Theor Appl Genet (1994) 87:947–956

THEORETICAL AND APPLIED GENETICS

RFLP variation and genealogical distance, multivariate distance, heterosis, and genetic variance in oats

H. Moser, M. Lee

Department of Agronomy, Iowa State University, Ames, IA 50011, USA

Received: 7 June 1993 / Accepted: 28 June 1993

Abstract. Patterns of restriction fragment length polymorphisms (RFLPs) have been proposed as estimators of genetic diversity among breeding lines and as predictors of heterosis and genetic variance. We evaluated these proposals by using a set of nine elite oat lines crossed in a diallel mating design without reciprocals. RFLP analysis was conducted using *Hin*dIII-digested DNA and a total of 107 probes from three different sources: 14 heterologous wheat cDNA clones, 17 oat genomic clones, and 76 oat cDNA clones. Of the 77 probes that produced high-quality autoradiographs, 26 detected polymorphisms among this set of lines, with an average of 2.6 variants per probe. RFLP-based genetic distance (*FD*) was calculated from these data by using Nei and Li's measure of genetic similarity, and was compared with two other measures of genetic divergence. Genealogical distance (*GD**) was obtained from the coefficients of parentage based on known parental pedigrees, and multivariate distance (*DI*) was calculated by using the first five principal components of the parental correlation matrix for 12 agronomic traits. *FD* was significantly correlated with *GD** ($r = 0.63$, $P < 0.01$), but not with *DI* ($r = -0.05$). Cluster analysis based on these three distance estimates did not produce equivalent groupings, but the *FD* and *GD** clusters were more similar to each other than to the *DI* clusters. These results indicate that: (1) sufficient variation exists for further application of RFLP technology to oats, (2) RFLPs could provide accurate estimates of genetic divergence among elite oat lines, and (3) it is unlikely that dispersed markers can predict heterosis or population genetic variance in oats. Further investigations will require more parental lines, a larger set of markers, and more information on the linkage relationships between RFLP markers and loci controlling the trait of interest.

Key words: *Avena sativa* L. – Genetic distance – RFLPs – Heterosis – Genetic variance

Introduction

Knowledge of genetic relations among elite lines and breeding materials of a crop species is essential for sustained and efficient cultivar imporvement. Information concerning genetic diversity among a set of lines can be useful as a general guide in the choice of parents for hybrid synthesis or population formation. However, prediction of the best combination of parents to form source populations for breeding is still, at best, an educated guess; consequently, the final evaluation of the progeny has to be determined through extensive field testing.

Ideal source populations for the selection of pure lines have both a high mean a large genetic variance for the traits of interest (Cox et al. 1985). The mean of a source population is usually associated with the parental means, but genetic variance is dependent, in part, on the degree of divergence between parents. Likewise, heterosis is associated not with per-se parental performance, but rather with the genetic divergence between the parents of the hybrid. Thus, the efficiency of most breeding programs could be improved if quantitative estimates of genetic divergence

Journal paper No. J-15302 of the Iowa Agriculture and Home Economics Experiment Station, Ames, IA 50011, USA. Project No. 2818 and 2447. Supported by Quaker Oats grant to M. Lee
Communicated by J. Mac Key
Correspondence to: H. Moser

could accurately predict the genetic variance of a population, or the mid-parent heterosis of a hybrid, without having to make the crosses and evaluate the progeny.

Breeders have proposed many estimates of genetic distance to quantify genetic relations among lines within a species. Three of the most common estimates are based on the coefficient of parentage probabilities calculated from pedigree information, the multivariate analysis of quantitative trait variation, or the analysis of molecular markers. Agreement among the three types of divergence estimates can vary because of sampling effects and failures in certain assumptions regarding the phenotypic expression of genetic variation, the relationships among ancestral lines, and selection during inbreeding (Cox et al. 1985; LeFort-Buson et al. 1986; Atchley et al. 1988; Melchinger 1993).

