

# Discrimination and assessment of black walnut (*Juglans nigra* L.) cultivars using phenology and microsatellite markers (SSRs)

Peng Zhao, Huijuan Zhou, Mark Coggeshall, Bill Reid, and Keith Woeste

**Abstract:** Black walnut (*Juglans nigra* L.), a large tree native throughout the eastern United States, produces a high-quality edible nut. Our goal was to maintain the integrity of black walnut breeding programs by verifying the identity of accessions. We sampled 285 ramets of 78 cultivars from the black walnut nut breeding orchards and clonal repositories at the University of Missouri and Kansas State University. We employed both phenotypic and genotypic methods to identify and differentiate cultivars. Phenotypes were evaluated using seven phenological traits. Cultivars varied for all traits among each of the 4 yr, but the best morphological characteristics for evaluating cultivar identity were bud break date and date of first pistillate bloom. Samples ( $n = 285$ ) were genotyped using 10 polymorphic microsatellite loci. The simple sequence repeats produced a total of 174 alleles and 17.2 alleles per locus. We detected 47 unique genotypes represented by more than one sample, including 128 instances of identical genotypes with different names (*synonyms*) and 106 instances of different genotypes with a shared name (*homonyms*). Our results indicated that multiple errors were committed during the propagation of these important cultivars. It may be difficult to determine which genotype is original to a cultivar name in the absence of a foundation plant materials collection or vouchered specimens. These results will assist black walnut breeders and producers by improving the integrity of breeding collections and by identifying the best phenological traits for rapid assessment of trueness to type.

**Key words:** nut breeding, trueness-to-type, genotyping, synonyms, homonyms.

**Résumé :** Le noyer noir (*Juglans nigra* L.) est un arbre indigène de grande taille qui pousse un peu partout dans l'est des États-Unis et produit une noix comestible de grande qualité. Les auteurs voulaient préserver l'intégrité des programmes d'hybridation de cet arbre en vérifiant l'identité des obtentions. Dans cette optique, ils ont échantillonné 285 ramets de 78 cultivars découverts dans les vergers d'hybridation du noyer noir et les dépôts de clones de l'Université du Missouri et de l'Université d'État du Kansas. Les chercheurs ont recouru au phénotype et au génotype pour identifier et différencier les cultivars. Le phénotype a été évalué à partir de sept caractères. Les caractères ont varié chez tous les cultivars, chacune des quatre années de l'étude. Les caractères morphologiques permettant d'identifier le mieux la variété sont la date du débourrement et la date de la première floraison pistillée. On a établi le génotype des échantillons ( $n = 285$ ) d'après dix microsatellites polymorphiques. En tout, les SSR ont produit 174 allèles, soit 17,2 allèles par locus. Quarante-sept génotypes uniques ont été décelés et étaient représentés par plus d'un échantillon, y compris 128 génotypes identiques mais portant des noms différents (*synonymes*) et 106 génotypes différents arborant le même nom (*homonymes*). Les résultats indiquent que de nombreuses erreurs ont été commises lors de la multiplication de ces importants cultivars. Faute de matériel végétal de fondation ou de spécimens confirmés, il pourrait s'avérer difficile de relier le génotype d'origine au nom d'un cultivar. Les résultats de cette étude aideront les améliorateurs et les producteurs en rehaussant l'intégrité des banques de matériel génétique et en indiquant les caractères phénologiques permettant le mieux d'établir rapidement la pureté génétique. [Traduit par la Rédaction]

**Mots-clés :** hybridation des arbres à noix, pureté génétique, génotypage, synonymes, homonymes.

Received 6 July 2017. Accepted 29 September 2017.

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

**M. Coggeshall and K. Woeste.** USDA Forest Service Hardwood Tree Improvement and Regeneration Center (HTIRC), Department of Forestry and Natural Resources, Purdue University, 715 West State Street, West Lafayette, IN 47907, USA.

**B. Reid.** Pecan Experiment Field, 8960 SW 90th Street, Chetopa, KS 67336, USA.

**Corresponding author:** Keith Woeste (email: [kwoeste@fs.fed.us](mailto:kwoeste@fs.fed.us)).

This work is free of all copyright and may be freely built upon, enhanced, and reused for any lawful purpose without restriction under copyright or database law. The work is made available under the [Creative Commons CC0 1.0 Universal Public Domain Dedication](https://creativecommons.org/licenses/by/4.0/) (CC0 1.0).

## Introduction

Black walnut (*Juglans nigra* L.), commonly known as eastern black walnut or American walnut, is native throughout the eastern United States from Massachusetts to Florida and west to Minnesota and Texas (Fowells 1965). It produces valuable timber, a high-quality edible nut, and is attractive to wildlife (Woeste and Michler 2011a, 2011b). Black walnut is an under-exploited crop that has been cultivated for at least 130 yr in the United States (Heiges 1896; Reid et al. 2004). Traits important for improved black walnut nut production include lateral bud fruitfulness, late leafing, resistance to anthracnose caused by *Gnomonia leptostyla* (Fr.) Ces. & Not., precocity, productivity, and improved nut quality (Reid 1990; Reid et al. 2004).

The cultivar ‘Thomas’, the earliest recorded clonal black walnut selected for nut production, has been propagated since 1881 (Heiges 1896; Reid et al. 2004; Michler et al. 2007). Since then, at least 700 nut cultivars have been named and recorded, mostly by small-scale, private nut growers and amateur breeders (Berhow 1962). For the most part, these cultivars were selected from wild populations solely on the basis of nut quality characteristics or as chance offspring of older cultivars such as ‘Thomas’ or ‘Ohio’ (Woeste 2004). The potential for genetic improvement in nut yield, percent kernel, and kernel quality is great (Coggeshall and Woeste 2009, 2010; Coggeshall 2011). A black walnut nut breeding program was initiated by the University of Missouri Center for Agroforestry in 1996 (Coggeshall and Woeste 2010). The black walnut germplasm collection at the University of Missouri currently contains 142 nut tree accessions used for breeding. To maintain the integrity of the breeding program, it was important to verify the identity of the accessions in this collection (Coggeshall and Woeste 2010) and a separate clonal repository at Kansas State University.

Phenological traits have many advantages as tools for cultivar discrimination; many are readily observable in the field and have high heritability. They also have the benefit of being biologically and (often) economically meaningful. Differences in phenology are particularly useful to black walnut breeders because although black walnut has high genetic diversity across its wide native range (Victory et al. 2006), its vegetative morphology is surprisingly homogeneous. It is not possible to tell most black walnut cultivars apart based on their vegetative form. Simple sequence repeat (SSR) markers, which are highly reproducible, multiallelic, and codominant (Varshney et al. 2005; Vilanova et al. 2012), can be used to identify important clones (Woeste et al. 2002; Foroni et al. 2005; Robichaud et al. 2006) and for breeding (Pollegioni et al. 2009). The advantages of using SSR markers for the detection of genetic variation and cultivar identification have been reported in other plant species (Zhang et al. 2012; Fajardo et al. 2013). The utilization

of both phenotypic and genotypic data is expected to provide the best method to select and cluster cultivars, clones, and seedlings in a crop breeding program (Fatahi et al. 2010; Ebrahimi et al. 2011).

