this study, plastid and nuclear genetic variation were surveyed in a representative sample of 80 indigenous pecan trees collected from throughout the native range of species. The primary goal was to assess the geographic distribution of nuclear- and plastid-encoded genetic variation in native pecans, and to identify evolutionary processes that might underlie these patterns. Our specific objectives were to: (1) assess geographic structure of genetic variation based on plastid and nuclear SSRs; (2) explore latitudinal trends in genetic variation; and (3) identify the geographic origins of Mexican *C. illinoensis* populations.

**Table 1** Sample localities and organization of 80 indigenous pecan accessions

Accession	Population no.	Population name	Latitude (°N)	Longitude (°W)
1	1	MX3	16.950608	96.757078
2	1	MX3	16.951017	96.756251
3	1	MX3	17.047174	96.790155
4	2	MX2	19.890810	103.586228
5	2	MX2	19.892520	103.588070
6	2	MX2	19.898786	103.596475
7	2	MX2	19.901079	103.595892
8	2	MX2	19.903901	103.596049
9	3	MX4	20.484237	99.222044
10	3	MX4	20.484709	99.220450
11	3	MX4	20.484985	99.223512
12	3	MX4	20.485628	99.221162
13	3	MX4	20.487144	99.222448
14	4	MX1	21.667270	99.539070
15	4	MX1	21.671560	99.539370
16	4	MX1	21.675034	99.540738
17	4	MX1	21.677638	99.543476
18	5	MX5	23.366721	99.475469
19	5	MX5	23.368471	99.479403
20	5	MX5	23.403687	99.382995
21	5	MX5	23.403723	99.382463
22	5	MX5	23.405077	99.381056
23	6	TX2	28.900406	99.778895
24	6	TX2	28.901271	99.783354
25	6	TX2	28.903418	99.780011
26	6	TX2	28.909848	99.780621
27	6	TX2	28.917739	99.795057
28	7	TX3	29.227754	100.488571
29	7	TX3	29.229297	100.481403
30	7	TX3	29.230118	100.485845
31	7	TX3	29.230255	100.481513
32	7	TX3	29.370957	100.884238
33	8	TX1	29.485231	97.463405
34	8	TX1	29.486244	97.466120
35	8	TX1	29.490094	97.466109
36	8	TX1	29.494697	97.460224
37	9	TX4	30.025400	101.169245
38	9	TX4	30.025908	101.169207
39	9	TX4	30.031706	101.171786
40	10	TX5	31.370275	100.267390
41	10	TX5	31.371668	100.267540
42	10	TX5	31.372814	100.267273
43	10	5X5	31.481551	100.467476
44	10	TX5	31.482114	100.470100
45	10	TX5	31.570800	100.011100
46	11	TX6	33.567941	94.202841
47	11	TX6	33.571370	94.204109
48	11	TX6	33.572334	94.200555
49	11	TX6	33.574408	94.205171

**Table 1** (continued)

Accession	Population no.	Population name	Latitude (°N)	Longitude (°W)
50	11	TX6	33.575331	94.201662
51	12	MS1	33.017763	91.039002
52	12	MS1	33.033793	91.030287
53	12	MS1	33.271546	90.856405
54	12	MS1	33.839308	91.027645
55	13	KY1	37.237483	88.463644
56	13	KY1	37.426200	87.906800
57	13	KY1	37.433100	87.922200
58	13	KY1	37.433700	87.918500
59	14	TN1	36.354200	89.410500
60	14	TN1	36.354827	89.428453
61	14	TN1	36.376600	89.501000
62	14	TN1	36.395257	89.481682
63	14	TN1	36.409907	89.361919
64	15	KS2	37.024931	95.046012
65	15	KS2	37.061523	95.059649
66	15	KS2	37.081993	95.017904
67	16	KS1	37.239429	97.011687
68	16	KS1	37.312209	97.028621
69	17	MO2	37.947363	94.334807
70	17	MO2	37.949715	94.353059
71	17	MO2	37.953287	94.347401
72	18	MO1	39.701346	93.304527
73	18	MO1	39.701688	93.305262
74	18	MO1	39.702152	93.306791
75	18	MO1	39.757901	93.608191
76	19	IL1	39.071800	90.642400
77	19	IL1	39.085500	90.616100
78	19	IL1	39.109146	90.493956
79	19	IL1	39.139900	90.617400
80	19	IL1	39.300437	90.595044

## Materials and methods

**Sampling and DNA extraction** Leaves for DNA extraction were collected from a pecan provenance orchard in Byron, Georgia. This collection was established from seed collected across the range of pecan in 1986 and 1987 (Grauke et al. 1989, Wood et al. 1998). Samples from 19 indigenous pecan populations were collected from across the USA and Mexico. Seed was collected from up to six trees in each population. The latitude and longitude of the original trees was estimated using collection records and Google Earth. We collected leaf samples from one seedling of each of 80 of the original mother trees to identify broad-scale patterns of genetic variation (Table 1, Fig. 1).

Young leaves were harvested and frozen at  $-80^{\circ}\text{C}$  until DNA extraction. Total genomic DNA was extracted using methods modified from those of Paterson et al. (1993). Frozen tissue was ground and placed in an extraction buffer [0.35 M glucose, 0.1 M Tris-HCl pH=8, 5 mM Na<sub>2</sub>EDTA pH=8, 2% (w/v) polyvinylpyrrolidone (PVP-40)] at pH=7.5 and a lysis buffer [0.1 M Tris-HCl pH=8, 1.4 M NaCl, 0.02 M Na<sub>2</sub>EDTA pH=8, 2% (w/v) CTAB, and 2% PVP-40]. At the time of DNA extraction, 1% (w/v) of ascorbic acid and 0.2% (v/v) of  $\beta$ -mercaptoethanol were added to both buffers. DNA was cleaned with chloroform/isoamyl alcohol (24:1) and precipitated with salt and isopropanol or ethanol. Successful extractions were also made using the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) following the manufacturer's instructions.

**SSR amplification** Eight primer pairs targeting microsatellite loci in organellar DNA were identified by previous workers (Bryan et al. 1999; Weising and Gardner 1999; Cheng et al. 2006). A preliminary study found three polymorphic loci [ccmp2 (Weising and Gardner 1999), ntcp40, and ntcp9 (Bryan et al. 1999)] that amplified and produced consistent fragment sizes across runs of *C. illinoensis* individuals. PCR amplifications were performed in PerkinElmer 9600 or 9700 thermal cyclers, using 10–30 ng of genomic DNA in a reaction volume of 12  $\mu$ l. The PCR mix contained 20 mM Tris–HCl (pH 8.4), 50 mM KCl, 2.5 mM MgCl<sub>2</sub>, 0.1 mM each of dATP, dGTP, dTTP and dCTP (Promega), 0.04 ml of Taq DNA polymerase (Invitrogen), and a mix of the forward and reverse primers in a final concentration of 0.25 pmol/ml. The thermal cycling protocol consisted of 3 min at 94°C followed by 35–40 cycles of 45 s at 94°C, 45 s at 55°C and 1 min at 72°C, and final elongation of 40–60 min at 72°C. In addition, 14 nuclear microsatellites (Grauke et al. 2003; Mendoza-Herrera et al. 2008), including six previously identified in *Juglans* (Woeste et al. 2002; Dangl et al. 2005) were amplified for the 80 accessions. After PCR, 3  $\mu$ l of each sample was loaded in a 2% agarose gel in 1 $\times$  TAE buffer and stained with ethidium bromide to verify amplification.