These relations have been investigated for several traits in oats. Cowen and Frey (1987a, b, c) studied the relations between genetic diversity among a set of nine elite oat lines and the amount of heterosis, transgressive segregation, and genetic variance expressed in their progeny. They calculated four types of genetic distances estimates: *GD** based on the coefficient of parentage, *DI* based on the multivariate analysis of parental data, and two estimates based on progeny data (*DII* and *DIII*). *GD** was most closely related to the population genetic variance, and *DI* most accurately predicted heterosis in one experiment, but none of the four distance estimates could be used to predict any of the three genetic parameters consistently or accurately.

Souza and Sorrells (1989, 1991a, b) evaluated the genetic relations among a diverse group of North American oat lines based on three measures of genetic diversity: the coefficient of parentage (COP), which is similar to *GD**; quantitatively-inherited morphological characters (QMC); and discretely-inherited morphological and biochemical characters (DMBC). COP was the best single predictor of both specific combining ability (SCA) and genetic variance, but the sum of all three measures, as a combined distance estimate, was a better predictor of SCA than any single distance measure. Even the best regression models, however, explained less than half of the variation in population genetic variances and no more than one-third of the variation in SCA. Also, contrary to expectations, the lowest SCA and smallest genetic variance estimates were often associated with the more divergent matings.

Analysis of variation in RFLPs among a set of lines can provide an estimate of genetic distance that has several advantages. The RFLP estimates are based on direct sampling of the genome, can be determined for any combination of lines, are free from environmental influences, and require no pedigree information or simplifying assumptions regarding relations among ancestral lines or selection parameters (Smith and Smith 1989). RFLP- and pedigree-based genetic distance estimates can be in high agreement (Smith et al. 1990; Melchinger et al. 1991), but RFLP-based genetic distances and heterosis seem to be associated only among parents that are closely related by pedigree (Melchinger 1993). Smith et al. (1990) observed highly-significant correlations between RFLP-based genetic distances and coefficients of parentage, F_1-yield, and heterosis in a group of 37 inbred maize lines representing a wide range of pedigree relationships. They concluded that RFLP-based distance estimates, combined with pedigree and quantitative trait locus (QTL) information, could accurately predict the best parental combinations for the production of high-yielding hybrids. Lee et al. (1989) observed a high correlation ($r = 0.74$) between RFLP-based distance estimates and specific combining ability in a diallel that included both related and unrelated crosses. Melchinger et al. (1990a, b) and Lee et al. (1989), however, found little association between RFLP distance and heterosis when they analyzed only unrelated maize inbreds as parents.

The objectives of our research were to: (1) investigate the extent of RFLP variation among the nine elite oat lines analyzed by Cowen and Frey (1987a, b), (2) determine the RFLP-based genetic distance (*FD*) among these lines and its association with *GD** and *DI*, and (3) determine the relationship between *FD* and heterosis and population genetic variance in this set of lines.

Materials and methods

Genetic materials

Nine elite oat lines or cultivars from four different midwestern U.S. oat (*Avena sativa* L.) breeding programs were included in this study (Table 1). Cowen and Frey (1987a, b) produced 35 of 36 possible F_1 hybrids among these nine lines in a half diallel without reciprocals. Forty-eight F_2-derived lines in the F_3 or F_4 were produced without selection from all possible matings (excluding reciprocals) to form 36 biparental source populations.

Field evaluation

Detailed procedures for field evaluation and estimation of genetic parameters are given in Cowen and Frey (1987a). We will briefly review the procedures here. Experiment 1 provided estimates of heterosis expressed in oat progenies when grown in competitive stands. The parents and F_1 hybrid progeny of each of eight matings (L1 × L9, L2 × L5, L2 × L6, L2 × L7, L2 × L8, L3 × L5, L4 × L7, and L7 × L9) were grown in hill plots in a randomized complete block design with two replications at three locations in 1983. Heterosis for grain yield (*GY*) was calculated as the difference between the F_1 hybrid mean and the mid-parent value. The significance of heterotic deviations was determined from an analysis of variance on the combined data from each mating.

