Low Correlation Between Genomic and Morphological Introgession Estimates in a Walnut Backcross

Keith Woeste¹ and Gale McGranahan
Department of Pomology, Wickson Hall, University of California, Davis, CA 95616

Robert Bernatzky
Department of Plant and Soil Science, French Hall, University of Massachusetts, Amherst, MA 01003

ADDITIONAL INDEX WORDS. Juglans regia, Juglans hindsi, Paradox, Mahalanobis' distance

ABSTRACT. A first backcross population of walnuts ([Juglans hindsi (Jeps.) Jeps. x Juglans regia L.] x J. regia) was used to evaluate the correlation between morphological (statistical) and genetic distance during introgression. Five traits based on leaf morphology were identified to quantify the morphology of the parental species, their F1 hybrids, and the backcrosses to each parent. These traits were used to evaluate the morphological similarity of first backcrosses to J. regia using Mahalanobis’ distance. The amount of genomic introgression of each backcross was estimated using 59 randomly amplified polymorphic DNA (RAPD) and 41 restriction fragment-length polymorphism (RFLP) genetic markers that identify polymorphisms between J. regia and J. hindsi. A smaller scaffold set of markers was also identified using published linkage data. The correlation between the measures of morphological and genomic introgression for the first backcrosses was low (0.23) but significant. The results suggest that selection based on morphology during backcrossing will not be an effective way to recover recurrent parent genome.

The morphological evaluation of populations has been used by geneticists and taxonomists to characterize introgression. The idea that morphological intermediacy in wild populations implies genetic introgression was introduced by Anderson (1949) and Fisher (1936). Since their path-breaking work, a variety of statistical distance measures have been used to study introgression, evaluate hybrid complexes, infer parentage or group classification, and estimate genetic distance (Broehman, 1987; Ietswaart and Feij, 1989; Namkoong, 1966; Yang and Selander, 1968; Zeyl and Lowcock, 1989). Mahalanobis’ distance (MD) (1936), a multivariate method for evaluating distance, is particularly appropriate for analyzing introgression because it removes the correlations among traits. It is the preferred method of analysis when the variance and covariance structures of the populations to be evaluated are unequal (Atchley et al., 1982a; Beer et al., 1993). Plant breeders have used the same distance measures to identify and classify germplasm and evaluate its genetic variability (Baum and Bailey, 1984; Goodman, 1968; Pereira-Lorenzo et al., 1996; Souza and Sorrels, 1991; Smith and Smith, 1989a). Plant breeders rarely use statistical distance measures to evaluate introgression, since replicating observations of many traits is expensive and the goal of most introgression programs is to backcross as quickly as possible.

Ecologists and plant breeders have also used genetic markers such as isozymes, randomly amplified polymorphic DNA (RAPD) markers, and restriction fragment-length polymorphisms (RFLPs) to evaluate genetic distance and predict the potential effects of combining gene pools (Dudley, 1994; Cowen and Frey, 1987; Kesseli et al., 1991; Moser and Lee, 1994; Smith and Smith, 1989b; Williams and St. Clair, 1993). Openshaw et al. (1994), Tanksley et al. (1981), and others have proposed that plant breeders could use molecular markers to reduce linkage drag and increase the efficiency of introgression. One method, outlined by Hillel et al. (1990) and Hospital et al. (1992), uses genetic markers to monitor genome introgression by selecting against the entire donor genome except a particular trait of interest. This approach has been called genomic selection (Hillel et al., 1990).

The Persian walnut (Juglans regia L.), and the Northern California black walnut (Juglans hindsi (Jeps.) Jeps.) are morphologically and genetically distinct, being members of different subsections of the genus Juglans. In this paper we apply a statistical distance measure (MD) and genomic selection to a first backcross to evaluate the introgression of Northern California black walnut into the cultivated Persian walnut. Because walnuts are long-lived trees with a long generation time, selection during backcrossing is potentially useful. In this paper we present data that indicates that statistical and genetic distance measures provide only weakly concordant estimates of the amount of introgression among first backcross trees.