The objective of this study was to identify and characterize the most important black walnut nut cultivars using both microsatellite markers and phenological traits.

## Materials and Methods

### Plant materials

In 1996, a total of 84 named black walnut cultivars, four ramets per clone, were planted at the Horticulture and Agroforestry Research Center (HARC), the University of Missouri in New Franklin, MO (39°01'N, 92°74'W) (Coggeshall 2002). Samples from ramets of 78 of these cultivars were collected from HARC by M. Coggeshall, and samples of 85 ramets of (many of the same) cultivars were collected from Kansas State University by B. Reid. These were genotyped using 10 nuclear microsatellite markers (Table 1).

### DNA extraction

Leaves were collected, placed in a plastic bag, and stored on ice until they were mailed to the Hardwood Tree Improvement and Regeneration Center (HTIRC, Lafayette, IN). DNA was extracted from leaf samples using the methods of Zhao and Woeste (2011) and stored at –80 °C. The concentration of DNA in the samples was quantified by measuring absorbance at 260 nm using a NanoDrop-8000 spectrophotometer (Thermo Fisher Scientific, Wilmington, DE).

### Microsatellite analysis

#### Screening of primer pairs

Primer pairs ( $n = 10$ ) designed for this research were derived by further sequencing a black walnut microsatellite library described by Woeste et al. (2002), Dangl et al. (2005) and Victory et al. (2006) (Table 2). The forward primer was labeled with NED, HEX, or 6-FAM fluorescent tag. We separated the primer pairs into four sets for multiplexing (Table 2).

#### PCR amplification

Genomic DNA was diluted to 100 ng  $\mu\text{L}^{-1}$  in TE buffer for storage. DNA was subsequently diluted to approximately 10 ng  $\mu\text{L}^{-1}$  with water for polymerase chain reaction (PCR) amplification. Polymerase chain reaction was performed in a total volume of 10  $\mu\text{L}$  containing 1  $\mu\text{L}$  of 10 ng  $\mu\text{L}^{-1}$  DNA, 1  $\mu\text{L}$  of 10 $\times$  *Taq* DNA polymerase reaction buffer [(100 mmol  $\text{L}^{-1}$  Tris-HCl, pH 8.8, 15 mmol  $\text{L}^{-1}$   $\text{MgCl}_2$ , 500 mmol  $\text{L}^{-1}$  KCl, and 0.01% (w/v) gelatin), Stratagene, La Jolla, CA], 1.25  $\mu\text{L}$  of 200 mmol  $\text{L}^{-1}$  dNTP, 1  $\mu\text{L}$  of 0.1 mg  $\text{mL}^{-1}$  BSA (bovine serum albumin, acetylated) (Promega, Madison, WI), 0.5  $\mu\text{L}$  of 10 mmol  $\text{L}^{-1}$  each primer, 4.5  $\mu\text{L}$  sterilized distilled water, and 0.25  $\mu\text{L}$  *Taq* polymerase (Promega or New England Biolabs,

**Table 1.** Black walnut (*Juglans nigra*) cultivars genotyped using 10 simple sequence repeat (SSR) loci.

Label <sup>a</sup>	Genotype <sup>b</sup>	N <sup>c</sup>
Beck	Singleton	1,1,0
Bowser	Bowser	9,5,4
Brown Nugget	Brown Nugget	5,1,4
Christianson	Christianson	3,1,2
Clermont	Hare	2,2,0
Clermont	Thomas	6,6,0
Clermont	Elmer Meyer2	6,2,4
Cochrane	Cochrane	2,2,0
Cooksey	Singleton <sup>d</sup>	1,1,0
Cranz	Cranz	4,3,1
Cranz	Sparrow	1,0,1
Crosby	Crosby	2,2,0
Cutleaf	Cutleaf	2,2,0
D.O.T	Singleton	1,0,1
Daniel	Singleton	1,0,1
Daniel	Sauber 1	2,2,0
Davidson	Davidson	8,8,0
Drake	Eldora	3,0,3
Dubois	Dubois	2,0,2
Dubois	Singleton	2,1,1
Eldora	Eldora	4,3,1
Elmer Meyer1	Elmer Meyer1	3,3,0
Elmer Meyer2	Elmer Meyer2	5,5,0
Emma K	Emma K	6,3,3
Farrington	Cranz	3,0,3
Football	Football	6,2,4
Grundy	Singleton	1,1,0
Hare	Hare	4,4,0
Harney	Singleton	1,1,0
Hay	Singleton	1,1,0
Hay	Kwik Krop	1,0,1
Hay	Christianson	2,2,0
Higbee	Singleton	1,0,1
Hybrid Walnut	Singleton	1,0,1
Jackson	Jackson	2,2,0
Kitty	Singleton	1,1,0
Knuvean	Knuvean	2,2,0
Krause	Krause	3,3,0
Kwik Krop	Kwik Krop	6,4,2
M-18 (K-07)	Singleton	1,0,1
McGinnis	McGinnis	5,2,3
Mintel	Brown Nugget	3,0,3
Mintel	Singleton	1,1,0
Mystery	Mystery	4,4,0
Neel	Neel	2,2,0
Ness	Ness	2,2,0
Rowher	Thomas	2,2,0
Ogden	Ogden	2,2,0
Ohio	Ohio	3,2,1
OK Selection	Singleton	1,1,0
Ozark King	Ozark King	2,2,0
Patterson	Emma K	2,2,0
Philops	Philops	2,2,0
Pound	Singleton	1,1,0
Pritchett	Singleton	1,1,0
Purdue137	Purdue137	2,2,0

**Table 1 (concluded).**

Label <sup>a</sup>	Genotype <sup>b</sup>	N <sup>c</sup>
Purdue 41	Purdue 41	2,2,0
R2T26	Cutleaf	2,2,0
R4T21	Singleton	2,2,0
R4T24	Singleton	1,1,0
Ridgeway	Ridgeway	2,2,0
Rupert	Rupert	5,3,2
Russell1	Singleton	2,2,0
Russell3	Russell 3	3,3,0
Russell3	Singleton	1,1,0
Sarcozie	Kwik Krop	2,2,0
Sauber1	Sauber1	3,3,0
Sauber1	Singleton	1,1,0
Sauber2	Eldora	3,3,0
Schesler	Schesler	4,4,0
Scrimger	Eldora	6,2,4
Shreve	Shreve	2,2,0
South Fork	South Fork	2,2,0
Sparks 127	Sparks 127	5,2,3
Sparks 129	Singleton	6,6,0
Sparks 147	Sparks 147	7,2,5
Sparks 177	Christianson	3,0,3
Sparrow	Eldora	6,4,2
Sparrow	Sparrow	1,0,1
Stabler	Stabler	2,0,2
Stabler	Singleton	1,0,1
Stambaugh	Thomas	8,8,0
Surprise	Surprise	11,6,5
Thatcher	Thomas	5,5,0
Thomas	Thomas	11,10,1
Thomas	Christianson	2,2,0
Thomas	Singleton	1,1,0
Tomboy	Tomboy	2,2,0
Vandersloot	Vandersloot	5,0,5
Victoria	Victoria	3,0,3
Victoria	Hare	2,2,0
Wiard	Wiard	2,2,0
Woody	Kwik Krop	2,2,0
Woody	Davidson	1,0,1
Total no. of samples	282	