Experiment 2 estimated the heterotic response for grain yield of 35 out of 36 possible matings among the nine parents. The parents and the F_1s of the 35 matings were grown as individual

Table 1. Cultivar name, state of origin, and pedigree of lines used in this study

No.	Name	State	Pedigree
L1	B605-1085	Iowa	Unknown
L2	Porter	Indiana	CI7684 Seln./3/Putnam*5/Minn313*s//Albion/4/Stout
L3	D226-30-8	Iowa	Y22-15-9//Clintford*4/B444 (*Avena sterilis*)
L4	Y22-15-9 (PI468112)	Iowa	Garland/B433 (*A. sterilis*)//Holden
L5	Ogle	Illinois	Brave//Tyler/Egdolon 23
L6	Y341-41 (CI9273)	Iowa	Clintford*6/B443 (*A. sterilis*)
L7	Bates	Missouri	Pettis/Florida 500
L8	Y349-23 (CI9277)	Iowa	Clintland/Garry seln. 5//CI8044*4/B443 (*A. sterilis*)
L9	Lang	Illinois	Tyler/Orbit

plants within plots in a completely-randomized design at one location. Parental lines were planted in four plots and the F_1 progeny were planted in one-to-three plots depending on the quantity of seed available. Mid-parent heterosis was calculated for grain yield and the standard error for heterosis was calculated for each mating based on the standard errors of the parents and progeny. The standard error for heterosis was used to determine the significance of the heterotic deviation for each mating.

Experiments 3 and 4 were designed to estimate genetic variance among F_2-derived lines for *GY*, above-ground biological yield (*BY*), straw yield (*SY*), harvest index (*HI*), height (*HT*), and heading date (*HD*). In experiment 3, 48 $F_{2:3}$ lines and the two parents from each population were grown in hill plots in a split-block design. Populations were assigned to main plots and individual lines to subplots. The experiment was grown at three locations in central Iowa, with two replications per location in 1983. Experiment 4 followed the same procedure as experiment 3 expect that $F_{2:4}$ lines and parents were grown at two locations in 1984. An analysis of variance was performed on each population (excluding parents) for both experiments separately, and for the combined data over 2 years for *GY*, *SY*, *BY*, and *HI*. A combined analysis of variance was conducted for *HT* and *HD*. The five year-locations were considered as unique environments in the combined analysis. Genetic variance components were estimated by equating observed mean squares to their expectations. Approximate standard errors for the genetic variances were obtained by the method of Bulmer (1957).

Cowen and Frey (1987a) calculated a generalized genetic variance (*GGV*) for each population. The genetic variance-covariance matrix, **G**, for three traits (*BY*, *GY*, and *HI*) was transformed to a diagonal matrix as follows:

$$\mathbf{D} = \mathbf{O}'\mathbf{G}\mathbf{O} \tag{1}$$

where **O** equals the orthogonal matrix of eigenvectors of **G**. The *GGV* is the product of the diagonal elements of **D**.

RFLP analysis

DNA was isolated from ground, lypholized leaf tissue (Saghai-Maroof et al. 1984) from a bulk of 15 5-week-old seedlings per line. Genomic DNA from each line was digested to completion with *Hin*dIII, loaded into an 0.7% agarose gel (10 µg DNA per lane) along with a set of lambda-derived molecular weight markers (Melchinger et al. 1991), electrophoresed at 23 V for 16–18 h, and transferred (Southern 1975) to MSI nylon membranes. A total of 107 clones was selected for probes from three different sources. Seventy-seven clones produced high quality autoradiographs. Fourteen wheat cDNA clones (W) were provided by Dr. M. D. Gale of the John Innes centre for Plant Research at the Cambridge Laboratory. These clones have been mapped in wheat and represent one probe for each set of homoeologous chromosomes. Four probes detected polymorphism in our oat lines and mapped to chromosome arms 1L, 2S, 3S, and 5L in wheat. Seventeen oat genomic clones (OG) shown to detect RFLPs in hexaploid oat crosses were selected from a group of clones provided by Dr. M. Sorrells from Cornell University. Seventy-six Iowa clones (ISU) were selected from a cDNA library prepared from roots of 8-day-old etiolated oat seedlings. These probes have detected RFLPs in diploid oat crosses between *Avena strigosa* and *A. wiesteii* (P. Rayaputi, ISU, personal communication). Mapping information regarding the OG or the ISU clones is not available. Inserts from all three sources were amplified by the polymerase chain reaction (PCR) for 36 cycles using "M13 forward" and "M13 reverse" as primers. Amplified DNA was run on a 1% low-melting-point agarose gel; inserts were excised from the gel, diluted to 3 ng μl^{-1} with double distilled water, labelled with ^{32}P-dCTP by the random primer reaction (Feinberg and Vogelstein 1983), and hybridized to DNA on the nylon membranes (Helentjaris et al. 1985).