Materials and Methods

CHOOSING TRAITS FOR MORPHOLOGICAL EVALUATION. We used the following criteria for choosing morphological traits for analysis of introgression: 1) they are easy to measure, 2) they have apparent quantitative inheritance, 3) the parental species must differ for the trait without overlap in range, 4) hybrids should be intermediate between their parents (Namkoong, 1966), 5) the traits should be vegetative and present in juvenile plants grown from seed to help evaluate heritability and early selection, and 6) the traits must not have been the target of selection. Identifying traits that meet these criteria was difficult; after the preliminary evaluation of >10 traits we identified 5 that could be reliably and easily scored: the ratio of the subterminal leaflet length/subterminal leaflet width (SL/SW), the ratio of the rachis length/number of leaflets (RL/N), the ratio of the basal leaflet length/basal leaflet width (BL/BW), the ratio of the subterminal leaflet length/rachis length (SW/RL), and the ratio of the petiole length/rachis length (PL/RL).
Table 1. Walnut and walnut hybrid populations used to evaluate interpopulation morphological variability.

<table>
<thead>
<tr>
<th>Population</th>
<th>Abbreviation</th>
<th>Location (Calif.)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Juglans regia</td>
<td>JR</td>
<td>Davis, Winters</td>
<td>29 genotypes, including old and recent cultivars, breeding lines and germplasm</td>
</tr>
<tr>
<td>Persian walnut</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Juglans hindsi</td>
<td>JH</td>
<td>Davis, Winters</td>
<td>12 germplasm accessions and 26 wild genotypes</td>
</tr>
<tr>
<td>F₁ interspecific hybrids (JH x JR)</td>
<td>F₁</td>
<td>Davis</td>
<td>15 genotypes of unknown parentage collected as naturally occurring hybrids and selected for fruitfulness or as rootstocks</td>
</tr>
<tr>
<td>First backcross [JH x JR] x JR</td>
<td>BC₁</td>
<td>Davis</td>
<td>Open-pollinated progeny of a single F₁, selected for fecundity in crosses with J. regia</td>
</tr>
<tr>
<td>First backcross [JH x JR] x JH</td>
<td>BCH</td>
<td>Davis</td>
<td>15 genotypes from seed of an open-pollinated F₁ selected for fecundity in crosses with J. hindsi</td>
</tr>
</tbody>
</table>

**PLANT POPULATIONS AND MORPHOLOGICAL DATA COLLECTION.** Mature J. regia, J. hindsi, their F₁ hybrids [Paradox (any F₁ hybrid between J. regia and J. hindsi is designated Paradox (McGranahan, personal communication))], BC₁, (F₁ x J. regia), BCH (backcrosses to J. hindsi, F₁ x J. hindsi), and juvenile second backcrosses to J. regia (BC₂) were characterized using the traits described above (Table 1). BC₁ was comprised of 84 random seedlings of a single, open-pollinated F₁ hybrid individual. All such F₁ hybrids are male sterile (Woeste et al., 1998). The plants described in this study were maintained by the Dept. of Pomology at the Univ. of California, Davis, according to current recommendations (Ramos, 1985: University of California Statewide Integrated Pest Management Project, 1982).

**MEASUREMENT METHOD.** On mature trees, morphological traits were measured to the nearest millimeter on randomly chosen, healthy appearing, fully expanded leaves that were exposed to full sun for at least part of the day. Leaves on shoots with actively growing tips were not measured. On juvenile plants, observations were made on healthy appearing, fully expanded leaves; usually the oldest leaves on the shoot for which precise measurements could be made. In all cases each observation is the mean of three measurements taken from three leaves per individual at the same time. Measured leaflets were always on the same side of the (pinnate compound) leaf. An effort was made to measure leaves from different sides of the canopy. Mature genotypes were evaluated for 2 to 3 years.

**ESTIMATES OF GENETIC PARAMETERS.** Estimates of the repeatability of the morphological traits were based on mature BC₁ trees evaluated in 1990–93. Neither all individuals nor all variables were evaluated in every year. To determine the significance of genotypic effects for the morphological traits, 18 randomly selected genotypes from each population were measured from 1990 through 1993. Repeatabilities and standard errors were calculated using the method of Becker (1985).

