GROWTH AND ISOZYME ALLELE FREQUENCY CORRELATIONS IN BLACK WALNUT.

F. H. KUNG, G. RINK, AND G. ZHANG

Abstract.--Average height and diameter of black walnut trees in two progeny test plantations in southern Illinois were measured at age 19. Seed from fast and slow growing families were collected in 1989. Isozymes extraction and analysis was applied to embryos removed from these nuts. When we correlated mother tree growth measurements with progeny isozyme allele frequencies by locus, none of the correlations were significant at the 5% level. However, we found a significant positive correlation at the 10% level between tree height and the frequency of a PGI2 allele. Using canonical correlation analysis, the results were more encouraging. The first canonical correlation was greater than 0.99 and the second one was 0.72. Diameter growth had a higher correlation with allele frequencies than height growth. Canonical correlation may be more practical than simple correlation analysis for studying relationships between isozyme allele frequencies and growth variables because of the polygenic inheritance pattern of growth traits.

Keywords: Juglans nigra electrophoresis, isozyme.

INTRODUCTION

Chemical composition of enzymes is determined by DNA strands in chromosomal genes; therefore, variation in chemical and/or physical enzyme composition reflects genetic variation. Electrophoresis can separate different proteins and enzymes on the basis of ionic charge, size, and shape. Separation occurs on gel media in response to variable enzyme migration rates at a given electric field. Interpretation of isozyme banding is based on the segregation patterns scored for enzyme polymorphism.

A preliminary study of isozyme variation in black walnut (*Juglans nigra* L.) was initiated in 1984 to determine which isozyme systems are best suited for discriminating among black walnut genotypes (Rink et *al.* 1989). We assayed 21 enzyme systems and found eight polymorphic enzymes which provided eleven loci with detectable isozyme variation: AAT1, ACO1, ACO2, PGM1, PGM2, PGM3, 6-PG2, ADH, FEST, ACP2, and PGI2. Another nine systems showed enzyme activity but lacked consistently reproducible results and were considered to be unsuitable. Those enzymes include DIA, GDH, BATA-GAL, GPT, G-2-D, EST, UGPP, SDH, and SKD. Four systems were apparently homozygotic (IDH G-6-P, ME, and MDH) and were also not useful.

Professor, Dept. of Forestry, Southern Illinois University, Carbondale, IL. 62901, Principal Research Geneticist, USDA Forest Service, North Central Forest Experiment Station, Carbondale, IL. 62901, and Lecturer, Dept. of Forestry, Guangxi Forestry College, Nanning, Guangxi, People's Republic of China.

Knowing the relationship between isozyme variation and fitness traits (growth, survival and reproduction) variation may improve the efficacy and efficiency of selective breeding. Potentially superior trees may be identified by isozyme banding pattern. The allele arrangement of superior trees might indicate whether pure lines or hybrid lines in certain enzyme systems should be maintained. For example, Werner and Moxley (1991) found a relationship between malate dehydrogenase (MDH) isozyme genotype and plant vigor in peach *[Prunus persica (L.)* Batsch]. Homozygous Mdh1-2/Mdh1-2 individuals showed the greatest vigor, and were significantly different in vigor from Mdh1-1/Mdh1-1 homozygotes and from Mdh1-1/Mdh1-2 heterozygotes.

Instead of comparing the distribution of height growth among known fixed genotypes as proposed by Werner and Moxley (1991), we explored the correlation between height growth and isozyme frequency. Although correlation and causation are not synonymous, the use of correlation may serve us as a tool for expanded exploratory data analysis.

METHODS

Average height and diameter of black walnut trees in two progeny test plantations in southern Illinois were measured at age 19. Seed from fast and slow growing families were collected in 1989. Height, in decimeters, and diameter at breast height, in millimeters, are presented in Table 1. The average number of trees measured was 10, and the average number of nuts collected was 29 per family.

Table 1. Growth and isozyme allele frequency of 13 black walnut families. The top nine families were fast growing and the bottom four were slow growing.