Calculating genetic distances

RFLP patterns were recorded for each probe and line combination by assigning a 1 for the presence of the band and 0 for the absence of a band. Nei and Li's *F* statistic (Nei and Li 1979) is a measure of genetic similarity and is calculated as:

$$F_{ij} = \frac{2N_{ij}}{N_i + N_j} \tag{2}$$

where N_{ij} equals the number of bands in common between parent *i* and parent *j*, and N_i and N_j equal the number of bands for parent *i* and parent *j*, respectively. The distance between two lines was calculated as $FD = 1 - F_{ij}$, which can range from 0, when all bands in two lines are identical, to 1, when there are no bands in common between two lines.

The precision of our *FD* estimates was evaluated by two methods. First, standard errors for *FD* (SE_{FD}) were calculated by using the jackknife, and 95% confidence intervals were calculated as $FD \pm t_{0.025,25}\ SE_{FD}$ (Efron 1982). Second, the probes were randomly divided into two equal sets of 13, the *FD* determined for each set of 13 probes, and the product-moment correlation calculated between the two estimates. This procedure was repeated four times.

Coefficients of kinship between two lines (f_{ij}) were calculated from their pedigree by using an iterative approach according to Cruden (1949) and Emik and Terrill (1949) under the assumption that the first known progenitors were unrelated ($f_{ij} = 0$; see Cowen and Frey 1987b). Selections from the same introduction were assigned half-sib relationships ($f_{ij} = 0.25$). Jacquard (1974) defined genealogical distance as $GD_{ij} = 1 - f_{ij}$. Since the *GD*

values among this set of lines tended to be low, Jacquard's suggested transformation was applied to the *GD* values to increase the separations of the matings:

$$GD^*_{ij} = e^{(1-f_{ij})} - 1. \quad (3)$$

Possible *GD** values range from 0, when two lines have the same pedigree, to 1.72, when they have no ancestors in common. The elite line B605–1085 was not included in this analysis of *GD** because it was selected from a composite cross and has an unknown pedigree.

A data set consisting of the following 12 agronomic traits was constructed for these nine lines to calculated *DI*, the multivariate-based distance: grain yield, biological yield, straw yield, harvest index, height, heading date, groat-protein percentage, groat percentage, test weight, lodging score, barley yellow dwarf virus reaction, and mean rust reaction. Principal component analysis was performed on the data set using the correlation matrix, and *DI* was calculated as the Euclidian distance between the first five principal components (see Cowen and Frey 1987b).

Statistical analysis

Simple correlations were computed to determine the relations among the three types of genetic distance estimates, and the relations between genetic distance and estimates of heterosis for grain yield and population genetic vaiances for *GY*, *BY*, *SY*, *HI*, *HT*, and *HD*. Two of the thirty-six matings in this diallel were between closely-related lines since Y22 is the female parent of D226 and 'Clintford' contributes prominently to the parentage of both Y341 and D226 (see Table 1). All correlations were calculated both with and without these two closely-related matings (Y22/D226, and Y341/D226).

Cluster analysis, using Ward's minimum variance method (Ward 1963), was performed based on *GD**, *FD*, and *DI* distance measures. All procedures were computed with SAS software (SAS 1988).