<sup>a</sup>Labeled cultivar name.<sup>b</sup>Cultivar based on multilocus genotype.<sup>c</sup>N, number of entries per cultivar. Obvious spelling variants were pooled; the first number indicates the total numbers of samples with the given labeled name assigned to the indicated genotype, the second number indicates the number of samples from ramets of that name from Horticulture and Agroforestry Research Center (HARC), and the third number indicates the number of samples from ramets of that name collected from Kansas State University's clonal repository.<sup>d</sup>Singleton refers to a genotype shared by no other sample and is thus either unverifiable if there are no other samples of the same cultivar name (e.g., 'Grundy'), or the result of a graft failure or mislabeling (e.g., Russell3).

**Table 2.** Characteristics of the 10 microsatellite loci used in this study.

Locus	GeneBank <sup>d</sup>	Alleles (bp)	$T_a^e$	$N^f$	$N_A^g$	$NE^{-1h}$
WGA06 <sup>a</sup>	AY333949	140, 142, 144, 146, 150, 152, 154, 156, 158, 160, 162	53.5	250	11	0.275
WGA24 <sup>a</sup>	AY333950	224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252	50.0	273	15	0.034
WGA27 <sup>a</sup>	AY333951	202, 204, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242	50.0	273	20	0.016
WGA32 <sup>a</sup>	AY333952	164, 166, 168, 172, 174, 176, 180, 182, 184, 186, 188, 192, 194, 196, 198, 202, 214, 216	53.5	265	18	0.018
WGA69 <sup>a</sup>	AY333953	162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182	50.0	260	11	0.192
WGA72 <sup>a</sup>	AY333954	144, 146, 148, 150, 152, 154, 156, 158, 160, 164	53.5	247	10	0.189
WGA76 <sup>c</sup>	AY636615	228, 230, 232, 234, 236, 238, 240, 242, 244, 260	50.0	281	10	0.055
WGA82 <sup>a</sup>	AY333956	154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 214	50.0	249	29	0.009
WGA89 <sup>b</sup>	AY352440	183, 185, 187, 189, 191, 195, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223	50.0	280	19	0.022
WGA90 <sup>b</sup>	AY352441	142, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176	50.0	279	17	0.022

<sup>a</sup>Locus previously published in [Woeste et al. \(2002\)](#).

<sup>b</sup>Locus previously published in [Robichaud et al. \(2006\)](#).

<sup>c</sup>WGA76 was not previously published, primer sequences were F: AGGGCACTCCCTTATGAGGT, R: CAGTCTCATTCCCTTTTCC.

<sup>d</sup>GeneBank accession number or identification number.

<sup>e</sup> $T_a$  indicates the annealing temperature of multiplex polymerase chain reaction (PCR).

<sup>f</sup> $N$  indicates number of genotyped sample.

<sup>g</sup> $N_A$  indicates mean number of alleles per locus.

<sup>h</sup> $NE^{-1}$  average nonexclusion probability for identity of two unrelated individuals.

Beverly, MA). Thermal cycling conditions were as follows: denaturation 3 min at 94 °C; 32 cycles of 15 s at 93 °C, 1 min at the annealing temperature for the primer  $T_a$  (Table 2), 30 s at 72 °C; and a final extension of 10 min at 72 °C at the end of the amplification.

### Genotyping

After PCR amplification, amplicons were diluted 1:10 in water and 1 µL of the diluted PCR product was combined with a master mix of 13.4 µL formamide (Invitrogen, Carlsbad, CA) and 0.6 µL Rox fluorescent size standard. The resulting solution was heated to 95 °C for 5 min in a thermal cycler, snap cooled, and sent to the Agricultural Genomics laboratory at Purdue University for analysis using an ABI 3100 capillary sequencer.

### Data analysis

Microsatellite allele sizes were determined from ABI 3100 data using Genemapper<sup>®</sup> software version 4.0 (Applied Biosystem, Foster City, CA). Controls (two positive and one negative) were included with each 96-well PCR plate to ensure accurate allele size scoring. Failed reactions were repeated. To standardize the data for producing the neighbor-joining (NJ) tree and principle components analysis (see below), all allele sizes were recalculated as deviations from the mean allele size at each locus. Original allele sizes (in bp) are shown in Table 2.

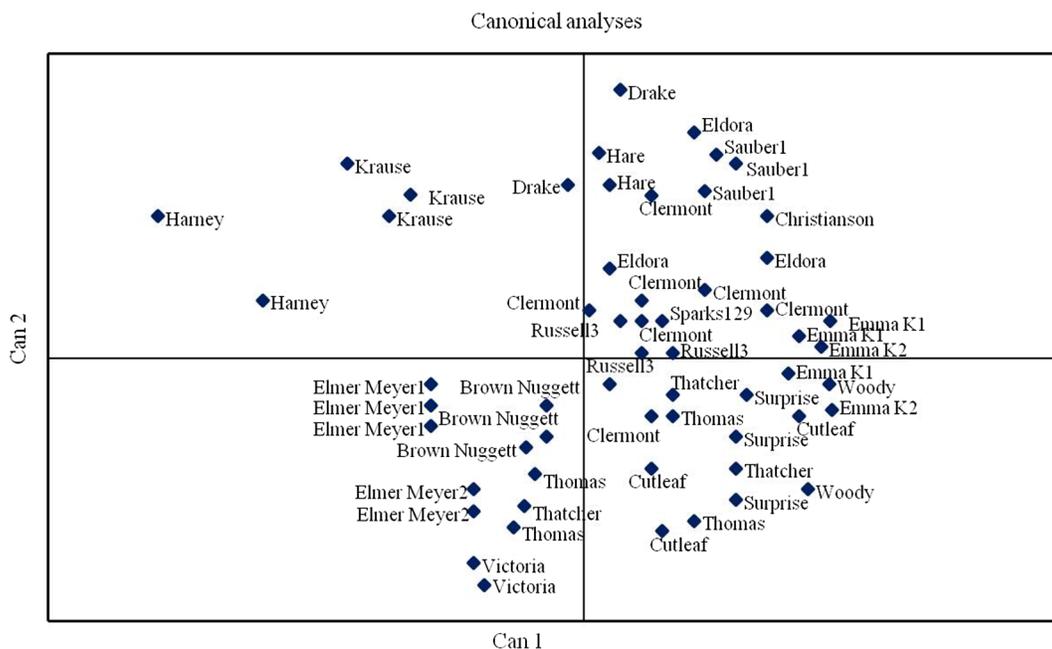
CERVUS3.0 and PopGen1.32 ([Marshall et al. 1998](#); [Kalinowski et al. 2007](#)) calculated the percentage of polymorphic loci (95% criterion; P), mean number of alleles per locus ( $N_A$ ), and polymorphic information content