FAM no.	HT dm.	DBH mm.	PGM21	PGM22	AC011	ACO21	ACO22	6PG21	6PG22	FEST1	ACP21	PGI21
			- - - - - - - - - - - - frequency -									
6300	94	128	0.50	0.50	0.50	0.75	0.25	0.75	0.25	1.00	1.00	1.00
6342	93	155	0.67	0.29	0.63	0.73	0.15	0.76	0.24	1.00	0.68	0.95
6344	84	100	0.81	0.17	0.68	0.61	0.39	0.64	0.25	0.94	0.58	0.98
6303	81	103	0.45	0.55	0.67	0.82	0.16	0.92	0.04	0.99	0.76	1.00
6325	76	119	0.75	0.25	0.45	0.82	0.15	0.93	0.07	0.98	1.00	0.96
6340	74	113	0.68	0.32	0.52	0.80	0.20	0.68	0.32	1.00	1.00	0.84
6325	72	98	0.57	0.43	0.80	0.53	0.47	0.90	0.10	0.69	0.69	0.80
6394	70	106	0.82	0.18	0.78	0.60	0.40	0.66	0.14	1.00	0.97	1.00
6340	65	118	0.70	0.30	0.53	0.69	0.18	0.65	0.12	0.95	0.92	1.00
6386	62	97	0.75	0.20	0.84	1.00	0.00	0.69	0.31	1.00	1.00	1.00
6382	61	92	0.58	0.40	0.53	0.53	0.45	0.59	0.41	0.99	0.58	0.95
6391	61	91	0.60	0.40	0.50	1.00	0.00	1.00	0.00	1.00	0.97	0.71
6386	52	75	0.40	0.60	0.60	1.00	0.00	0.40	0.60	1.00	0.50	0.60

Nuts were dehusked and stored at 2-5°C until needed, then cracked open and embryos removed. Embryos from each nut were individually macerated in a leaf extraction buffer described by Marty *et al.*(1984). Filter paper wicks (2 x 10

mm.) were soaked with the resulting liquid and then inserted in the starch gel. Electrophoretic methods used in this study are described by Marty et *al.* (1984).

Interpretation of isozyme banding was on the basis of segregation patterns. Embryo genotype data were analyzed using the Multilocus Estimation Program of Ritland and Jain (1981). Independent and variable allele frequencies in progeny are presented in Table 1.

STATISTICAL ANALYSIS

The CORR procedure (SAS 1982), the RSQUARE and the CANCORR procedures (SAS 1985) were used for data analysis. The CORR procedure computes Pearson product moment correlation between variables. The RSQUARE procedure finds subsets of independent variables that best predict a dependent variable by linear regression. We defined allele frequencies as independent and height or diameter as dependent variables. The CANCORR procedures analyzes the relationship between two sets of variables. Each set can contain several original variables. Our variable set named "growth" contained measurements on height and diameter, while the variable set "gene" contained allele frequencies.

Because of the low number (n=4) of slow growing families, we found that the relationship between growth and allele frequency was not significantly different from that of the fast growing families. Therefore, the two sets were pooled together as a single data base.

RESULTS AND DISCUSSION

Because the model used for the RSQUARE and the CANCORR procedures must be a full rank model, not all observed allele frequencies can be included. Therefore, we eliminated allele frequencies which were fixed. Fixation may be due to limited sample size or statistical dependency between alleles for a given locus. For example, we found the allele frequency ADH1 = 1.0 in each of the 13 families. In the previous study the allele frequency ADH1 =0.923 and ADH2 =0.077 among 948 embryos from 26 open-pollinated progenies (Rink et *al.* 1989). To avoid statistical dependency, we systematically excluded the last allele frequency in each locus. For example, PGM2 was interpreted to have three alleles: PGM21 PGM22 and PGM23, the PGM23 allele became fixed once the other two were included. Therefore, PGM21 and pgm22 were used but PGM23 was not studied.

None of the correlations between growth and allele frequency in Table 2 were significant at the 5% level. However, at the 10% level we found a positive correlation between the frequency of a PGI2 allele and growth. Taller and bigger mother trees tended to produce embryos with greater PGI2 allele frequencies.

Phosphoglucose isomerase (PGI) is known to be an enzyme which catalyzes the interconversions of glucose-6-phosphate and fructose-6-phosphate, both important substrates for respiration. However, the exact role alleles PGI1 and PGI2 have in this process or in the accumulation of biomass for height and diameter in black walnut trees is not known.