Results and discussion

Variation in RFLPs

Of the 77 probes that produced high quality autoradiographs, 26 (33%) detected RFLPs among the nine parents used in this study (Table 2). Most of the probes revealed two-to-five bands in each lane, but there were exceptions; some probes revealed only one band per lane, and one probe revealed eight-to-ten bands per lane. Probes that revealed polymorphism tended to have more bands per lane, on the average, than those that revealed only monomorphic bands. Twenty-five of the probes that detected polymorphic bands also detected one-to-six monomorphic bands in the same pattern.

The 26 probes that revealed polymorphism in this set of lines detected a total of 68 variant banding patterns, for an average of 2.6 variants per probe. Sixteen probes (62%) distinguished only two variant types (Fig. 1). Six probes detected polymorphic sequences in only one of the nine lines, the other eight lines being monomorphic. In all, 22 of 68 variants were found in only one line (Fig. 2).

Table 2. Summary of RFLP results

Source of probe	Number of probes[a]		Average number bands per lane		Average number of variants per polymorphic probe
	Mono	Poly	Mono	Poly	
Wheat	5	4	2.6	2.8	2.5
Cornell	9	4	1.4	3.0	2.5
Iowa	37	18	2.3	3.7	2.7
Total	51	26	2.2	3.6	2.6

[a] Number of probes that detected monomorphic (Mono) or polymorphic (Poly) banding patterns

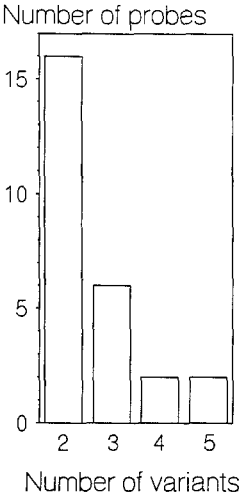

Fig. 1. Distribution of the number of RFLP patterns (variants) as detected by 26 probes among nine elite oat lines

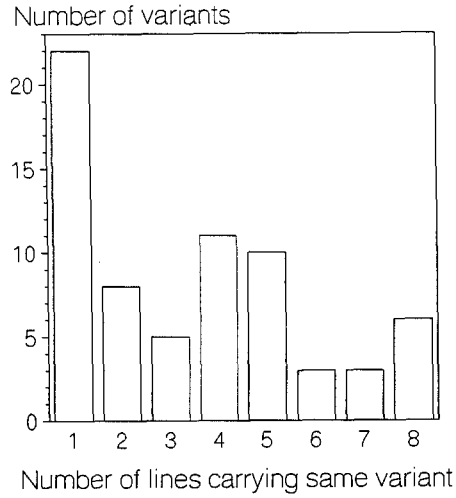

Fig. 2. Distribution of the number of parents with the same RFLP variant from a total of 68 variants. Twenty-two variants were present in only one of the nine lines. Fifty-one probes detected the same variant in all nine lines (i.e., were monomorphic)

These probes had been preselected for their ability to detect polymorphism in wheat, or in diploid or hexaploid oats. The degree of polymorphism in these lines, which do not represent the extremes of genetic diversity among cultivated oats, is similar to that of other self-pollinated plant species, such as tomato or barley (Helentjaris et al. 1985; Graner et al. 1990), and seems large enough to warrant further investigation into the application of RFLP technology to oats.

Associations among genetic-distance estimates

Although all but two of the matings in this study were fairly divergent by pedigree (average $GD^* = 1.23$), all *FD* estimates were below 0.22, with an average of 0.13 (Table 3). The frequent presence of monomorphic bands lowered the *FD* estimates among these lines. Ogle/Lang was the least divergent mating ($FD = 0.05$), whereas B605/Bates and Porter/Bates were the most divergent matings ($FD = 0.22$) based on RFLP data. B605, an Iowa line whose pedigree is unknown, seems most divergent from Ogle, Lang and Bates, and relatively closely related to the three elite lines from Iowa (Y349, Y341, and Y22). Bates, with an average *FD* of 0.19 and a GD^* of 1.51, is clearly the most divergent line based on both RFLP-and pedigree-based expectations. The *FD* values generally followed pedigree expectations; the correlation between *FD* and GD^* was a significant but moderate 0.63 ($P < 0.01$) (Fig. 3). Neither *FD* nor GD^* were correlated with *DI* ($r = -0.02$ and $r = -0.05$, respectively).