(PIC). To accept two genotypes as identical, we required a perfect match of allele sizes at all loci. SAS software version 9.1 (SAS Institute, Inc., Cary, NC) was used for canonical discriminant analysis. The CANDISC procedure univariate test statistics and  $F$  test were performed with SAS software. The first two canonical variables for each cultivar were plotted to show the patterns of discrimination among cultivars (Fig. 1). CANDISC accepts missing data, and although within-class distributions should be approximately normal, it can be used descriptively with non-normal data ([SAS Institute 1990](#)). The genetic analysis was similar to the methods described in [Zhao et al. \(2013\)](#). Dendrograms were constructed based on microsatellite genotypes using NJ in the software PHYLIP version 3.5c ([Felsenstein 1995](#)) and the dendrogram was drawn with TreeView version 1.5 ([Page 1996](#)). The overall genetic structure of the sampled trees was investigated using principal components analysis (PCA) and the software GenALEX 6.5 ([Peakall and Smouse 2012](#)).

### Phenological measurements

From 2003 to 2006, ramets of 21 black walnut cultivars growing at HARC were selected and evaluated for seven phenological traits over four consecutive growing seasons (Table 3). 'Davidson', the earliest leafing cultivar, was used as a reference standard. Thus, all cultivars were characterized for bud break days after 'Davidson' (BBDAD), first pistillate bloom (Julian days, FPLD), first pollen shed, days after 'Davidson' (FPD), season length in days (LD), harvest date, days after 'Davidson' (HDAD), peak pistillate bloom days after 'Davidson' (PPBDAD),

**Fig. 1.** Plot of black walnut clones onto first two canonical variables composed of linear combinations of phenological traits. Repeated names indicate observations from the same tree in multiple years.



and peak pollen shed days after ‘Davidson’ (PPDAD) (Table 3). The method’s details were described by Coggeshall and Woeste (2010). The phenological data were used to distinguish 21 black walnut cultivars using canonical correlation analysis (SAS Institute, Inc.); canonical variables are orthogonal linear combinations of the phenological traits, similar to principal components. Although multiple ramets (grafted copies) were rated for each cultivar, there was missing data for many ramets in some years; as a result, 17.5% of the data was missing.

## Results

### Analysis of phenology data

Phenological traits among the 21 cultivars observed over four consecutive growing seasons varied from a minimum of 26 d (bud break) to a maximum of 47 d (season length) (Table 3). Bud break ranged from 0 to 26 d after ‘Davidson’, FPLD ranged from Julian day 108 to 135, HDAD ranged from 0 to 43 d, PPBDAD date ranged from 0.0 to 23.0 d, and PPDAD ranged from –11.0 to 22.0 d. All 21 cultivars were protogynous, meaning FPLD date was earlier than FPD date for a given tree. First pistillate bloom ranged from 109 to 135 Julian days; the latest flowering cultivars were ‘Elmer Meyer2’ and ‘Hare’ at 128–135 d; the FPLD dates for ‘Drake’ ranged from day 117 to 130, which resulted in the largest standard deviation in all cultivars, although the range for ‘Thatcher’, which turned out to be a synonym of ‘Thomas’, was actually larger (20 d). Season length, defined as the period from FPLD receptivity to date of harvest, ranged from 113 to 160 d; the latest flowering cultivar (‘Hare’) also had the longest season length, from

153 to 160 d; the longest flowering time was observed in cultivar ‘Kwik Krop’ (18 d); the cultivars with the shortest pistillate flowering length were ‘Elmer Meyer1’ and ‘Emma K2’ (2 d) (Table 3).

Univariate tests run as part of canonical discriminant analysis showed BBDAD, FPLD, FPD, LD, PPBDAD, and PPDAD were highly significant for discriminating difference among cultivars when the data were assigned to trees based on labeled cultivar name (Table 4). These traits should be useful for distinguishing among clones despite intraclonal and interannual variability in phenology. HDDAD was the only trait that was not reliable for distinguishing among this set of cultivars (Table 4).

When ramets evaluated in the phenology study were reassigned to their correct clone based on genotype data, we re-ran the analysis and found that, as before, all traits but HDDAD were useful for distinguishing clone identity. We plotted the phenology data for each of the 21 cultivars on a biplot of the two best canonical variables. The biplot showed that despite year-to-year variation, in general, cultivars with disparate phenologies (e.g., ‘Eldora’ and ‘Brown Nugget’) could be identified, as could entries with similar phenology (‘Surprise’ and ‘Thatcher’). ‘Drake’, ‘Krause’, ‘Harney’, and ‘Victoria’ were separated from other cultivars, indicating their phenology was different than the majority of the cultivars (Fig. 1).

### Analysis of genotypic data

#### Polymorphism and genetic diversity

To maintain the integrity of the black walnut nut breeding program, it was important to verify the identity of accessions in both germplasm collections. Ten

**Table 3.** Range of phenological trait means observed from 2003 to 2006 to characterize 21 black walnut (*Juglans nigra* L.) nut cultivars at the University of Missouri nut breeding orchard.