Estimates of the narrow-sense heritability (h²) of the morphological traits were obtained using a regression of offspring on biparental means (Falconer, 1981). BC₁ progeny were in 54 full-sib families composed of an average of 3.3 seedlings per family. BC₂ was produced from crosses of J. regia ‘Chandler’ (used as a male) to the members of the BC₁, which was a mixture of half-sibs and full-sibs all derived from a single open-pollinated F₁. Biparental means were based on single observations and progeny values were means of single observations of all the progeny produced by a cross. An adjustment was made for differences in parent and offspring grand means by using deviations from trait means as dependent and independent variables (Fernando and Gianola, 1988). Standard errors for heritability were estimated as in Falconer (1981). All observations for heritability estimates were log₁₀ transformed before analysis. The offspring and parents were grown in separate but similarly situated blocks in the Univ. of California, Davis, pomology fields.

**MAHALANOBSİS DISTANCE.** The statistical distance between two populations i and j was calculated as Mahalanobis’ generalized distance Eq. [1]:

\[
MD_{ij} = (x_i - x_j)^T S^{-1} (x_i - x_j) / \alpha P
\]  

[1]

where P was the number of variables (5), \( x_i \) was the vector of character means for the reference population i (J. regia), and \( S^{-1} \) was the estimate of the common within population phenotypic covariance matrix for mature J. regia germplasm. The MD of any individual from the reference population was calculated as an individual’s mean value for each of the traits during all the years data was collected. For some genotypes this mean was composed of many observations over several years. When between-population comparisons were made, MDs were log₁₀ transformed to make them conform more closely to the assumptions of ANOVA (Fernandez, 1992).

**GENETIC MARKERS AND GENOMIC INTROGRESSION.** Genomic introgression was evaluated in 76 BC₁, by using 59 RAPD and 41 RFLP genetic markers that identify J. regia and/or J. hindsi loci polymorphic in the BC₁ (Woeste et al., 1996; Fjellstrom and Parfitt, 1994). The RFLP loci were identified by Fjellstrom (1993), and segregated as codominant markers in the BC₁. The RAPD markers were fixed (present) in J. hindsi but absent from J. regia (fixed null). The RAPDs segregated as present (in the heterozygous state) or absent in the BC₁. The markers are in 14 linkage groups covering about 1000 cM of the estimated 1600 cM genetic map of walnut. Only markers that did not deviate significantly from the expected 1:1 segregation in the BC₁ were used in the analysis (Woeste et al., 1996; Fjellstrom, 1993). The percent recipient parent genome (PRPG) of each BC₁ (Eq. [2]) was calculated as the proportion of the BC₁ genome derived entirely from J. regia (0.5) plus the product of the proportion of the genome segregating for J. regia and J. hindsi alleles (0.5) and the proportion of the loci homozygous for J. regia expressed as a percent.

\[
PRPG = \left(0.5 + \left(0.5 \times \text{no. homozygous loci/no. loci observed}\right)\right) \times 100
\]  

[2]

A data set of 100 markers and a subset of 50 RAPD and RFLP markers spaced ≤20 cM from the next marker were used to calculate PRPG. We refer to the set of 50 markers as a “scaffold” set. Selection on markers is theoretically more efficient when evenly spaced marker sets are used (Hospital et al., 1992). Not all markers were scored in every BC₁; an average of 67 markers per genotype was scored when the full data set was used, and 33.0 markers per genotype were scored when the scaffold data set was used.

**STATISTICAL ANALYSES.** The data were analyzed using SAS

Results

There were highly significant differences between *J. regia* and *J. hindsii* for the morphological traits for both mature and juvenile plants (in all cases *P* ≤ 0.0001). The morphology of *F*₁ hybrids was intermediate for each trait, and the backcross means were intermediate between the *F*₁ and the parental species, although the population ranges overlapped extensively (data not shown). As the amount of *J. regia* genome in hybrids increased, leaflets were shorter and wider (more oblong and less lanceolate); there was greater space between leaflets, the ratio of leaflet width to leaf length increased and the distance from the base of the rachis to the first leaflet increased as a proportion of the total leaf length.

Variance for SL/SW, BL/BW, SW/RL and PL/RL was significant both among and within populations (Table 2). There was significant variance for RL/N only among populations. For SL/SW, SW/RL and PL/RL more than half of the variance was among populations, reflecting the ease with which mature members of these populations can be visually distinguished. The *BC₁* population was significantly different from *J. regia* and the *F*₁ hybrids for all traits but BL/BW. The proportion of variance within populations due to differences among genotypes varied from 9% to 47%. Repeating measures over years as we did for this analysis does not lead to strong conclusions since the variance for genotypes within populations was completely confounded with genotype x year interactions. Confounding of this type may have been significant for some of the populations but was not found to be so in an analysis that included only *J. regia* cultivars (data not shown).