The cluster analyses based on these distance estimates did not produce equivalent groupings, but the clusters formed from GD^* and *FD* did contain some similarities; D226 and Y22 clustered closely together, Ogle and Lang clustered with Y349, and Bates was the only line in a group that was relatively distant from all other lines (Fig. 4). However, Y341 and Porter were not assigned to the same group in the pedigree analysis as in the RFLP analysis. Since Y22 and Y341 are both closely related to D226 (see Table 1), these three lines should cluster together based on pedigree expectations. The groupings based on DI were different from those based on either GD^* or *FD*.

The discrepancy between GD^* and *FD* can be expected because both methods are subject to sampling errors (in this case, possibly too few probes and

Table 3. Coefficient of parentage (*f*), genealogical distance (GD^*), RFLP-based distance (*FD*), and multivariate distance (*DI*) of 36 oat crosses

Parent 1	Parent 2	*f*	GD^*	*FD*[a]	*DI*
B605-1085	Lang	–	–	0.16	3.1
B605-1085	Y349-23	–	–	0.12	5.1
B605-1085	Bates	–	–	0.22	5.1
B605-1085	Y341-41	–	–	0.12	4.2
B605-1085	Ogle	–	–	0.17	3.7
B605-1085	Y22-15-9	–	–	0.09	4.7
B605-1085	Y226-30-8	–	–	0.10	2.7
B605-1085	Porter	–	–	0.12	4.4
Porter	Lang	0.11	1.44	0.15	5.5
Porter	Y349-23	0.25	1.12	0.10	4.9
Porter	Bates	0.06	1.56	0.22	5.1
Porter	Y341-41	0.21	1.21	0.13	4.4
Porter	Ogle	0.20	1.23	0.15	4.7
Porter	Y22-15-9	0.15	1.33	0.09	7.2
Porter	D226-30-8	0.18	1.27	0.08	5.9
D226-30-8	Lang	0.21	1.21	0.16	1.6
D226-30-8	Y349-23	0.25	1.12	0.10	4.6
D226-30-8	Bates	0.06	1.56	0.20	4.9
D226-30-8	Y341-41	0.59	0.50	0.10	5.3
D226-30-8	Ogle	0.18	1.27	0.15	3.5
D226-30-8	Y22-15-9	0.61	0.48	0.06	5.8
Y22-15-9	Lang	0.22	1.19	0.15	4.9
Y22-15-9	Y349-23	0.24	1.13	0.12	6.5
Y22-15-9	Bates	0.07	1.53	0.19	5.8
Y22-15-9	Y341-41	0.23	1.16	0.15	5.3
Y22-15-9	Ogle	0.18	1.27	0.14	7.4
Ogle	Lang	0.26	1.10	0.05	3.9
Ogle	Y349-23	0.23	1.17	0.06	4.2
Ogle	Bates	0.10	1.47	0.16	4.6
Ogle	Y341-41	0.18	1.27	0.11	3.9
Y341-41	Lang	0.20	1.22	0.08	4.6
Y341-41	Y349-23	0.26	1.10	0.10	3.4
Y341-41	Bates	0.05	1.58	0.18	4.3
Bates	Lang	0.09	1.48	0.16	3.6
Bates	Y349-23	0.10	1.47	0.17	4.3
Y349-23	Lang	0.25	1.11	0.07	3.8
Mean		0.20	1.23	0.13	4.6
Minimum		0.05	0.48	0.05	1.6
Maximum		0.61	1.58	0.22	7.4
Standard deviation		0.13	0.26	0.04	1.2