Labeled cultivar	Genotype	BBDAD		FPLD		FPD		LD		HDAD		PPBDAD		PPDAD	
		Day	SD	Julian Day	SD	Julian Day	SD	Julian Day	SD	Day	SD	Day	SD	Day	SD
Brown Nugget	Brown Nugget	2–5	1.5	112–115	1.7	122–130	3.9	140–144	2.8	0–15	10.6	0–6	3.0	3–9	3.6
Christianson	Thomas Meyer	15–19	2.8	ND	ND	ND	ND	ND	ND	21–26	3.5	ND	ND	ND	ND
Clermont(L)1	Clermont	14–21	3.2	124–128	1.9	134–139	2.2	127–144	8.4	9–21	11.0	9–15	2.8	12–16	2.1
Clermont(L)2	Hare	13–23	4.7	124–129	2.5	139–143	2.6	136–153	9.3	9–39	15.1	12–18	3.0	11–5	2.5
Cutleaf	Cutleaf	9–20	5.0	124–130	3.2	139–140	0.6	ND	ND	ND	ND	13–19	3.1	13–21	3.9
Drake	Drake	8–23	6.4	117–130	7.2	135–136	0.7	147–155	5.7	28–41	9.2	12–18	3.3	12–16	2.5
Eldora	Sparrow	13–16	2.5	122–132	5.3	126–138	5.2	113–117	2.5	0–1	0.7	7–20	6.5	8–13	2.1
Elmer Meyer1	Elmer Meyer1	6–18	4.8	124–126	1.2	138–139	0.7	135–137	3.5	0–15	8.4	8–12	2.3	10–15	3.2
Elmer Meyer2	Elmer Meyer2	18–21	3.6	128–135	3.5	143–145	1.0	124–127	2.5	14–15	0.7	17–23	2.8	18–22	2.1
Emma K1	Emma K	3–16	6.4	112–116	2.2	125–128	1.7	145–150	3.7	6–21	7.5	2–3	0.8	–2 to 3	2.5
Emma K2	Emma K	4–14	5.2	109–116	3.8	120–126	2.6	142–144	1.4	7–43	18.1	2–3	0.6	–5 to 2	2.8
Hare	Hare	13–26	5.4	128–135	3.8	139–143	2.0	ND	ND	ND	ND	16–21	2.3	14–21	3.3
Krause	Krause	3–4	0.5	120–128	3.3	109–115	2.9	128–135	4.6	8–14	4.2	10–12	1.0	–11 to 8	3.4
Russell3	Sparks127	1–4	1.5	ND	ND	114–123	4.2	ND	ND	ND	ND	ND	ND	–6 to 2	1.7
Sauber1	Sauber1	11–24	5.9	122–129	3.1	134–139	3.0	136–143	3.3	8–27	9.7	9–18	4.6	9–15	3.2
Sparks129	Sparks129	11–23	5.3	ND	ND	ND	ND	ND	ND	21–39	12.7	ND	ND	ND	ND
Surprise	Surprise	10–12	1.0	122–125	1.3	133–136	1.5	137–147	7.1	21–29	5.7	8–15	2.8	8–5	2.9
Thatcher	Thomas	12–26	6.0	112–132	4.3	137–144	3.1	145–155	4.8	12–43	17.3	7–23	6.8	13–20	4.0
Thomas	Thomas	15–26	5.1	124–130	3.1	137–143	3.0	140–156	8.5	8–43	17.6	8–19	5.0	11–19	3.7
Victoria	Hare	8–23	5.3	123–131	3.6	132–138	2.6	153–160	4.6	41–43	1.4	7–21	6.2	8–16	3.7
Woody	Kwik Krop	0–4	1.8	108–115	3.6	121–130	4.0	141–159	12.7	ND	ND	0–4	2.0	2–4	1.2

**Note:** BBDAD, bud break days after ‘Davidson’; FPLD, first pistillate bloom; FPD, first pollen shed; LD, season length in Julian days; HDAD, harvest date, days after ‘Davidson’; PPBDAD, peak pistillate bloom date, days after ‘Davidson’; PPDAD, peak pollen shed days after ‘Davidson’; Day, number of days after reference cultivar ‘Davidson’; SD, standard deviation; ND, no data.

**Table 4.** The CANDISC procedure univariate test statistics and *F* statistics using phenotypic data based on labeled cultivar name.

Variable <sup>a</sup>	Total SD	Pooled SD	Between SD	R <sup>2</sup>	R <sup>2</sup> /(1-R <sup>2</sup> )	F value	Pr > F
BBDAD	7.5748	5.3920	6.4496	0.6960	2.2892	3.43	0.0054**
FPLD	6.7842	3.6615	6.2900	0.8252	4.7218	7.08	<0.0001***
FPD	6.3425	4.5826	5.9835	0.7123	2.5218	3.68	<0.0001***
LD	10.8459	5.3270	10.2372	0.8553	5.9091	8.86	<0.0001***
HDDAD	13.5045	11.5158	10.3482	0.5637	1.2920	1.94	0.0827
PPBDAD	6.5000	4.6775	5.5078	0.6893	2.2185	3.33	0.0064**
PPDAD	8.9455	3.4055	8.7239	0.9130	10.4997	15.75	<0.0001***

**Note:** Asterisks (\*\*, \*\*\*) indicate highly or very highly significant ( $p < 0.01$ ,  $p < 0.0001$ ), respectively. SD, standard deviation; BBDAD, bud break days after 'Davidson'; FPLD, first pistillate bloom; FPD, first pollen shed; LD, season length in Julian days; HDDAD, harvest date, days after 'Davidson'; PPBDAD, peak pistillate bloom date, days after 'Davidson'; PPDAD, peak pollen shed days after 'Davidson'.

<sup>a</sup>Variable names are defined in Table 3.

microsatellite markers were used to characterize 285 samples with 78 different labeled names (Table 1). The SSRs produced a total of 174 alleles. The nonexclusion probability (i.e., the probability that any two genotypes in the sample would not be differentiated based on 172 alleles of the 10 SSR loci described) was  $2.6 \times 10^{-14}$ . Thus, the high polymorphism levels detected (17.2 alleles per locus) and low-average nonexclusion probability of each locus (range from 0.009 to 0.275) indicated that these microsatellite markers were suitable for the genetic "fingerprinting" of black walnut cultivars and assessing clonal heterogeneity.

#### Cultivar assignment and clustering of cultivars

After genotyping all samples, we identified 47 black walnut genotypes that were shared by at least two samples, although in many cases, samples that shared a genotype did not come from ramets that had the same labeled name. So, for example, the genotype we associated with 'Brown Nugget' was shared by eight samples, five from ramets labeled 'Brown Nugget', and three from ramets labeled 'Mintle'. Each sample's new or corrected name based on shared genotype was added to Table 1. In addition, there were 32 samples that had unique genotypes (shared by no other samples); these were designated 'singleton' in Table 1. Thus, many ramets that shared cultivar names did not share genotypes (homonyms), and conversely, some genotypes had multiple cultivar names (synonyms). An additional example will illustrate the point. Some ramets of 'Victoria', 'Hare', and 'Clermont' had identical genotypes. Other ramets of 'Clermont' had the same genotype as 'Thomas', 'Stambaugh', and 'Rowher'.

Of the samples with unique genotypes (indicated as singleton in Table 1), 14 were from ramets bearing a name (often a cultivar name) for which no other samples were available. These included samples of 'Beck', 'Cooksey', 'D.O.T', 'Grundy', 'Harney', 'Higbee', 'Hybrid Walnut', 'Kitty', 'M-18(K-07)', 'OK Selection', 'Pound', 'Pritchett', 'R4T21', and 'R4T24'. These samples were

examples of a single ramet with a unique name associated with a unique genotype. The remaining 18 samples placed in the singleton category came from ramets with a label that identified them as a cultivar. Despite their label, however, this group of singletons had genotypes that were not shared by any other ramets with the same cultivar name. Examples of this type included samples of 'Daniel' (one sample), 'Dubois' (two samples), 'Hay' (one sample), 'Intel' (one sample), 'Russell1' (two samples), 'Russell3' (one sample), 'Sauber1' (one sample), 'Sparks' (six samples), 'Stabler' (two samples), and 'Thomas' (one sample) (Table 1). Each of these 18 samples represents a homonym (one name representing multiple genotypes).