We used estimates of genetic parameters to verify that the traits used to evaluate morphological introgression had a genetic basis. These parameters estimate the extent to which hybrid morphology for the traits represents genomic mixing and not simply random environmental effects. repeatability (R) is the upper bound of the broad-sense heritability (Falconer, 1981). The repeatability of all the traits was moderately high in the *BC₁*, ranging from 0.53 for RL/N to 0.76 for BL/BW (Table 3). During the years of evaluation slightly over half of the differences among BC₁ genotypes for each of the traits were attributable to genotypic differences.

Narrow-sense heritability (h²) is an estimate of the proportion of the total variance due to additive genetic effects. We estimated the narrow-sense heritabilities of the five morphological traits by regression of offspring values on biparental means (Table 3). The estimates were significant for SL/SW and SW/RL. Very high heritabilities were found for other characters in walnut based on plot means (Hansche et al., 1972). Heritabilities of 0.3 or 0.4 are commonly observed for morphological traits (Cheverud, 1988). Our estimates of the genetic parameters (Table 3) were certainly biased by the genetic structure of the populations used in the analysis, probably leading to overestimation of the parameters. For example, the estimate of the heritability of the trait SW/RL was larger than unity, the theoretical upper bound for heritability. This inaccuracy may be due to the small size of the populations used to generate the estimate, aberrant segregation in the backcrosses or selection during the breeding process. All of these factors were likely present in some *BC₁* families since many of the female parents (*BC₃*)s) had very low fertility and produced offspring with poor vigor. The heritability estimates were biased by several other factors as well: because we used a single male parent ("Chandler") and a *BC₁* population derived from a single, open-pollinated female, the variance of the parents (the denominator in the heritability estimate) was reduced. Since the male parent(s) was presum-

Table 2. ANOVA for population effects for the morphological traits used to calculate Mahalanobis' distance, the distribution of the variance among sources of variation, and mean separation of the populations.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Source¹</th>
<th>df</th>
<th>Mean square</th>
<th>F</th>
<th>Variance (%)</th>
<th>Mean separation</th>
</tr>
</thead>
<tbody>
<tr>
<td>SL/SW</td>
<td>Population</td>
<td>2</td>
<td>0.1505</td>
<td>22.5***</td>
<td>56.7</td>
<td>* **</td>
</tr>
<tr>
<td></td>
<td>Genotype (population)</td>
<td>18</td>
<td>0.0067</td>
<td>5.15***</td>
<td>31.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>63</td>
<td>0.0013</td>
<td>12.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RL/N</td>
<td>Population</td>
<td>2</td>
<td>0.0524</td>
<td>9.57***</td>
<td>31.5</td>
<td>* **</td>
</tr>
<tr>
<td></td>
<td>Genotype (population)</td>
<td>18</td>
<td>0.0055</td>
<td>1.77</td>
<td>14.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>63</td>
<td>0.0031</td>
<td>33.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BL/BW</td>
<td>Population</td>
<td>2</td>
<td>0.0255</td>
<td>4.55**</td>
<td>26.1</td>
<td>NS **</td>
</tr>
<tr>
<td></td>
<td>Genotype (population)</td>
<td>18</td>
<td>0.0056</td>
<td>4.67**</td>
<td>47.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>25</td>
<td>0.0012</td>
<td>26.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SW/RL</td>
<td>Population</td>
<td>2</td>
<td>0.0907</td>
<td>47.9***</td>
<td>77.6</td>
<td>*** ***</td>
</tr>
<tr>
<td></td>
<td>Genotype (population)</td>
<td>18</td>
<td>0.0189</td>
<td>1-7***</td>
<td>10.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>63</td>
<td>0.0046</td>
<td>11.5</td>
<td>*** ***</td>
<td></td>
</tr>
<tr>
<td>PL/RL</td>
<td>Population</td>
<td>2</td>
<td>0.1143</td>
<td>42.3***</td>
<td>72.4</td>
<td>*** ***</td>
</tr>
<tr>
<td></td>
<td>Genotype (population)</td>
<td>18</td>
<td>0.0027</td>
<td>2.69*</td>
<td>9.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>44</td>
<td>0.0010</td>
<td>18.6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

¹See text for descriptions of the morphological traits.
²Populations were *Juglans regia* (JR), *J. regia* x *J. hindsii* (*F*₁), and the first backcross to *J. regia* (*BC₁*).
³Genotypes are nested within population.
⁴,, , ⁵Non-significant or significant at *P* = 0.05, 0.01, 0.001, respectively. Population mean separation was by LSD.