PCR fragments were visualized using an ABI Prism Genetic Analyzer 3130 (Applied Biosystems, Foster City, CA, USA). For analysis in the capillary system, forward primers were fluorescently labeled at the 5' end using either 6-FAM or HEX. Two microliters of sample was diluted 1:30 with deionized water prior to sizing. Ten microliters 400-ROX internal size standard in deionized formamide at 2.5% was added to each sample to estimate fragment size using GeneScan and Genotyper software v 3.7 (Applied Biosystems). Samples were multiplexed when possible. Fragments were scored using Genotyper software v3.7. The size of amplified products was calculated based on the 400ROX internal standard. Alleles at each SSR locus were called by the size in base pairs expressed as a whole number. The range of sizes within each allele class was facilitated by FlexiBinV2 (Amos et al. 2007). Some samples had two replicates to monitor the reproducibility between runs and many were run multiple times. Repeated runs included known standards of diverse haplotypes to insure consistency.

**Data analysis** Plastid and nuclear SSR datasets were analyzed separately using Genalex 6.41 (Peakall and Smouse 2006). For the nuclear loci, the number of alleles per locus ( $N_a$ ), effective number of alleles ( $N_e$ ), Shannon's information index (I), observed heterozygosity ( $H_o$ ), expected heterozygosity ( $H_e$ ), unbiased expected heterozygosity (UHe), and the fixation index ( $F=H_e-H_o/H_e$ ) were estimated. For each plastid locus,  $N_a$ ,  $N_e$ , I, Diversity ( $h$ , =

1-(sum of squared population allele frequencies)), and unbiased diversity ( $uh$ , =  $(N/(N-1)) * h$ ) were calculated.

The 80 samples included in this study were collected from 19 geographically distinct populations (Table 1). Two distinct approaches were taken to analyze the data: (1) analyses were completed for a priori populations, but sample numbers were low ( $N=2-5$ ), and (2) a Bayesian framework was used to assign individuals to clusters, regardless of population of origin. In the first approach, standard population genetic parameters (listed above) were generated for each population:  $N_a$ , number of private (unique) alleles ( $P_a$ ), and number of polymorphic loci (PL). Population structure was assessed using AMOVA and Mantel tests were employed to test the null hypothesis that there is no association between geographic distance and genetic distance. Analyses were performed using Genalex 6.41 (Peakall and Smouse 2006). In the second approach, we used the model-based clustering method STRUCTURE version 2.3.2 (Pritchard et al. 2007) to assign individuals to homogeneous clusters (K populations), regardless of their geographic origin. Separate analyses were conducted for the organellar SSR markers and the nuclear SSR markers. Following a burn-in of  $1 \times 10^5$  Markov chain Monte Carlo (MCMC) iterations, five independent runs for  $K=1-10$  (three for the nuclear data) were run using  $1 \times 10^6$  MCMC iterations. For the organellar and nuclear SSR datasets, we specified that allele frequencies were independent and set  $\lambda$  to 1. Two separate analyses were run for each dataset using: (1) admixture ancestry model and (2) no admixture model. The most likely number of populations ( $K$ ) was identified based on the highest value of Ln P(D), a model-choice criterion that estimates the posterior probability of the data. Clusters identified based on the STRUCTURE results of the organellar and nuclear SSR dataset were treated as populations in subsequent analyses (“cpSSR-defined clusters” and “nSSR-defined clusters”). Using the cpSSR-defined clusters, we estimated genetic diversity and population structure for the cpSSR data and for the nuclear data. Then, we used the nSSR-defined clusters to estimate genetic diversity and structure of the nSSR data and the cpSSR data. This approach allowed us to compare levels of nuclear and plastid SSR genetic variation and population structure for clusters identified with each marker type (cpSSR and nSSR).

To explore relationships among individuals, a distance-based approach was implemented using the program DARwin (Perrier and Jacquemoud-Collet 2006). The dissimilarity index was calculated for the single-count data. Tree construction was completed using the weighted neighbor-joining method (Saitou and Nei 1987) and exported for visualization in the program FigTree v1.3.1 (downloaded Feb. 2011 from <http://tree.bio.ed.ac.uk/software/figtree>).



## Results

The number of alleles per locus ranged from four to five for the plastid SSRs and from 2 to 22 for the nuclear SSRs; these and additional locus-specific measures are reported in Table 2 (and Supplemental Table 1). All loci were polymorphic for both datasets (Table 2). Analyses of the 19 original populations revealed variation in number of polymorphic loci and number of alleles per population (Table 3; other population genetic parameters not reported due to low sample numbers). The mean percentage of polymorphic loci per population was  $24.56\% \pm 7.14\%$  for cpSSRs and  $77.07\% \pm 2.48\%$  for the nuclear SSRs (Table 3). The average number of alleles per locus was  $1.298 \pm 0.075$  for cpSSRs and  $3.226 \pm 0.123$  for nSSRs. Private cp SSR alleles were identified in 3 of the 19 a priori populations: MX2 (two private alleles), MX5 (one private allele), and TX 3 (one private allele). Private nSSR alleles were identified in 16 of the 19 a priori populations (Table 3). AMOVA indicated that 74% of the plastid molecular variance was distributed among a priori populations, with 26% distributed within a priori populations ( $P < 0.001$ ). In contrast, 11% of the nuclear molecular variance was distributed among populations, with 89% of the variance distributed within populations ( $P < 0.001$ ) (Table 4). Mantel tests rejected the null hypothesis of no association between

geographic distance and genetic distance, indicating a significant correlation between geography and genetics for the sampled populations based on plastid ( $P = 0.000$ ,  $r = 0.461$ ) and nuclear SSR data ( $P = 0.000$ ,  $r = 0.165$ ).

**STRUCTURE analyses based on cpSSR data** A priori population designations were removed and Bayesian analyses were used to determine how the 80 individuals clustered together based on their cpSSR profiles. The cpSSR data clustered samples into three groups (admixture model  $K = 3$ ,  $\text{LnP}(D) = -131.2$ ; no admixture model  $K = 3$ ,  $\text{LnP}(D) = -116.1$ ). Bar plots for the admixture model are shown in Fig. 2a (accessions listed in their original order as in Table 1) and 2b (accessions listed by q score). There were no differences between the admixture and no admixture models for the cpSSR data (data not shown). Individuals were considered part of a group if more than 80% of their profile grouped with a given cluster. STRUCTURE analysis based on plastid SSRs unambiguously assigned 92.5% of individuals (74/80 individuals) to one of the three clusters (Fig. 2). Six individuals were assigned nearly equally to two clusters (#s 20, 22 (MX5) 34 (TX1), 46 (TX1), 55 (KY1), 64 (KS2)) (Fig. 2a,b). Cluster one (blue in Fig. 2, southern cluster; average latitude  $19.94453^\circ$  N, longitude  $99.98127^\circ$  W) included 15 individuals collected from four populations in Mexico (MX1,

**Table 2** Diversity statistics for 80 native pecan accessions by genome and locus

Genome	Locus	$N_a$	$N_e$	Index	$h$	$uh$	$H_o$
Plastid	ccmp2	4	2.282	0.958	0.562	0.569	
	ntcp40	5	2.790	1.161	0.642	0.650	
	ntcp9	4	1.484	0.668	0.326	0.330	
	Mean/SE	$4.333 \pm 0.211$	$2.185 \pm 0.380$	$0.929 \pm 0.143$	$0.510 \pm 0.095$	$0.516 \pm 0.061$	
Nucleus	cin22	3	1.416	0.565			0.213
	cin13	6	2.062	0.995			0.400
	cin20	2	1.064	0.139			0.038
	cin4	12	3.197	1.626			0.667
	cin23	3	1.375	0.520			0.238
	ca10	2	1.013	0.038			0.013
	ga38	13	4.912	1.992			0.713
	ga39	16	3.833	1.942			0.663
	ga41	3	1.396	0.544			0.150
	wga242	19	9.229	2.450			0.800
	wga118	16	6.995	2.293			0.725
	wga321	22	14.050	2.801			0.875
	wga4	22	9.202	2.552			0.650
	I-A05	4	1.391	0.544			0.313
	Mean/SE	$10.214 \pm 2.071$	$4.367 \pm 1.080$	$1.357 \pm 0.261$			$0.461 \pm 0.080$