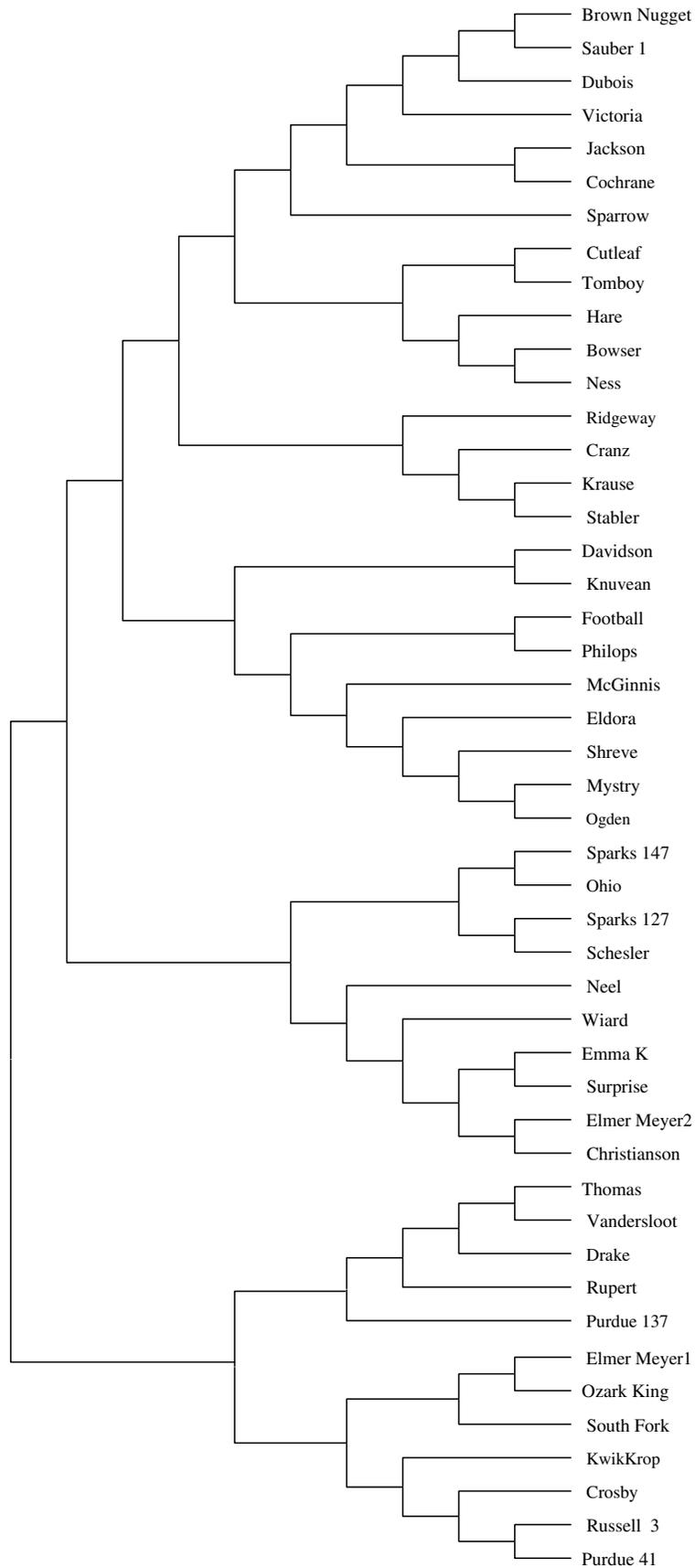
We performed PCA based on the 47 genotypes within 253 samples. It showed that 65.5% of the variance for SSR alleles was explained by the first two axes (PC1 = 43.2% + PC2 = 23.2%) (Fig. 2). The results of PCA showed two main clusters of genotypes, one high on the y axis and the other low, with other genotypes scattered between. Three of the most important cultivars, 'Emma K', 'Kwik Krop', and 'Sparrow' showed little variation on the y axis but were spread across nearly the entire x axis.

In total, 128 samples were grouped into 12 genotypes (cultivars) that contained samples with different label names (synonyms). We observed that 106 of the samples shared a labeled name with at least one other sample although their genotypes differed (homonyms). Samples from ramets with a unique name and unique genotype might appear to indicate a "true" genotype/cultivar name association, but caution is warranted because there may be other trees bearing these names with genotypes that do not match those in either the Kansas or Missouri collections sampled for this study. In all, 253 samples (88.8% of the 285 total) had a labeled name and a genotype that matched at least one other sample with the same name and genotype (Table 1).

After genotyping and cultivar assignment, 47 genotypes were clustered using NJ (Fig. 3). The resulting dendrogram showed three clearly defined clusters of genotypes (i.e., 'Thomas' to 'Purdue 41', 'Sparks 147' to



**Fig. 3.** Dendrogram (neighbour-joining) of 47 black walnut genotypes based on 10 microsatellite loci.



*Juglans* (protandry versus protogyny) is regulated by a single gene (so even closely-related germplasm can be segregating for this trait). Within protandrous and protogynous morphs, there is wide phenological variability. There is no published data relating genetic and phenological variance in walnut, but the authors have observed that half-sibs can have highly divergent phenology.

The biplot of canonical ordinates (Fig. 1) reflects phenological similarities within and among clones over the years. The relative positions of cultivars on the biplot do reflect their phenotypic relatedness, but not their genotypic relatedness, as is shown in the NJ tree (Fig. 3). Trees may have similar phenology but be unrelated, so it is not expected that the biplot of canonical ordinates would array the cultivars similar to a genotypic clustering method such as NJ. Figure 1 does, however, show the relative phenotypic similarity of clones. Clones that plot near to each other are difficult to distinguish based on phenology, while those that are distant are easy to separate based on phenology. Knowledge of the centroid and the variance for common cultivars permits breeders to determine which cultivars can be visually distinguished and which cannot. In the end, no phenotypic trait will ever have the discriminatory power of a genotype, as all phenotypes are affected by environment. The effect of year on phenology in black walnut is also shown in Fig. 1.

Phenology and genetic distance are not correlated in black walnut, so we did not expect Fig. 1 to correspond with the NJ tree (Fig. 3) except in cases of synonymy. When the phenological traits were integrated into canonical variables and the observations were plotted onto the biplot, clonal phenology could be distinguished, regardless of year-to-year variation in the weather. This approach to cultivar discrimination could be practical for black walnut, but a large data set including more data from more clones would be needed if the goal was to determine clonal identities with confidence based solely on differences in phenology.

The PCA and NJ analyses have different goals and they are based on different analytical methods, so PCA and NJ analyses do not always produce results that appear compatible. The reason is PCA identifies axes that are linear combinations of traits (i.e., SSR alleles) that maximize the variance among entries. Its goal is to separate entries as much as possible. If we associate position in the PCA plot with relatedness, then we are assuming that entries with the greatest differences in allele sizes are the least related—which may not be correct. Neighbor-joining clusters entries based on shared alleles, so its goal is to join genotypes that are similar. If we associate position in the PCA plot with relatedness, then we are assuming that allele sharing is equivalent to relatedness, which is often correct but does not account for homoplasy.