Table 3. Heritability and repeatability estimates of the morphological traits used to calculate Mahalanobis' distance.

<table>
<thead>
<tr>
<th>Trait</th>
<th>h² ± SE</th>
<th>R ± SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>SL/SW</td>
<td>0.45 ± 0.19</td>
<td>0.628 ± 0.044</td>
</tr>
<tr>
<td>RL/N</td>
<td>0.11 ± 0.19</td>
<td>0.530 ± 0.051</td>
</tr>
<tr>
<td>PI/RL</td>
<td>0.15 ± 0.19</td>
<td>0.589 ± 0.047</td>
</tr>
<tr>
<td>SW/RL</td>
<td>1.25 ± 0.19***</td>
<td>0.595 ± 0.054</td>
</tr>
<tr>
<td>BL/BW</td>
<td>0.41 ± 0.19</td>
<td>0.759 ± 0.116</td>
</tr>
</tbody>
</table>

*See text for description of the traits.
**Based on 54 BC1 families.
***Significant at P = 0.05 or 0.001, respectively.

ably heterozygous for genes affecting morphology, the covariance of the offspring and parents (the numerator) was probably decreased. The relative importance of these biases in the estimates is not known.

MAHALANOBIS' DISTANCE AS A MEASURE OF INTROGRESSION. We used MD to create a single composite morphological score based on the five traits described above. MD was calculated as the morphological similarity of walnut genotypes or populations to the mean J. regia morphotype and is thus the statistical measure of morphological introgression. The BC1, F1, and BCH hybrid populations represent a range of mixing of the J. hindsii and J. regia genomes; assuming normal segregation they contain 75%, 50%, and 25% J. regia genome, respectively. The more introgressed a population, the smaller its mean MD from J. regia (Table 4). As measured by MD, J. regia, the F1, hybrids and J. hindsii are morphologically distinct, without overlap in range. J. regia was the most variable for MD and J. hindsii the least variable, but all the populations were significantly distinct morphologically. The distribution of the MD of the BC1 is roughly bell-shaped with short tails and a skew toward the male (J. regia) parent (Fig. 1). The median and mean MD of the BC1 were 1.35 and 1.34, respectively, after log transformation, with a standard deviation of 0.27. The most extreme MD values in the BC1 overlapped with the range of MD of the parental types (J. regia and the F1 hybrids). Thus, the majority of the BC1 were morphologically intermediate between J. regia and the F1s, but some of the BC1 were morphologically indistinguishable from the parental types. Some of the variability of the BC1 was environmental, since the mean MD of the BC1 was significantly different in 1992 and 1993 based on Student's t-test (P = 0.013). Despite this variability, there were significant differences for MD among BC1 genotypes (P ≤ 0.001) and the rank correlation between the MD of 30 mature BC1 genotypes in 1992 and 1993 was 0.718 (P ≤ 0.0001). The MD had a repeatability of (R = 0.42 ± 0.21) based on observations of 30 BC1 genotypes in 2 years.

The J. regia population was the most variable for MD as measured by cv, and J. hindsii the least variable (Table 4). At least two factors probably contributed to the large difference in cv between the two parental populations: a diverse J. regia germplasm was used in the analysis whereas the germplasm base for J. hindsii is quite narrow, and the variables which comprise the MD may better describe the morphological variation within J. regia than J. hindsii. The variability for MD within the F1 was quite low. The genetic variability within these hybrids is not well characterized and it is possible the F1s used in the study are related since they were naturally occurring hybrids and only a handful of genotypes of J. regia are widely planted in California. Moreover it is not known if all or only a subset of J. hindsii and J. regia genotypes can hybridize.