In most cases, the pattern of correlations of allele frequencies with height growth and with diameter growth were similar. We interpreted the algebraic sign

Table 2. Correlation analysis for growth and allele frequency.

isozyme locus	Between allele frequency and Ht.	Between allele frequency and Dbh,
	Correlation Coefficients	
PGM21	0.1193	0.2763
PGM22	-0.1261	-0.2910
ACO11	-0.0661	-0.2066
ACO21	-0.3206	-0.2196
ACO22	0.2651	0.0451
6PG21	0.3384	0.2562
6PG22	-0.3077	-0.2851
FEST1	-0.0085	0.1350
ACP21	0.1113	0.3035
PGI21	0.5405	0.5325

Table 3. R-square of regression models for dependent variable HT and for dependent variable DBH

	R-square	Variables in HT Model
1	0.292	PGI21
2	0.364	6PG21 PGI21
3	0.416	6PG21 6PG22 PGI21
4	0.452	ACO21 6PG21 6PG22 PGI21
5	0.472	ACO21 ACO22 6PG21 6PG22 PGI21
6	0.482	ACO21 ACO22 6PG21 6PG22 FEST1 PGI21
7	0.497	PGM21 PGM22 ACO21 ACO22 6PG21 6PG22 PGI21
8	0.513	PGM21 PGM22 ACO21 ACO22 6PG21 6PG22 ACP21 PGI21
9	0.522	PGM21 PGM22 ACO11 ACO21 ACO22 6PG21 6PG22 ACP21 PGI21
10	0.538	PGM21 PGM22 ACO11 ACO21 ACO22 6PG21 6PG22 FEST1 ACP21 PGI21

	R-square	Variables in DBH Model
1	0.283	PGI21
2	0.429	ACO21 ACO22
3	0.572	ACO21 ACO22 ACP21
4	0.597	ACO21 ACO22 ACP21 PGI21
5	0.705	ACO21 ACO22 6PG21 6PG22 ACP21
6	0.742	ACO21 ACO22 6PG21 6PG22 ACP21 PGI21
7	0.765	ACO11 ACO21 ACO22 6PG21 6PG22 FEST1 ACP21
8	0.767	ACO11 ACO21 ACC22 6PG21 6PG22 FEST1 ACP21 PGI21
9	0.789	PGM21 PGM22 ACO11 ACO21 ACO22 6PG21 6PG22 FEST1 ACP21
10	0.805	PGM21 PGM22 ACO11 ACO21 ACO22 6PG21 6PG22 FEST1 ACP21 PGI21

of the correlation as an indicator of vigor. For example, in a two allele system, fast growing families displayed greater frequencies of PGI21, ACP21, and ACO12; and slow growing families had greater frequencies of PGI22, ACP22 and ACO11. Similarly, in a 3-allele system, fast growing was associated with more PGM21, ACO22 and 6PG21, and slower growth with more PGM22 ACO21 and 6PG22. However, there was some indication that increasing FEST1 frequency is associated with increased diameter but not height growth.

Because growth of forest trees is considered to be under the control of many additive genes, it is logical that variation in growth may potentially be modelled using a combination of several isozyme systems. The RSQUARE procedure calculates R^2 for all possible combinations of subset variables. However, from the given 10 loci we can construct 10 regression models using one locus, 45 models for 2 loci, 120 models for 3 loci. It is not necessary to report every possible model here, so only the top ones are presented in Table 3.

Using greater numbers of enzyme systems results in a better fit of the regression. Tree diameter can be predicted better than tree height from the allele frequencies in the multiple regression model. When all 10 frequencies were used, the R-square for the DBH model was 0.805, but for the HT model, it was only 0.538.

Because simple and multiple correlation are special cases of canonical correlation in which one or both sets contain a single variable, we wanted to explore the general relationship between growth and allele frequency in greater detail. Although the 10 allele frequencies may be statistically independent but biologically correlated variables as are tree heights and tree diameters, canonical analysis enabled us to extract canonical variates that are uncorrelated to each other and which can produce the highest correlation between growth and allele frequency. Since the number of canonical variables always equals to the smaller number of the original variables in the two sets, we had two canonical variables from each set. We named the canonical variables extracted from the 10 alele frequencies as GENE1 and GENE2, and those extracted from height and diameter as GROWTH1 and GROWTH2.