Mean and standard error provided across all loci by genome

$N_a$  number of alleles per locus,  $N_e$  effective number of alleles,  $I$  Shannon's Information index,  $h$  diversity for plastid loci =  $1 - (\text{sum of squared population allele frequencies})$ ,  $uh$  unbiased diversity for plastid loci =  $(N/N-1) * h$ ,  $H_o$  observed heterozygosity

**Table 3** Diversity indices for 19 a priori native pecan populations

Population	No.	Latitude (°N)	Longitude (°W)	cpSSRs			nSSRs		
				PL (of 3)	Na	Pa	PL (of 14)	Na	Pa
MX3	3	16.98293	96.76783	0	1.00	0	8	2.07±0.34	0
MX2	5	19.89742	103.59254	2	2.33±0.67	2	9	2.29±0.32	1
MX4	5	20.48534	99.22192	3	2.00	0	10	3.21±0.54	4
MX1	4	21.67288	99.54066	0	1.00	0	8	3.00±0.59	1
MX5	5	23.38954	99.42028	1	1.67±0.67	1	11	3.64±0.59	2
TX2	6	28.90602	99.78299	0	1.00	0	10	3.71±0.64	2
TX3	5	29.25768	100.56431	1	1.33±0.33	1	11	3.64±0.56	0
TX1	5	29.48850	97.46440	1	1.33±0.33	0	9	3.21±0.58	2
TX4	4	30.02767	101.17008	0	1.00	0	12	2.79±0.37	2
TX5	6	31.44154	100.29181	0	1.00	0	12	4.21±0.66	1
TX6	5	33.57228	94.20287	2	1.67±0.33	0	11	3.57±0.65	0
MS1	4	33.29060	90.98834	2	1.67±0.33	0	13	3.64±0.59	6
KY1	4	37.38262	88.05279	1	1.33±0.33	0	11	2.64±0.37	1
TN1	5	36.37816	89.43671	0	1.00	0	12	3.86±0.64	0
KS2	3	37.05615	95.04119	1	1.33±0.33	0	12	3.29±0.40	1
KS1	2	37.27582	97.02015	0	1.00	0	10	2.29±0.27	1
MO2	3	37.95012	94.34509	0	1.00	0	13	2.86±0.39	0
MO1	4	39.71577	93.38119	0	1.00	0	12	2.93±0.36	1
IL1	5	39.14136	90.59298	0	1.00	0	11	4.14±0.74	1
Mean%PL				24.56%	1.30±0.08		77.1%	3.23±0.12	
SE				7.14%			2.48%		

Population designations follow Table 1. Latitude (°N) and longitude (°W) are decimal degree averages over the number (*N*) of accessions within a population

*PL* number of polymorphic loci, *Na* number of alleles, *Pa* number of private alleles

MX2, MX3, MX4) (Fig. 1). Cluster two (red in Fig. 2, central cluster; average latitude 30.29129° N, longitude 98.61908° W) included 30 individuals. These included all individuals from populations TX1, TX2, TX3, TX4, and TX5, with isolated individuals from KY1, KS2, MS1, MX5, and TX6 (Fig. 1). Six of the 30 individuals [20, 22 (MX5), 34 (TX1), 46 (TX6), 55 (KY1), 64 (KS2)] were assigned primarily to cluster two, but approximately 40–45% of their profile grouped with cluster three (Fig. 2b). Cluster three (green in Fig. 2; the northern cluster, average latitude 34.93399° N, longitude 93.17874° W) included 35 individuals. These included all individuals from IL1, KS1, MO2, MO1, and TN1 as well as most individuals from MS1, KS2, KY1, MX5, and TX6 (Fig. 1). Isolated individuals from MX2 and MX4 were also in this cluster. One individual that grouped with cluster three [#18 (MX5)] had a small portion of its profile group with cluster 2.

Ten individuals (12.5%) clustered at least partially with a group that differed from the region where they were collected (Fig. 2a). Five individuals collected in Mexico grouped with the northern group [07 (MX2), 11 (MX4), 18, 19, 21 (MX5)]; two individuals collected in Mexico

grouped with both the central group and the northern group [20, 22 (MX5)]. Individual 34 (TX1) clustered with both the central and the northern groups. Two individuals collected in the northern portion of the range clustered with both the northern and central clusters: 55 (KY1) and 64 (KS2).

Estimates of genetic variation for the three cpSSR-defined clusters (Table 5) reveal that plastid diversity is highest in the southernmost cpSSR-defined cluster, and decreases from south to north. To assess the impact of the six individuals that group with clusters two and three [20, 22 (MX5), 34 (TX1), 46 (TX6), 55 (KY1), 64 (KS2)], diversity parameters were estimated without these individuals. With these individuals removed, the south–north decrease in cpSSR genetic variation did not change in direction or magnitude (data not shown). Eighty-six percent of plastid molecular variance was attributed to among-cp cluster differences, with 14% due to within-cp cluster differences ( $P < 0.001$ ; Table 4).

Estimates of nSSR variation for the cpSSR-defined clusters exhibit a different geographic pattern; nuclear SSR variation increases from south to north (Table 5). AMOVA showed that 7% of the nuclear molecular variance

**Table 4** Analysis of molecular variance (AMOVA) by population and genome for 80 native pecan accessions

Population definition	SSR dataset	Source	<i>df</i>	SS	MS	Est. var.	Percentage (%)	<i>P</i> value
A priori	Plastid	Among pops.	18	2,612.958	145.164	31.869	74%	<0.001
		Within pops.	61	695.667	11.404	11.404	26%	
		Total	79	3,308.625	43.273		100%	
	Nucleus	Among pops.	18	215.263	11.959	0.999	11%	
		Within pops.	61	473.750	7.766	7.766	89%	
		Total	79	689.013		8.765	100%	
cpSSR clusters	Plastid	Among pops.	2	2,661.373	1,330.686	52.238	86%	<0.001
		Within pops.	77	647.252	8.406	8.406	14%	
		Total	79	3,308.625		60.644	100%	
	Nucleus	Among pops.	2	46.884	23.442	0.597	7%	
		Within pops.	77	642.129	8.339	8.339	93%	
		Total	79	689.013	8.936		100%	
Nuclear SSR clusters	Plastid	Among pops.	2	2,124.455	1,062.227	46.970	75%	<0.001
		Within pops.	77	1,184.170	15.379	15.379	25%	
		Total	79	3,308.625	62.349	100%	<0.001	
	Nucleus	Among pops.	2	66.875	33.428	1.137	12%	
		Within pops.	77	622.156	8.080	8.080	88%	
		Total	79	689.013		9.217	100%	

*P* (rand>=data) for PhiPT is based on permutation across the full data set. PhiPT=AP/(WP+AP), where AP is the Est. Var. Among Pops, WP is the Est. Var. within populations

SS sum of squares, MS mean square, Est. Var. estimated variance

in cpSSR-defined clusters was among-populations with 93% from within-populations ( $P<0.001$ ; Table 4).