PERCENT RECIPIENT PARENT GENOME. The percent recipient parent genome (PRPG) is an estimate of genomic introgression based on the segregation of genetic markers derived from J. hindsii and J. regia in the BC1. The PRPG of the BC1 ranges from a high of 87.1% recipient parent genome to a low of 65% with a mean of 74% (Fig. 2) approximating the theoretically expected distribution (Hillel et al., 1990). Although the distribution appears roughly symmetric, there were more heterozygous than homozygous loci in the BC1 (1297 versus 1189) and the difference was significant when tested against an expected 1:1 segregation (χ² = 4.6, P ≤ 0.05) (Woeste, 1996). This difference skewed the PRPG slightly toward the F1 (female) parent. The standard deviation of the PRPG (50%) is also slightly higher than the 36% predicted by Hillel et al. (1990) for first backcrosses. The PRPG of the BC1 was also calculated using a scaffold set of 50 loci. Rank correlations between the larger and smaller data sets were high (r = 0.72) and significant (Table 5).

COMPARISON OF MD AND PRPG. Although the MD and PRPG each produce reasonable descriptions of the introgression in the BC1, the correlation between the two methods while statistically significant is strikingly low (r = 0.23; Table 5). Backcrosses that were statistically quite similar to J. regia, i.e., that were virtually identical to J. regia morphologically, often contained as much as 25% J. hindsii genome. The rank correlation between MD and PRPG based on a scaffold set of markers was also quite low (r = 0.14) and nonsignificant. In our population two single morphological traits SL/SW and SW/RL, had equal or better correlations with PRPG than the MD (Table 5).

Discussion

In this study the PRPG and MD were investigated as measures of linkage drag. These measures could then be used for selecting among backcrosses. The results demonstrate that the parental species and their hybrids have significantly different morphologies. The estimates of genetic parameters, as well as the significant differences for the traits between and within populations, suggest that the observed morphological differences have a genetic basis. When combined into a single, statistical measure of introgression, MD, the ability to measure a composite morphological score is increased, and the type of variability is better described.

Table 4. Statistical summary and means separation of the log-transformed Mahalanobis' distance (MD) from mature Juglans regia to individual, mature

<table>
<thead>
<tr>
<th>Population</th>
<th>n³</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean⁴</th>
<th>SD</th>
<th>CV (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>J. regia</td>
<td>24</td>
<td>-0.08</td>
<td>1.18</td>
<td>0.60 a</td>
<td>0.30</td>
<td>50</td>
</tr>
<tr>
<td>Backcross (BC1)</td>
<td>80</td>
<td>0.64</td>
<td>1.84</td>
<td>1.34 b</td>
<td>0.27</td>
<td>20</td>
</tr>
<tr>
<td>F1 hybrids</td>
<td>11</td>
<td>1.57</td>
<td>1.91</td>
<td>1.76 c</td>
<td>0.10</td>
<td>5.7</td>
</tr>
<tr>
<td>Backcross (BCH)</td>
<td>14</td>
<td>1.82</td>
<td>2.34</td>
<td>2.07 d</td>
<td>0.16</td>
<td>7.7</td>
</tr>
<tr>
<td>J. hindsii</td>
<td>15</td>
<td>2.11</td>
<td>2.49</td>
<td>2.32 e</td>
<td>0.09</td>
<td>3.9</td>
</tr>
</tbody>
</table>

³Sample size is the number of genotypes evaluated in each population, not the number of observations.
⁴Means followed by the same letter are not significantly different, P ≤ 0.05, by REGWQ test.


261
MD, the traits reflected the expected genomic mixing of *J. hindsii*, *J. regia*, their F₁ hybrids and the backcross populations. The differences in the coefficient of variation (C) of the MD for the parental populations (Table 4) suggests that the component traits of the MD may better represent the variability within *J. regia* than *J. hindsii*. If so, this might be an advantage to a breeding program with the goal of selecting those BC₁s morphologically most similar to *J. regia* for further backcrossing and eliminating as much of the *J. hindsii* genome as possible. Alternatively, one might attempt to capture the genetic variability introduced by *J. hindsii* by identifying backcrosses that are morphologically more like *J. hindsii* but with suitable horticultural qualities. Since in this study the MD and PRPG did not correspond, a simple, rapid morphological selection among the backcrosses does not seem promising. Several of the traits used in this analysis were strongly phenotypically correlated (Woeste, 1995). Evaluating a larger number of genetically uncorrelated traits might have produced a better agreement between the MD and PRPG, but the work involved in identifying these traits, scoring the populations and analyzing the data is considerable.