The first canonical correlation between GROWTH1 and GENE1 was 0.99, and the second canonical correlation between GROWTH2 and GENE2 was 0.72 (Table 4). The canonical variable GROWTH1 is associated mainly with diameter, and GROWTH2, with height. Therefore, the factor GROWTH1 may be associated with cambial growth and GROWTH2, with shoot growth. The canonical variable GENE1 is associated mainly with allele PGM21, PGM22, ACO11 and ACP21, while GENE2 is associated with ACO21, ACO22, 6PG21, 6PG22, and PGI21. Although the genetic factors GENE1 and GENE2 are likely responsible for biochemical functions common to isozyme systems in each group, exact interpretation of these genetic factors awaits the availability of a larger data base.

The relationship between simple correlation and canonical correlation can be explained as follows. According to path analysis (Li, 1975), correlation between two variables is the sum or product of all paths between them: paths arranged in series are multiplicative while paths in parallel are additive. For example, there are two routes from PGI21 to DBH: one goes through GENE1-GROWTH1, the other goes through GENE2-GROWTH2. Therefore, the correlation between PGI21

and DBH is

$$0.2982*0.9942*0.7854+0.6765*0.7156*0.6190 = 0.2328+0.2997=0.5325$$

which is in agreement with the simple correlation given in Table 2. Similarly, the simple correlation between PGI21 and HT can be calculate from the canonical structure diagram through the routes of GENE1-GROWTH1 and GENE2-GROWTH2:

$$0.2982*0.9942*0.2367+0.6765*0.7156*0.9716=0.0702+0.4703=0.5405$$

Table 4. Canonical correlation analysis and canonical structure between growth and allele frequency.

Canonical Correlation Analysis

Canonical Variables	Canonical Correlation	Approx. Standard Error
GROWTH1-GENE1	0.9942	0.0033
GROWTH2-GENE2	0.7156	0.1408

Canonical Structure

Set original variable	Correlation with their own canonical variables		Correlation with the opposite canonical variables	
	GROWTH1	GROWTH2	GENE1	GENE2
HT	0.2367	0.9716	0.2353	0.6953
DBH	0.7854	0.6190	0.7809	0.4430
	GENE1	GENE2	GROWTH1	GROWTH2
PGM21	0.3175	0.0641	0.3157	0.0459
PGM22	-0.3338	-0.0684	-0.3319	-0.0490
ACO11	-0.2608	-0.0069	-0.2593	-0.0050
ACO21	-0.0245	-0.4527	-0.0243	-0.3240
ACO22	-0.1961	0.4477	-0.1950	0.3204
6PG21	0.0643	0.4650	0.0640	0.3328
6PG22	-0.1411	-0.3949	-0.1403	-0.2826
FEST1	0.2227	-0.0877	0.2214	-0.0628
ACP21	0.3687	0.0353	0.3666	0.0253
PGI21	0.2982	0.6765	0.2965	0.4841

Although the above two simple correlations are similar in size (0.5325 vs. 0.5405), and in paths (GENE1-GROWTH1 and GENE2-GROWTH2), their components are

quite different. The route from GPI21 toward DBH via GENE1-GROWTH1 contributed 0.2328/0.5325=44% to the simple correlation, but only 0.0702/0.5405=13% toward HT. Thus, the canonical correlations, GENE1-GROWTH1 and GENE2-GROWTH2, are the common links between any genetic variables on one side and any growth variables on the other side. They are the most important components and principal paths. However, the simple correlation is also determined by the strength between the set variables and their own canonical variable.