**STRUCTURE analyses based on nuclear SSR data** Bayesian analyses of the 80 *C. illinoensis* individuals based on nSSR data indicated that the most likely number of clusters was three for both the admixture model [ $K=3$ ;  $\text{LnP}(D)=-3,033.2$ ; Fig. 3] and the no-admixture model [ $K=3$ ,  $\text{LnP}(D)=-3,030.7$ ; data not shown]. The three clusters identified based on nSSR data corresponded to southern ( $N=13$ , blue in Fig. 3; average latitude  $19.678372^\circ$  N, longitude  $100.04894^\circ$  W), central ( $N=48$ , red in Fig. 3; average latitude  $31.414147^\circ$  N, longitude  $98.811790^\circ$  W), and northern ( $N=19$ , green in Fig. 3; average latitude  $35.099938^\circ$  N, longitude  $93.260284^\circ$  W).

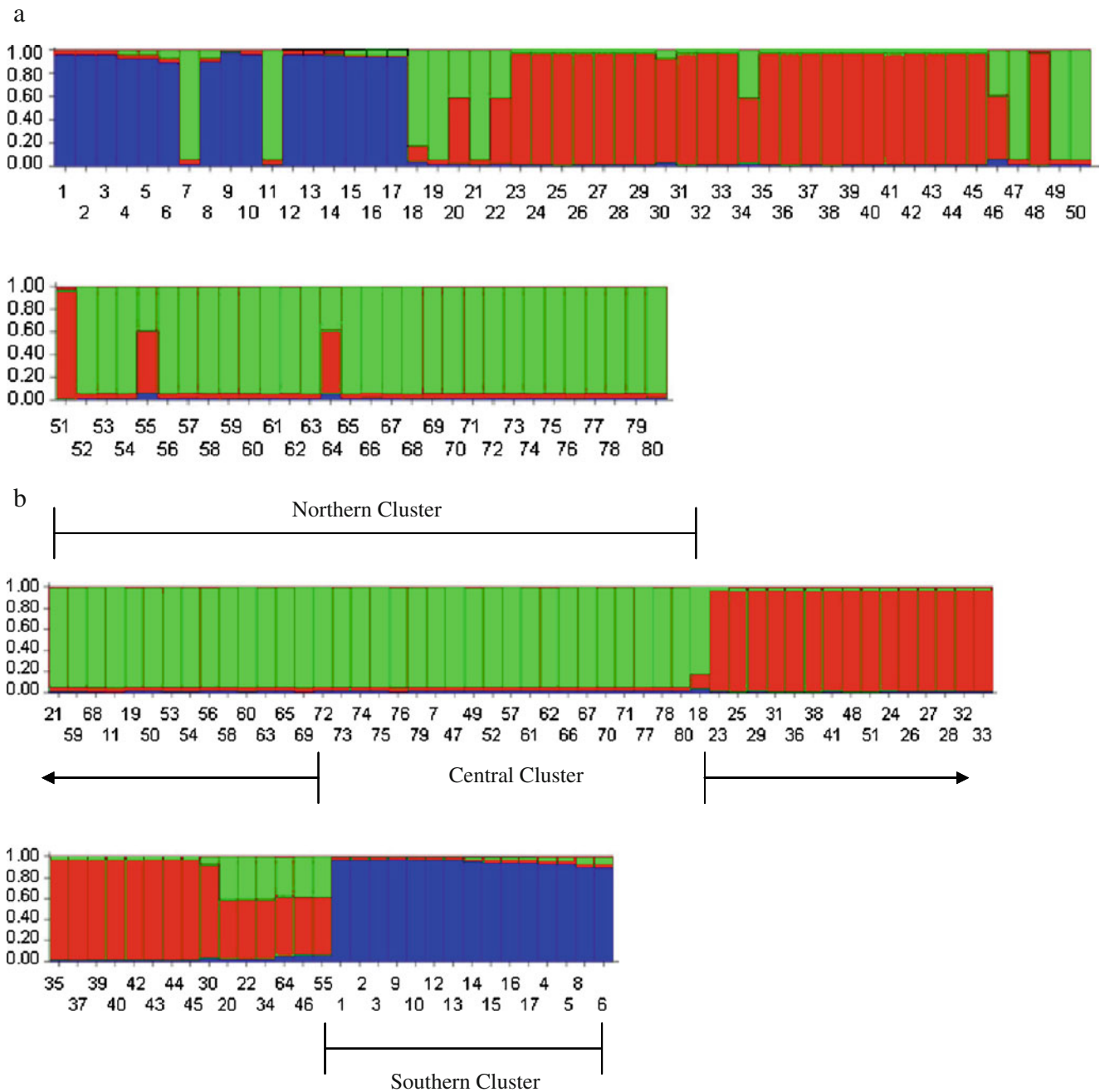
Although STRUCTURE analyses of cpSSRs and nSSRs both identified three geographically distinct clusters (southern, central, and northern), the size and composition of those clusters differed depending upon whether the clusters were defined based on chloroplast or nuclear SSR data. The primary difference between STRUCTURE analyses based on cpSSR and nSSR data is the size of the central cluster. Based on cpSSR data, the central cluster included 30 individuals but included 48 individuals based on nSSR data. Corresponding but asymmetrical size reductions are observed in the southern and northern populations: based

on cpSSR data the southern cluster included 15 individuals but included only 13 individuals based on nSSR data; the northern cluster included 35 individuals based on cpSSR data, but only 19 individuals based on nSSR data.

Estimates of nuclear genetic variation for the three nSSR-defined clusters (Table 5) revealed no consistent geographic pattern in levels of genetic variation. Twelve percent of nuclear molecular variance was attributed to among-nSSR-defined cluster differences, with 88% due to within-cluster differences ( $P<0.001$ ; Table 4). Plastid diversity decreases from south to north in the nSSR-defined clusters (Table 5) and is independent of nSSR variation. AMOVA identified that 75% of the plastid molecular variance resulted from among nSSR-defined cluster differences, with 25% derived from within-population differences ( $P<0.001$ ; Table 4).

A neighbor-joining tree based on plastid SSR data identified the same three clusters as STRUCTURE (Fig. 4). Individuals assigned to multiple groups in the STRUCTURE analysis were resolved with the cluster that had the majority representation within its profile (31, 48, 57, 66, 20, 22, 35, 36) with one exception. Individual 18, collected in Mexico (MX5 population), grouped with the northern cluster in the STRUCTURE analysis of the cpSSR data. In the neighbor-joining tree, this individual was resolved at the base of the Mexican group. The neighbor joining analysis of the nuclear SSR data produced complex assemblages of individuals with mixed geographic origins,





**Fig. 2** STRUCTURE analysis for cpSSRs using Admixture model ( $K=3$ ,  $\text{LnP}(D)=-131.2$ ) for 80 *C. illinoensis* individuals. Numbers correspond to accessions listed in Table 1. Colors correspond to southern (blue), central (red), and northern (green) clusters **a**

*Accession numbers* are shown below profiles (see Table 1 for detailed description). Accessions are listed in their original order, regardless of q score **b** Accessions are ordered based on q score

with no resolution of genetic structure related to geographic origin (data not shown).