There are no published pedigrees in black walnut—nearly all trees listed in Table 1 and Fig. 3 are selections

from the wild and they likely reflect a nearly random sample of the wild black walnut germplasm. In most cases, clusters of genotypes in the NJ tree (Fig. 3) do not reflect (recent) common ancestors, but NJ trees of this type are useful because in some cases cultivars that are joined very closely may in fact be synonyms or related as chance seedlings arising as (putative) offspring of a cultivar. For example, ‘Sparks#147’ and ‘Sparks#127’ were both developed by the same hobbyist, possibly using ‘Ohio’ as a parent.

The number of detected alleles for each SSR locus ranged from 11 to 29, with an average of 17.20 (Table 2), about the same as that found by Zhao et al. (2013) in 20 black walnut progenies from a timber breeding program (17.92 alleles per locus). Thus, nut cultivars showed high levels of neutral genetic diversity, levels similar to those of wild trees. The high levels of polymorphism observed for the SSRs in this collection of clones resulted in high power to discriminate among cultivars and closely-related clones. The PCA represents the genetic structure of the samples, whereas the NJ tree shows their relatedness. We also used the PCA to help identify examples of potential homonyms and synonyms that could later be verified by direct comparison of genotypic data. The current analysis yielded 47 unique black walnut genotypes, some of which, based on NJ, clustered near the foundational selections ‘Ohio’, ‘Thomas’, ‘Emma K’, and ‘Elmer Myer1’, possibly reflecting pedigree.

## Conclusion

Simple sequence repeat genotyping was an efficient tool for cultivar genetic fingerprinting, the identification of off-types, and for assessing inter- and intra-cultivar variation. The 253 black walnut trees sampled and genotyped yielded consensus genotypes for at least 26 black walnut cultivars (‘Thomas’, ‘Christianson’, ‘Elmer Meyer2’, ‘Eldora’, ‘Hare’, ‘Kwik Krop’, ‘Cranz’, ‘Emma K’, ‘Brown Nugget’, ‘Davidson’, ‘Surprise’, ‘Bowser’, ‘Sparks147’, ‘Sparks127’, ‘Sauber’, ‘McGinnis’, ‘Vandersloot’, ‘Rupert’, ‘Victoria’, ‘Mystery’, ‘Cutleaf’, ‘Drake’, ‘Russel3’, ‘Krause’, ‘Ohio’, and ‘Schesler’). We observed two major types of inconsistencies: (i) identical genotypes with different names (synonyms) involving about 116 samples, representing 40.7% of all entries, and (ii) different genotypes that shared the same cultivar name (homonyms), observed for about 101 samples or 35.43% of all samples. The most likely reasons for these results were graft failure, misidentification of a tree from an orchard map, or labeling error during propagation or sampling. Errors such as these may have been compounded over time, as the vast majority of the scion wood from which the collections at HARC and Kansas State University were propagated was obtained from private growers rather than a formal breeding program, and thus likely reflects past labeling or propagation errors. In some cases, cultivars from one repository (Kansas or Missouri) had a consistent genotype that was

different from the genotype of the same cultivars from the other site (Table 1). Errors in propagation at one or more locations do not mean that cultivars cannot be distinguished, but errors within and among clone banks make it challenging to arrive at a consensus about the true genotype of a cultivar. These black walnut SSR genotypes will be useful in black walnut breeding programs and for others who wish to understand the identity of material they have propagated or purchased. Growers of black walnuts and black walnut researchers will benefit from being able to verify the identity of the trees in their orchards.

## Acknowledgements

We wish to thank C. Michler, M. Kremer, H. Bergeman, and W. Johnson for their technical assistance, J. McKenna and B. Beheler for sample collection, and J. Gosewehr for office support. Mention of a trademark, proprietary product, or vendor does not constitute a guarantee or warranty of the product by the US Department of Agriculture (USDA) and does not imply its approval to the exclusion of other products or vendors that also may be suitable. We wish to thank the support of the National Natural Science Foundation of China (Grant Nos. 31200500 and 41471038) and the Program for Excellent Young Academic Backbones funding by Northwest University. We also thank USDA Agricultural Research Service for contributing to the Missouri University black walnut breeding program and the maintenance of collections there. Genotypes were generated at the Purdue University Genomics Core Facility.