[a] Standard errors for *FD* for each cross ranged from 0.02 to 0.04

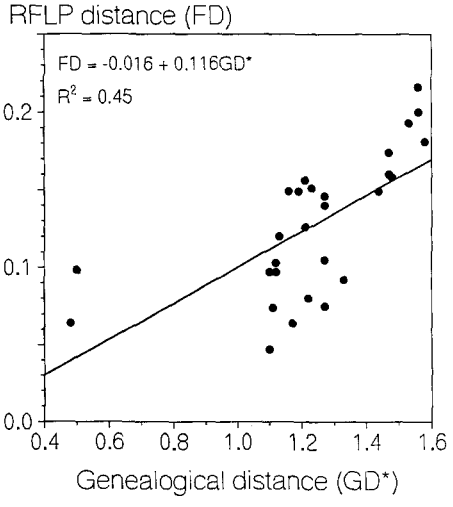

Fig. 3. Relationship between genealogical distance (GD^*) and RFLP-based distance (*FD*) in 28 matings among eight oat lines. The *vertical bar* represents 2 × the average standard error for *FD* as calculated by the jackknife

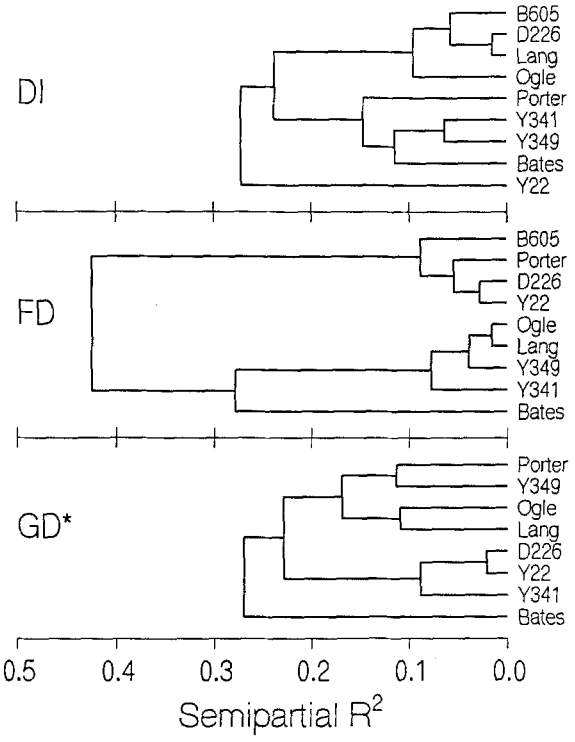

Fig. 4. Ward's minimum variance cluster analysis of lines based on multivariate distance (*DI*), RFLP-based distance (*FD*), and genealogical distance (*GD**). B605-1085 was not included in the *GD** analysis because it has an unknown pedigree

genotypes). *GD** is based on the probability that two lines share identical alleles at a locus and requires the assumption that each parent contributes equally to the progeny. Genetic drift during inbreeding can cause random deviations in the parental contribution, although the effects are probably not great because of the large number of factors that segregate in the progeny (St. Martin 1982).

Molecular markers represent a sample of a plant's genome, yet they are used to infer similarities of the entire genome among a set of lines or populations. The number of markers or probes used in an analysis, and the variability of the reference population, will directly affect the precision of the resulting estimates. Messmer et al. (1991) and Smith et al. (1992) have indicated that at least 100 probes (loci) are required to provide acceptable estimates of true genetic relationships (*D'*) among maize inbred lines. However, maize is relatively very polymorphic. We evaluated the precision of our estimates, based on only 26 probes, through two methods – calculation of standard errors by the jackknife method and correlations between random subsets of 13 probes (see Materials amd methods). Standard errors for most *FD* estimates were 0.03 and ranged from 0.02 to 0.04, resulting in 95% confidence intervals (*CI*) ranging from 0.08 to 0.16. These represent fairly large *CI*s given the low *FD* values in this study, but there were some significant differences in *FD* estimates among these crosses. Smith et al. (1992) reported confidence intervals of 0.30 for Nei and Li's F statistic based on 30 probes in maize. Correlations between subsets of probes ranged from 0.53 to 0.69 ($P < 0.01$); thus, two sets of 13 probes provide similar estimates of genetic distance. Doubling the number of probes to 26 (the number of probes used in this study) would probably improve the precision of the estimates to some degree. Messmer et al. (1991) conducted a similar analysis using two sets of 27 probes with two restriction enzymes in maize and obtained a correlation of only 0.17 ($P < 0.05$). The wider *CI*s and the smaller correlations between subsets of probes indicates that maize inbreds may require more loci to obtain accurate estimates of genetic distance because of the greater variation in RFLP loci. The magnitude of the standard errors for *FD* and the significant, but moderate, correlations between subsets of the 26 probes used in this study indicate detection of significant differences in *FD* estimates among these crosses, but more probes would have improved the precision of the results to some degree.