**Discussion**

Data presented here provide important insights into the geographic distribution of genetic variation in natural pop-

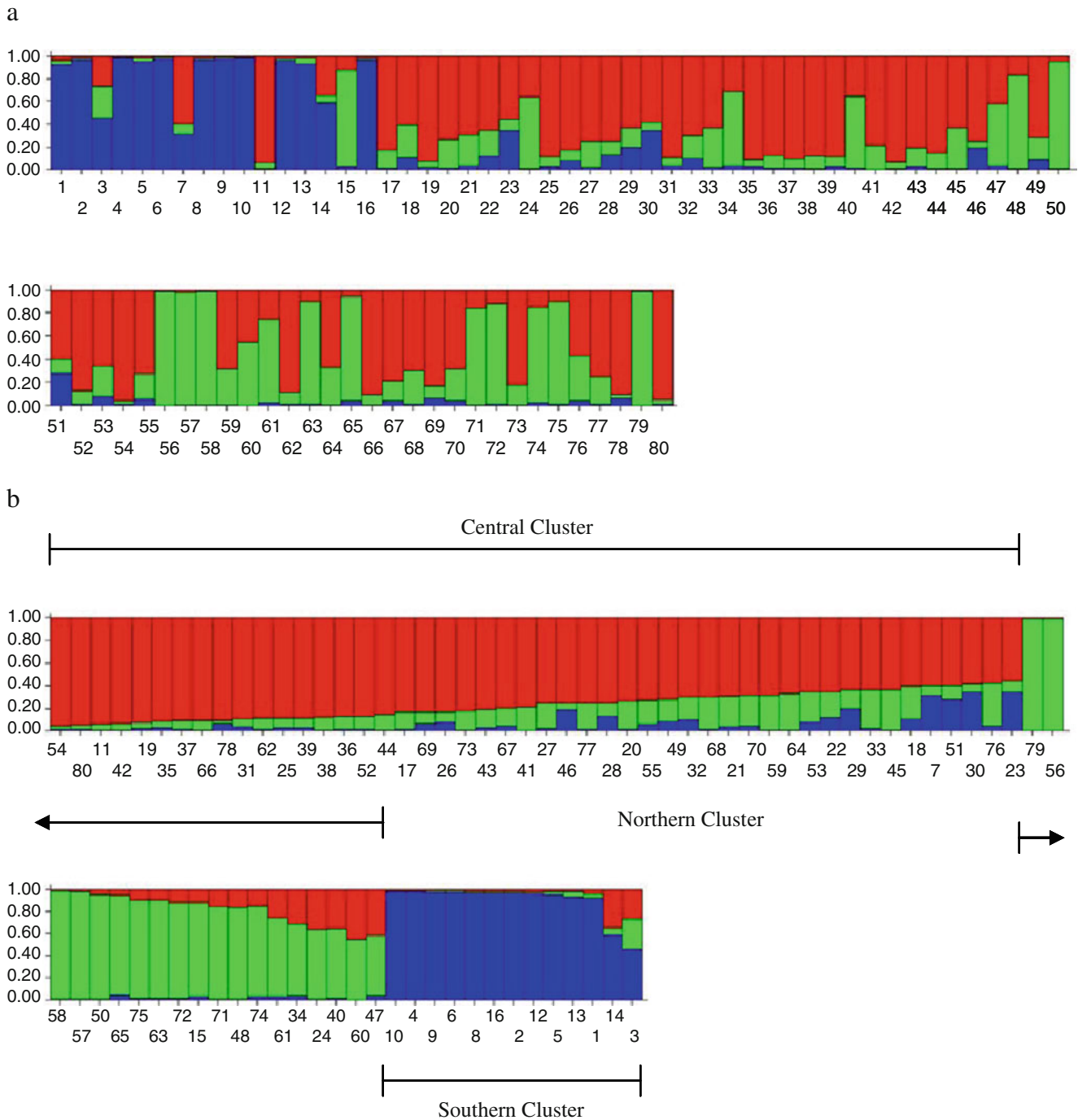
ulations of *C. illinoensis*. Previous studies of this same germplasm collection used allozymes to analyze levels of genetic variation and to estimate outcrossing rates (Rüter et al. 1999; Rüter et al. 2000). Our analyses of microsatellite data corroborate results from previous studies, and expand upon these earlier contributions by identifying differences in the geographic structure of nuclear and organellar genetic variation.

**Table 5** Diversity indices of 80 native pecan accessions by cluster within genomes, with clusters based on either cpSSRs or nSSRs

Cluster defined by	Genome	Cluster	Avg. Lat. (°N)	Avg. Long. (°W)	No.	%P	PA	Na	Ne	I	h	Ho
cpSSR	Plastid	South	19.9445	99.9813	15	100	6	2.33±0.33	1.94±0.36	0.69±0.17	0.44±0.17	0.44±0.11
		Central	30.2913	98.6191	30	66.67	3	2.00±0.58	1.18±0.09	0.26±0.14	0.14±0.07	0.14±0.07
		North	34.9340	93.1787	35	33.33	1	1.33±0.33	1.02±0.02	0.04±0.04	0.19±0.19	0.19±0.19
		Mean ± SE				66.67%±19.25%		1.89±0.26	1.38±0.18	0.33±0.12	0.20±0.07	0.20±0.07
cpSSR	Nucleus	South	19.9445	99.9813	15	78.57	4	4.78±0.97	2.96±0.63	0.94±0.22	0.35±0.08	0.35±0.08
		Central	30.2913	98.6191	30	92.86	15	7.50±1.52	4.17±0.97	1.27±0.26	0.50±0.09	0.50±0.09
		North	34.9340	93.1787	35	100	28	8.43±1.66	4.10±0.98	1.31±0.24	0.48±0.08	0.48±0.08
		Mean ± SE				90.48%±6.30%		6.90±0.83	3.74±0.50	1.17±0.14	0.44±0.05	0.44±0.05
nSSR	Plastid	South	19.6784	100.0489	13	100	4	2.33±0.33	1.93±0.32	0.70±0.16	0.45±0.10	0.45±0.10
		Central	31.4141	98.8118	48	100	2	3.00±0.58	1.77±0.36	0.58±0.24	0.37±0.16	0.37±0.16
		North	35.0999	93.2603	19	100	0	2.33±0.33	1.38±0.14	0.45±0.13	0.26±0.08	0.26±0.08
		Mean ± SE				100		2.56±0.24	1.69±0.17	0.58±0.10	0.36±0.07	0.36±0.07
nSSR	Nucleus	South	19.6784	100.0489	13	78.57	3	4.21±0.80	2.73±0.56	0.88±0.20	0.34±0.08	0.34±0.08
		Central	31.4141	98.8118	48	100	50	9.71±1.90	4.53±1.10	1.37±0.26	0.50±0.09	0.50±0.09
		North	35.0999	93.2603	19	85.71	3	5.29±1.05	3.36±0.81	1.04±0.22	0.46±0.08	0.46±0.08
		Mean ± SE				88.10%±6.30%		6.40±0.84	3.54±0.50	1.10±0.13	0.43±0.05	0.43±0.05

Mean and standard error provided across all loci by genome

*Avg. Lat.* mean over the number of observations (*N*) in degrees North Latitude, *Avg. Long.* mean over *N* in degrees West Longitude, *%P* percent polymorphic loci, *PA* private alleles, *Na* number of alleles per locus, *Ne* effective number of alleles, *I* Shannon's information index, *h* diversity for plastid loci=1-(sum of squared population allele frequencies), *Ho* observed heterozygosity=number heterozygous/*N*