The relationship between the PRPG and the breeding value of the backcrosses is unclear. Selection against *J. hindsii* genome may not be necessary if traits such as shell thickness are strongly affected by only a few genes. In fact, promising variability for nut quality emerged early in the backcrossing (Woeste, 1998). Specific associations between markers and traits of interest might be useful, but because the *J. regia X J. hindsii* cross is male sterile and F₁ and BC₁ plants often have very low fertility, these associations could be due to linkage disequilibrium and they would not be present in other crosses. Adding more markers would improve our ability to sample the hybrid genome only marginally since any additional markers would be linked to and in linkage disequilibrium with markers that already exist.

The low correlation between statistical and genetic distance in this walnut BC₁ population is similar to the results of studies by Smith and Smith (1989b) and Beer et al. (1993). In these and other studies involving lines of field crops, researchers typically used large numbers (>10) of agronomic characters as the basis of their morphological evaluation. Despite the large data sets that have been generated in this fashion, the correlation between MD and genetic markers is typically low, in the range of r = 0.20 to 0.30. The results from (sometimes distantly) related populations of field crops are true of backcrosses as well. Selection during backcrossing for recurrent parent genome based on morphology would be an unreliable approach in this walnut BC₁ population. These results also correspond with those from ecological studies. Typically

![Graph](image1.png)

**Fig. 1.** Distribution of Mahalanobis’ distance (MD) for 80 first backcross walnuts. Arrows indicate mean MD for *Juglans regia* and *J. hindsii* X *J. regia* F₁ populations.

![Graph](image2.png)

**Fig. 2.** Distribution of percent *Juglans regia* genome for 76 first backcross walnuts based on 100 randomly amplified polymorphic DNA (RAPD) markers and restriction fragment-length polymorphisms (RFLPs).

<table>
<thead>
<tr>
<th>Correlation</th>
<th>Percent <em>Juglans regia</em> genome</th>
<th>Percent <em>Juglans regia</em> genome (scaffold)</th>
<th>MD</th>
</tr>
</thead>
<tbody>
<tr>
<td>SL/SW</td>
<td>-0.23</td>
<td>-0.22</td>
<td>0.21**</td>
</tr>
<tr>
<td>RL/N</td>
<td>-0.20</td>
<td>-0.06</td>
<td>-0.52***</td>
</tr>
<tr>
<td>BL/BW</td>
<td>-0.07</td>
<td>-0.19</td>
<td>0.19</td>
</tr>
<tr>
<td>SW/RL</td>
<td>0.31**</td>
<td>0.24**</td>
<td>-0.62***</td>
</tr>
<tr>
<td>PL/RL</td>
<td>0.22</td>
<td>0.07</td>
<td>-0.63***</td>
</tr>
<tr>
<td>Percent <em>Juglans regia</em> genome</td>
<td></td>
<td>0.72***</td>
<td>0.23**</td>
</tr>
<tr>
<td>Percent <em>Juglans regia</em> genome (scaffold)</td>
<td></td>
<td></td>
<td>0.14</td>
</tr>
</tbody>
</table>

*a n ≥ 50 for all correlations.

*b See text for a description of the traits.

* * * = Significant at * P ≤ 0.05 or 0.0001, respectively.

Table 5. Rank correlations among five morphological traits, percent *Juglans regia* genome, percent *Juglans regia* genome based on a scaffold set of markers, and Mahalanobis’ distance (MD) for first backcross walnuts.
ecologists consider morphological intermediacy to be reliable evidence of hybridization, but molecular data is often considered to be a more sensitive measure of hybridity, especially when the genetic relationships among populations are complex (Dowling et al., 1989).

In general, when morphological and genetic distance measures are used to evaluate the same population, the results, while sometimes congruent (Atchley et al., 1982b; Villani et al., 1992), are often surprisingly discordant (Atchley et al., 1988; Smith and Smith, 1989b; Lewontin, 1984, 1986) despite the well-accepted theoretical underpinnings of each method. This disparity accentuates the insight that morphological differentiation of populations and individuals is complex; not determined in a simply additive fashion as the sum of genes from each parent or parental population. This was true of the traits considered here, even though they were specifically chosen for this characteristic.

**Literature Cited**