## References

- Berhow, S. 1962. Black walnut varieties—a reference list. Northern Nut Growers Association Annual Report 53: 63–69.
- Coggeshall, M.V. 2002. Black walnut cultivar improvement program at the University of Missouri. Annu. Rep. North. Nutr. Growth Assoc. 93: 93–96.
- Coggeshall, M.V. 2011. Black walnut: a nut crop for the midwestern United States. HortScience, 46: 340–342.
- Coggeshall, M.V., and Woeste, K.E. 2009. Microsatellite and phenological descriptors identify eastern black walnut cultivars in Missouri, USA. In International Symposium on Molecular Markers in Horticulture 859. pp. 93–98.
- Coggeshall, M.V., and Woeste, K.E. 2010. The use of microsatellite and phenological descriptors to identify a collection of eastern black walnut cultivars in Missouri, USA. International Symposium on Molecular Markers in Horticultural Species. Corvallis, OR. pp. 93–98.
- Dangl, G.S., Woeste, K., Aradhya, M.K., Koehmstedt, A., Simon, C., Potter, D., Leslie, C.A., and McGranahan, G. 2005. Characterization of 14 microsatellite markers for genetic analysis and cultivar identification of walnut. J. Am. Soc. Hort. Sci. 130: 348–354.
- Ebrahimi, A., Fatahi, R., and Zamani, Z. 2011. Analysis of genetic diversity among some Persian walnut genotypes (*Juglans regia* L.) using morphological traits and SSRs markers. Sci. Hortic. 130: 146–151. doi:10.1016/j.scienta.2011.06.028.
- Fajardo, D., Morales, J., Zhu, H., Steffan, S., Harbut, R., Bassil, N., and Zalapa, J. 2013. Discrimination of American cranberry cultivars and assessment of clonal heterogeneity using microsatellite markers. Plant Mol. Biol. Rep. 31: 264–271. doi:10.1007/s11105-012-0497-4.
- Fatahi, R., Ebrahimi, A., and Zamani, Z. 2010. Characterization of some Iranians and foreign walnut genotypes using morphological traits and RAPD markers. Hortic. Environ. Biotechnol. 51: 51–60.
- Felsenstein, J. 1995. PHYLIP (Phylogeny Interference Package) version 3, 5 c. Department of Genetics, University of Washington, Seattle, WA.
- Faroni, I., Rao, R., Woeste, K., and Gallitelli, M. 2005. Characterization of *Juglans regia* L. with SSR markers and the evaluation of genetic relationships among cultivars and the 'Sorrento' landrace. J. Hortic. Sci. Biotech. 80: 49–53. doi:10.1080/14620316.2005.11511890.
- Fowells, H.A. 1965. Silvics of forest trees of the United States. Agriculture Handbook No. 271. U.S. Department of Agriculture (USDA), Washington, DC.
- Gleeson, S.K. 1982. Heterodichogamy in walnuts: inheritance and stable ratios. Evolution, 36(5): 892–902. doi:10.1111/j.1558-5646.1982.tb05461.x. PMID:28567826.
- Heiges, S.B. 1896. Nut culture in the United States. USDA Gov. Printing Office, Washington, DC.
- Kalinowski, S.T., Taper, M.L., and Marshall, T.C. 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol. Ecol. 16: 1099–1106. doi:10.1111/j.1365-294X.2007.03089.x. PMID:17305863.
- Kostamo, K., Toljamo, A., Antonius, K., Kokko, H., and Kärenlampi, S.O. 2013. Morphological and molecular identification to secure cultivar maintenance and management of self-sterile *Rubus arcticus*. Ann. Bot. 111: 713–721. doi:10.1093/aob/mct029. PMID:23456688.
- Marshall, T.C., Slate, J., Kruuk, L.E.B., and Pemberton, J.M. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. Mol. Ecol. 7: 639–655. doi:10.1046/j.1365-294x.1998.00374.x. PMID:9633105.
- Michler, C.H., Woeste, K.E., and Pijut, P.M. 2007. Black walnut. Pages 189–198 in C. Kole, ed. Forest trees. Genome mapping and molecular breeding in plants series. Vol. 7. Springer-Verlag, Berlin, Germany.
- Page, R.D.M. 1996. TREEVIEW: an application to display phylogenetic trees on personal computers. Comp. Appl. Biol. Sci. 12: 357–358. doi:10.1093/bioinformatics/12.4.357. PMID:8902363.
- Peakall, R., and Smouse, P.E. 2012. GenALEX 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. Bioinformatics, 28: 2537–2539. doi:10.1093/bioinformatics/bts460. PMID:22820204.
- Pollegioni, P., Woeste, K., Mugnozza, G.S., and Malvolti, M.E. 2009. Retrospective identification of hybridogenic walnut plants by SSR fingerprinting and parentage analysis. Mol. Breed. 24: 321–335. doi:10.1007/s11032-009-9294-7.
- Ramos, D.E., ed. 1997. Walnut production manual. Vol. 3373. UCANR Publications, Novato, CA. 320 p.
- Reid, W. 1990. Eastern black walnut: potential for commercial nut producing cultivars. Pages 327–331 in J. Janick and J.E. Simon, eds. Advances in new crops. Timber, Portland, OR.
- Reid, W., Coggeshall, M.V., and Hunt, K.L. 2004. Cultivar evaluation and development for black walnut orchards. Pages 18–24 in C.H. Michler, P.M. Pijut, J. Van Sambeek, M. Coggeshall, J. Seifert, K. Woeste, and R. Overton, eds. Black walnut in a new century. Proc. 6th Walnut Council Research Symposium, Gen. Tech. Rep. NC-243, U.S. Dept. Agric. For. Serv. North. Central Res. Stn, St. Paul, MN.

- Robichaud, R.L., Glaubitz, J.C., Rhodes, O.E., Jr., and Woeste, K. 2006. A robust set of black walnut microsatellites for parentage and clonal identification. *New Forest*, **32**: 179–196. doi:[10.1007/s11056-005-5961-7](https://doi.org/10.1007/s11056-005-5961-7).
- SAS Institute. 1990. SAS/STAT 9.2 user's guide. SAS Institute Inc., Cary, NC.
- Varshney, R.K., Graner, A., and Sorrells, M.E. 2005. Genic microsatellite markers in plants: features and applications. *Trends Biotechnol.* **23**: 48–55. doi:[10.1016/j.tibtech.2004.11.005](https://doi.org/10.1016/j.tibtech.2004.11.005). PMID:[15629858](https://pubmed.ncbi.nlm.nih.gov/15629858/).
- Victory, E.R., Glaubitz, J.C., Rhodes, O.E., and Woeste, K.E. 2006. Genetic homogeneity in *Juglans nigra* (Juglandaceae) at nuclear microsatellites. *Am. J. Bot.* **93**: 118–126. doi:[10.3732/ajb.93.1.118](https://doi.org/10.3732/ajb.93.1.118).
- Vilanova, S., Manzur, J.P., and Prohens, J. 2012. Development and characterization of genomic simple sequence repeat markers in eggplant and their application to the study of diversity and relationships in a collection of different cultivar types and origins. *Mol. Breed.* **30**: 647–660. doi:[10.1007/s11032-011-9650-2](https://doi.org/10.1007/s11032-011-9650-2).
- Woeste, K. 2004. An online database of *Juglans* cultivar names and origins. *HortScience*, **39**: 1771.
- Woeste, K., and Michler, C. 2011a. *Juglans*. Pages 77–87 in C. Kole, ed. *Wild crop relatives: genomic and breeding resources, forest trees*. Springer-Verlag, Berlin, Germany.
- Woeste, K., and Michler, C. 2011b. Black walnut and its wild relatives. Pages 77–87 in C. Kole, ed. *Wild crop relatives—genomic and breeding resources. Forest Trees*. Springer-Verlag, Berlin, Germany.
- Woeste, K., Burns, R., Rhodes, O., and Michler, C. 2002. Thirty polymorphic nuclear microsatellite loci from black walnut. *J. Heredity*, **93**: 58–60. doi:[10.1093/jhered/93.1.58](https://doi.org/10.1093/jhered/93.1.58). PMID:[12011179](https://pubmed.ncbi.nlm.nih.gov/12011179/).
- Zhao, P., and Woeste, K. 2011. DNA markers identify hybrids between butternut (*Juglans cinerea* L.) and Japanese walnut (*Juglans ailantifolia* Carr.). *Tree Genet. & Genom.* **7**: 511–533. doi:[10.1007/s11295-010-0352-4](https://doi.org/10.1007/s11295-010-0352-4).
- Zhang, Q., Li, J., Zhao, Y., Korban, S.S., and Han, Y. 2012. Evaluation of genetic diversity in Chinese wild apple species along with apple cultivars using SSR markers. *Plant Mol. Biol. Rep.* **30**: 539–546. doi:[10.1007/s11105-011-0366-6](https://doi.org/10.1007/s11105-011-0366-6).
- Zhao, P., Zhang, S.X., and Woeste, K. 2013. Genotypic data changes family assignment and family rank for growth and quality traits in a black walnut (*Juglans nigra* L.) progeny test. *New Forest*, **44**: 357–368. doi:[10.1007/s11056-012-9343-7](https://doi.org/10.1007/s11056-012-9343-7).