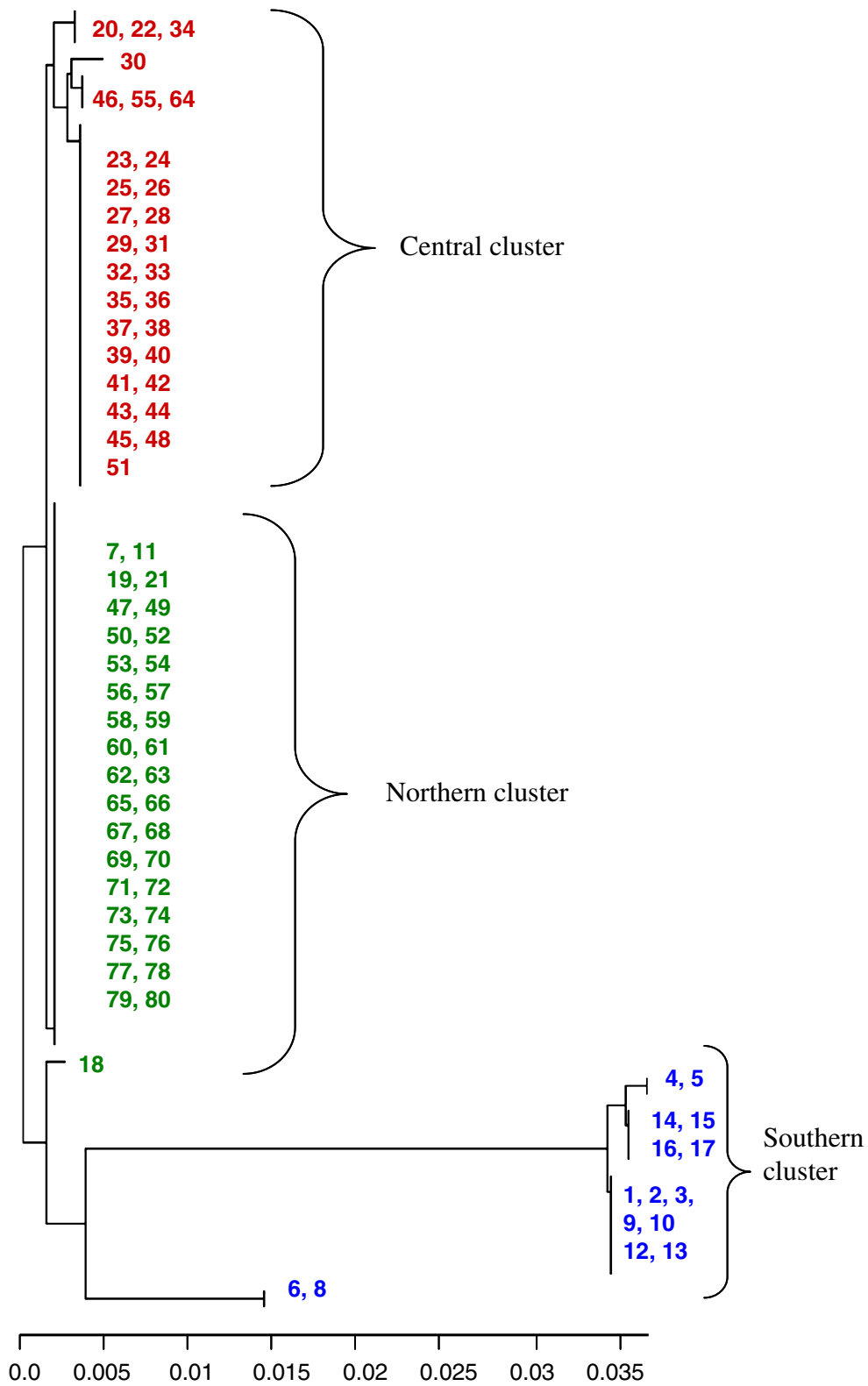


**Fig. 3** STRUCTURE analysis of nuclear SSRs using the Admixture model ( $K=3$ ;  $\text{LnP}(D)=-3033.2$ ). Accession numbers (Table 1) are shown below profiles. Colors correspond to southern (blue), central

(red), and northern (green) clusters **a** Accessions are listed in their original order, regardless of q score **b** Accessions are listed according to q score

*Levels of genetic variation in native C. illinoensis populations* As expected, estimates of nuclear genetic variation based on microsatellites are higher than levels of genetic variation observed in provenance, wild, or cultivated pecan trees based on allozyme data (Rüter et al. 1999). Compared to estimates of microsatellite variation in other

native tree populations [e.g., *Fagus japonica* (Hiraoka and Tomaru 2009); *Fraxinus mandshurica* var. *japonica* (Hu and Ennos 1999); *Juglans nigra* (Victory et al. 2006) and *Picea abies* (Meloni et al. 2007)], native pecan populations harbor relatively lower levels of genetic variation. Low levels of genetic variation could reflect relatively recent



**Fig. 4** Neighbor-joining tree based on dissimilarity index for single counts data from cpSSR. Tip labels are accession numbers from Table 1, color coded consistently with STRUCTURE analysis as shown in Fig. 2: Southern=*blue*, Central=*red*, Northern=*green*.

Labels adjacent to the same vertical bar are at the same level, regardless of position and are grouped chronologically for ease of reference

genetic reductions in diversity due to some constraint, such as disease pressure. Alternatively, rapid range expansion could also result in reduced levels of diversity.

*Contrasting geographic patterns among organellar vs. nuclear SSRs: implications for dispersal* Evolutionary analyses based on nuclear and organellar DNA often yield contrasting patterns. In phylogenetic analyses, incongruence between genomes has been interpreted as evidence for hybridization between distinct taxa (Acosta and Premoli 2010), namely through plastid or cytoplasmic capture (Tsitrone et al. 2003). Intraspecific population-level analyses based on data from multiple genomes provide an opportunity to compare movement of genetic material carried by pollen (nuclear DNA) relative to genetic material carried by seeds (nuclear and plastid DNA) (Ennos 1994; Hu and Ennos 1999). Theoretical and empirical studies have shown that organellar DNA exhibit elevated levels of population structure relative to nuclear DNA (reviewed in Latta 2006). This has been attributed to the smaller effective population size of organellar DNA and limited dispersal ability of seeds relative to pollen. Data presented here reveal that microsatellite loci from the plastid and the nucleus display strikingly different patterns of genetic variation (Figs. 2, 3; Table 5). Among-population differentiation of plastid SSRs ranged from 74–86%, compared to 7–12% for the nuclear SSRs (Table 4). This pattern of highly structured plastid SSRs and highly unstructured nuclear SSRs was consistent across different groupings of individuals (e.g., using a priori populations, cpSSR-defined clusters, nSSR-defined clusters). These data suggest that plastid SSRs are useful tools for identifying population structure in *C. illinoensis* and hold promise for ongoing efforts to identify and conserve a representative sample of *C. illinoensis* germplasm in ex situ collections.

Comparative analyses of organellar and nuclear genetic structure in crop wild relatives may shed light on historical use patterns as well as contemporary gene flow. In this study, the relative dearth of among-population genetic structure displayed by nuclear microsatellite loci, and the extensive geographic structure exhibited by plastid SSRs, suggest that pollen is the primary agent of gene flow in native *C. illinoensis* populations (Table 4)

Extensive pollen-mediated gene flow and limited seed-mediated gene flow is expected in wind-pollinated tree populations with nuts dispersed by gravity or rodents, as in *Carya* (Namkoong and Gregorius 1985). However, in tree species where fruits are dispersed long distances by birds (Vander Wall 2001) or by humans, the geographic structure of cpDNA might well be altered. Although plastid data presented here indicated significant geographic structure which implies limited seed dispersal, long-distance seed

dispersal cannot be discounted completely for *C. illinoensis*. Our analyses of cpSSR data identified a few individuals that may be the descendants of seed dispersed over long distance. For example, two individuals collected in southern portions of the range (#7 from MX2 and #11 from MX4) clustered with the northern groups based on cpSSR data (Fig. 2) and with the central group based on nSSR data (Fig. 3). These data suggest periodic long-distance dispersal via seeds.

Additional evidence for long-distance dispersal of pecan comes from the archeological record. Frequent occurrences of pecan in archeological sites in the Upper Mississippi Valley indicate that pecan was used by people characterized by the Dalton artifact complex (Smith 1992). These early Holocene foragers are described as “opportunistic dispersal agents for plant propagules, and their small camps created ephemeral disturbed soil opportunities for some pioneer species” (Smith 1992, p. 282). Abrams and Nowacki (2008) suggest that Native Americans influenced North America's forest composition by actively and passively promoting production of nuts and acorns in preference to other competing vegetation, using the tools of fire and tree girdling. Historic accounts of pecan usage indicate that people have been actively selecting and establishing plantings in new locations since the colonial period (True 1919).