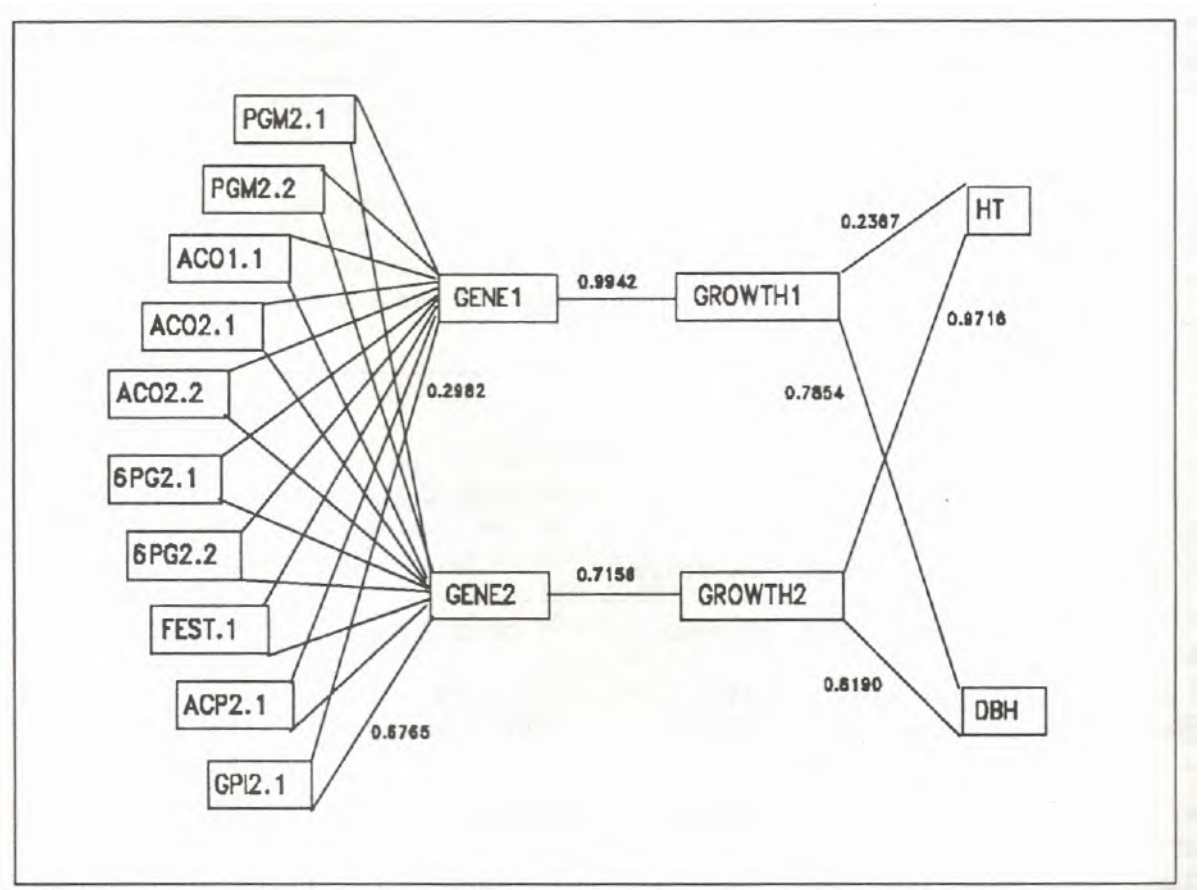

Figure 1. Path diagram and canonical structure for allele frequencies on the left and growth traits on the right.

CONCLUSIONS

Variation in height and diameter growth of black walnut trees seems to be related to phosphorylated enzyme frequencies. Diameter growth has a higher correlation with allele frequencies than height growth. Canonical correlation analysis is a better statistical tool to study the relationship between vigor and isozyme frequency than simple or multiple correlation.

ACKNOWLEDGMENTS

We would like to thank the Illinois Department of Conservation and Southern Illinois University for providing the study sites for the plantings from which we collected nuts for this report. This research work was partially supported by McIntire-Stennis Fund.

LITERATURE CITED

Li,C.C. 1975. Path Analysis - a primer. The Boxwood Press, Pacific Grove. Calif. 346 pp.

Marty,T.L., D.M.O'Malley, and R.P.Guries. 1984. A manual for starch-gel electrophoresis: A new microwave edition. Univ. Wisc. Staff. Pap. Ser. No. 20. 23 pp.

Rink,G., E.R.Carroll, and F.H.Kung. 1989. Estimation of Juglans nigra L. mating system parameters. For. Sci. 35(2):623-627.

Ritland,K., and S.Jain. 1981. A model for the estimation of outcrossing rate and gene frequencies using N independent loci. Heredity. 47:35-52.

SAS. 1982. SAS User's Guide: Basics 1982 edition. SAS Inst. Inc. Cary. NC. 923 pp.

SAS. 1985. SAS User's Guide: Statistics. Version 5 ed. SAS Inst. Inc. Cary. NC. 957 pp.

Werner,D.J., and D.F.Moxley. 1991. Relationship between Malate Dehydrogenase isozyme genotype and plant vigor in peach. *J.* Amer. Soc. Hort. Sci. 116(2):327-329.