*Latitudinal gradients in genetic variation may reflect unique aspects of postglacial recolonization* The geographic distribution of genetic diversity in native populations is the result of evolutionary processes acting on the raw material of progenitor populations over time. A previous analysis of North American pecan populations (Rüter et al. 1999) reported higher levels of genetic variation in northern and central *C. illinoensis* populations, and lower levels of genetic variation in southern and eastern populations based on nuclear-encoded allozyme loci. Nuclear microsatellite data presented here are consistent with this finding, but plastid data reveal the opposite pattern, with plastid diversity decreasing with latitude (Table 5). We suggest that contrasting patterns in plastid and nuclear genetic diversity resulted from unique aspects of the movement of seeds versus pollen during postglacial recolonization.

In North American and European temperate forests, glacial activity played an important role in shaping contemporary patterns of genetic variation. Fossil studies have shown that populations were restricted in refugia during glacial times, which were followed by periods of expansion as populations tracked warming climates (e.g., (Magri et al. 2006)). Molecular signatures of recolonization have been documented in numerous taxa and include increased levels of genetic variation in putative refuges



and decreasing levels of genetic variation moving away from putative refugial areas (Hewitt 1996). Geographic patterns of plastid diversity presented here conform to expectations of recolonization, and display higher levels of genetic variation and greater numbers of private alleles in southern areas relative to northern regions. During the Pleistocene epoch, repeated episodes of glacial advance restricted the northern distribution of *Carya* in North America to about 35° N at 20,000 BP (Delcourt and Delcourt 1987). As glaciers receded, hickories migrated north with rates of advance as high as 354 m/year, and reached the northernmost interglacial limit west of the Appalachian Mountains of 45° N by 8,000 years BP (Delcourt and Delcourt 1987). The homogeneity of plastid haplotypes in the northern region is consistent with the establishment of northern populations from a limited group of sources. Human-mediated dispersal may have contributed to the rapid movement of *C. illinoensis* up the Mississippi River floodplain, an area where Archaic people were active (Delcourt and Delcourt 1987; see discussion above).

In contrast to the inverse relationship observed between plastid genetic variation and latitude, northern *C. illinoensis* populations exhibit the same or greater levels of nuclear SSR variation compared to southern populations. Relatively elevated levels of nuclear genetic variation at northern latitudes may reflect a unique aspect of postglacial recolonization: interspecific gene flow in newly colonized areas. Petit et al. (2003) proposed that pollen-mediated gene flow may facilitate expansion/invasion of a species into habitats already occupied by closely related species. In these cases, pollen of the expanding species lands upon stigmas of established species in a given area, yielding F1 hybrids. This scenario is consistent with observations of highly geographically structured plastid variation but unstructured nuclear variation, if *C. illinoensis* was the initial *Carya* species to colonize areas in the north. In this case, other *Carya* species expanding into the northern regions following the establishment of *C. illinoensis* may have contributed pollen to the gene pool of *C. illinoensis* because these species would have lacked mates in newly colonized areas. Presumably, similar patterns would not be observed in more southern populations because these species would not have been mate-limited in refugial areas.

*Geographic origins of Mexican C. illinoensis populations* The origins of Mexican *C. illinoensis* populations have been debated for over 60 years (e.g., Manning 1949). Previous studies based on allozyme data (Rüter et al. 1999) and nuclear microsatellite data presented here demonstrate that Mexican pecan populations harbor lower levels of nuclear genetic variation than northern *C. illinoensis* populations, an observation that is consistent with a relatively recent introduction of the species into the

southern portions of its range. However, plastid microsatellite data reveal that Mexican *C. illinoensis* populations harbor more private alleles and higher levels of variation than other North American pecan populations. The plastid data suggest that Mexican populations are not the product of recent introductions, but may represent the descendants of relatively ancient *C. illinoensis* populations. If they represent refugia, the Mexican populations would be the source for more northern materials.

More than 50 plant species occupy disjunct distributions between the eastern USA and northeastern Mexico (Graham 1999), including *C. illinoensis*, *Carya myristiciformis* (nutmeg hickory), and *Carya ovata* (shagbark hickory). Two competing hypotheses have been proposed to explain the disjunction: (1) Mexican populations were introduced from the north during the late Cenozoic cooling (Miocene and Pliocene); and (2) Mexican populations were part of an ancient rainforest that has persisted since the early Eocene (Graham 1999). Milne (2006) identified some of the challenges associated with untangling vicariance from long-distance dispersal in resolving questions of long-standing plant disjunctions and noted that intraspecific haplotype variation could offer valuable insight, if examined with appropriate molecular tools. The use of molecular tools for dating phylogenies in broad contexts requires their development within groups that demonstrate known patterns of vicariance, documented by other hard evidence. The genus *Carya* provides opportunities for the development of such markers, due to its disjunct distribution between the southeastern USA and Asia that has been well-established with the fossil record (Manchester 1999). The use of plastid SSR markers in this collection reveals levels of intraspecific variation that is distinct from the pattern of nuclear variation and reveals structure related to geographic origin. Recognition of the structure of population diversity should contribute to the development of improved strategies for the conservation of that diversity (Namkoong et al. 2004), and to the development of more precise molecular tools to characterize those populations.

## References

- Abrams MD, Nowacki GJ (2008) Native Americans as active and passive promoters of mast and fruit trees in the eastern USA. *Holocene* 18:1123–1137
- Acosta MC, Premoli AC (2010) Evidence of chloroplast capture in South American *Nothofagus* (subgenus *Nothofagus*, Nothofagaceae). *Mol Phylogenet Evol* 54:235–242
- Amos W, Hoffman JI, Frodsham A, Zhang L, Best S, Hill AVS (2007) Automated binning of microsatellite alleles: problems and solutions. *Mol Ecol Notes* 7:10–14
- Braun EL (1961) Woody plants of Ohio. Ohio State University Press, Columbus, p 362

- Bryan GJ, McNicol J, Ramsay G, Meyer RC, Jong WSD (1999) Polymorphic simple sequence repeat markers in chloroplast genomes of Solanaceous plants. *Theor Appl Genet* 99:859–867
- Caicedo AL, Stinchcombe JR, Olsen KM, Schmitt J, Purugganan MD (2004) Epistatic interaction between *Arabidopsis* FRI and FLC flowering time genes generates a latitudinal cline in a life history trait. *Proc Nat Acad Sci USA* 101:15670–15675
- Charlesworth B, Charlesworth D, Barton NH (2003) The effects of genetic and geographic structure on neutral variation. *Ann Rev Ecol Evol Systemat* 34:99–125
- Cheng YJ, Meng HJ, Guo WW, Deng XX (2006) Universal chloroplast primer pairs for simple sequence repeat analysis in diverse genera of fruit crops. *The Journal of Horticultural Science and Biotechnology* 81:132–138
- Dangl GS, Woeste K, Aradhya MK, Koehmstedt A, Simon C, Potter D, Leslie C, McGranahan GH (2005) Characterization of 14 microsatellite markers for genetic analysis and cultivar identification of walnut. *The Journal of Horticultural Science and Biotechnology* 130:348–354
- Delcourt PA, Delcourt HR (1987) Long-term forest dynamics of the temperate zone: a case study of late-quaternary forests in Eastern North America. Springer, New York
- Dering JP (1977) Baker Cave plant macrofossil report. Unpublished manuscript on file with the Baker Cave Project Texas Archaeological Research Laboratory, University of Texas at Austin
- Ennos RA (1994) Estimating the relative rates of pollen and seed migration among plant populations. *Heredity* 72:250–259
- Fagan BM (1992) *People of the earth: an introduction to world prehistory*. Harper Collins Publishers, New York
- Flores D, Ford M (2009) Mexico tree nuts annual 2009. USDA FAS Global Agricultural Information Network (GAIN) Report Number MX9061
- Graham A (1999) The Tertiary history of the northern temperate element in the northern Latin American biota. *Am J Bot* 86:32–38
- Grauke LJ, Payne JA, Wood BW (1989) North American pecans: a provenance study. Annual Report of the Northern Nut Growers Association 80:124–131
- Grauke LJ, Iqbal MJ, Reddy AS, Thompson TE (2003) Development of microsatellite DNA markers in pecan. *Journal of the American Society for Horticultural Science* 128(3):374–380
- Harlan JR (1992) *Crops and man*. American Society of Agronomy, Inc., Crop Science Society of America, Inc. Madison, Wisconsin
- Harper RM (1928) *Economic botany of Alabama. Part 2. Catalogue of the trees, shrubs and vines of Alabama, with their economic properties and local distribution*. Geological survey of Alabama. University of Alabama
- Hester TR (1981) Late Paleo Indian occupations at Baker Cave, southwestern Texas. *Bull Tex Arch Soc* 53(101):119
- Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biol J Linn Soc* 58:247–276
- Hiraoka K, Tomaru N (2009) Population genetic structure of *Fagus japonica* revealed by nuclear microsatellite markers. *Int J Plant Sci* 170:748–758
- Hu XS, Ennos RA (1999) Impacts of seed and pollen flow on population differentiation for plant genomes with three contrasting modes of inheritance. *Genetics* 152:441–450
- Jump AS, Marchant R, Peñuelas J (2009) Environmental change and the option value of genetic diversity. *Trends Plant Sci* 14:51–58
- Latta RG (2006) Integrating patterns across multiple genetic markers to infer spatial processes. *Landscape Ecol* 21:809–820
- Magri D, Vendramin GG, Comps B, Dupanloup I, Geburek T, Gömöry D, Latalowa M, Litt T, Paule L, Roure JM, Tantau I, van der Knaap WO, Petit R, de Beaulieu J-L (2006) A new scenario of the quaternary history of European beech populations: palaeobotanical evidence and genetic consequences. *New Phytol* 171:199–221
- Manchester SR (1999) Biogeographical relationships of North American Tertiary floras. *Ann Mo Bot Gard* 86:472–522
- Manning WE (1949) The genus *Carya* in Mexico. *J Am Arb* 30:425–432
- Manning WE (1962) Additional notes on *Juglans* and *Carya* in Mexico and Central America. *Bul Torrey Bot Soc* 89(2):110–113
- Meloni M, Perini D, Binelli G (2007) The distribution of genetic variation in Norway spruce (*Picea abies* Karst.) populations in the western Alps. *J Biogeogr* 34:929–938
- Mendoza-Herrera, MA, Grauke LJ, Binzel M (2008) Confirmation of SSR markers within a structured collection of pecan [*Carya illinoensis* (Wangenh.) K. Koch], cultivars. Plant and Animal Genome XVI Conf., San Diego, CA. Jan. 12–16, 2008 (Abstract P140).
- Milne RI (2006) Northern hemisphere plant disjunctions: a window on tertiary land bridges and climate change? *Ann Bot* 98:465–472
- Miranda F, Sharp AJ (1950) Characteristics of the vegetation in certain temperate regions of eastern Mexico. *Ecology* 31:313–333
- Mitchell-Olds T, Willis JH, Goldstein DB (2007) Which evolutionary processes influence natural genetic variation for phenotypic traits? *Nat Rev Genet* 8:845–856
- Namkoong G, Gregorius HR (1985) Conditions for protected polymorphisms in subdivided plant populations. 2: seed versus pollen migration. *Am Nat* 125:521–534
- Namkoong G, Lewontin RC, Yanchuk AD (2004) Plant genetic resource management: the next investments in quantitative and qualitative genetics. *Genet Resour Crop Evol* 51:853–862
- Paterson AH, Brubaker C, Wendel JF (1993) A rapid method for extraction of cotton (*Gossypium* spp.) genomic DNA suitable for RFLP or PCR analysis. *Plant Mol Biol Rep* 11:122–127
- Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in excel population genetic software for teaching and research. *Mol Ecol Notes* 6:288–295
- Perrier X, Jacquemoud-Collet JP (2006) DARwin software. <http://darwin.cirad.fr/darwin>
- Petit RJ, Bodenes C, Ducousso A, Roussel G, Kremer A (2003) Hybridization as a mechanism of invasion in oaks. *New Phytologist* 161:151–164
- Pritchard JK, Wen X, Falush D (2007) Documentation for structure software: Version 2.2. <http://pritch.bsd.uchicago.edu/software>.
- Rüter B, Hamrick JL, Wood BW (1999) Genetic diversity within provenance and cultivar germplasm collections versus natural populations of pecan (*Carya illinoensis*). *J Hered* 90:521–528
- Rüter B, Hamrick JL, Wood BW (2000) Outcrossing rates and relatedness estimates in pecan (*Carya illinoensis*) populations. *J Hered* 91:72–75
- Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425
- Sloan DB, Barr CM, Olson MS, Keller SR, Taylor DR (2008) Evolutionary rate variation at multiple levels of biological organization in plant mitochondrial DNA. *Mol Biol Evol* 25:243–246
- Smith BD (1992) *Rivers of change: essays on early agriculture in eastern North America*. Smithsonian Institution Press, Washington, DC
- Soltis DE, Morris AB, McLachlan JS, Manos PS, Soltis PS (2006) Comparative phylogeography of unglaciated eastern North America. *Mol Ecol* 15:4261–4293
- Stone DE (1962) Affinities of a Mexican endemic, *Carya palmeri*, with American and Asian hickories. *Am J Bot* 49:199–212
- Styles BW, Ahler SR, Fowler ML (1983) Modoc rock shelter revisited. In: Phillips JL, Brown JA (eds) *Archaic hunters and gatherers in the American midwest*. Academic, New York, NY, pp 261–297

- Taylor WA (1905) Promising New Fruits. In: Hill GW (ed) Yearbook of the United States Department of Agriculture, 1904. Government Printing Office, Washington DC, pp 399–416
- Thompson TE, Grauke LJ (1991) Pecans and hickories (*Carya*). In: Moore JN, Ballington JR (eds) Genetic resources of temperate fruit and nut crops. International Society of Horticultural Sciences, Wageningen, pp 839–904
- Thompson TE, Romberg LD (1985) Inheritance of heterodichogamy in pecan. *J Hered* 76:456–458
- True RH (1919) Notes on the early history of the pecan in America. *Annu Rep Smithsonian Inst* 1917:435–448
- Tsitrone A, Kirkpatrick M, Levin DA (2003) A model for chloroplast capture. *Evolution* 57:1776–1782
- Vander Wall SB (2001) The evolutionary ecology of nut dispersal. *Bot Rev* 67:74–117
- Weising K, Gardner RC (1999) A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. *Genome* 42:9–19
- Welch PD, Scarry CM (1995) Status-related variation in foodways in the Moundville chiefdom. *Am Antiquity* 60:397–419
- Woeste K, Burns R, Rhodes O, Michler C (2002) Thirty polymorphic nuclear microsatellite loci from black walnut. *J Hered* 93:58–60
- Wood BW, Grauke LJ, Payne JA (1998) Provenance variation in pecan. *Journal American Society for Horticultural Science* 123:1023–1028