A survey of the literature reveals that the correlation between pedigree-based distances and molecular marker- or RFLP-based distances improves as more probes or marker loci are employed in the analysis. Ehoibu et al. (1990) obtained a nonsignificant correlation of 0.41 based on six isozyme probes in *Drosophila*; Cox et al. (1985) obtained correlations ranging fron 0.15 to 0.45 based on 13 isozyme and morphological markers in soybean; Smith and Smith (1989) obtained a correlation of 0.51 based on 31 isozyme loci in maize; Melchinger et al. (1991) estimated a correlation of 0.74 based on RFLP analysis with 83 probes in maize; and Smith et al. (1990) obtained a correlation of 0.90 using 257 probe-enzyme combinations in maize (all r values, except those of Ehibu et al. 1990, were significant at $P < 0.01$). The association between *GD** and *FD* observed the present study ($r = 0.63$ with 26 probes) is consistent with the literature, considering the number of probes or loci used in each case.

In addition to possible sampling errors, *GD** and *FD* contain assumptions that bias the genetic distance estimates in opposite directions. *GD** overestimates *D'* because it ignores possible relations among ancestral lines and selection for common linkage blocks, whereas *FD* tends to underestimate *D'* because it assumes that two bands with the same mobility in a gel are identical when they could actually be different. It is difficult to evaluate the validity of these assumptions in our study because the exact allelic nature of the variation detected by these probes in unknown.

Genetic-distance estimates based on the variation of, or the association among, quantitative traits (*DI*)

assume that phenotypic variation accurately reflects true genetic relationships. However, measurement of genotypic divergence through phenotypic variation is a complex statistical problem that requires not only accurate estimates of means, variances, and covariances, but also consideration of evironmental sources of variance, nonadditive genetic effects, intercorrelations among traits, heterogeneity among variance and covariance estimates, and genotype-by-environment interactions (Atchley et al. 1988). As a result of the complex nature of the relation between phenotypic and genotypic variation, lines that are phenotypically different may vary at only a few loci, and lines very similar in appearance or performance may be quite genetically dissimilar. The complete lack of association between *DI* and *GD** or *FD* in our study strongly suggests that this particular set of traits does not accurately reflect the true genetic distance (*D'*) among these lines.

Atchley et al. (1988) investigated the relationship among genealogical, molecular, and several univariate and multivariate genetic-distance estimates in ten inbred mouse strains. Although molecular-based distance estimates accurately reflected genealogical divergence among the strains, only one of the three multivariate distances was associated with molecular distance and none of them was associated with genealogical divergence. Souza and Sorrells (1989, 1991a, b, c) have summarized the genetic relationships among North American oat cultivars based on genealogical distance (COP), multivariate analysis of quantitative traits (QMC), and discrete morphological and biochemical characteristics (DMBC). Cluster analyses based on the three different estimates all gave different results, but DMBC and COP clusters were more similar to each other than to QMC clusters. Traits such as heading date and vernalization requirement, which are governed by relatively few loci with large effects in oats, dominated the phenotypic estimates of genetic distances (QMC). LeFort-Buson et al. (1986) obtained a moderate but nonsignificant correlation of 0.41 between pedigree-based distance and multivariate distance (Mahalonobis' D^2) in rapeseed. And, finally, Smith and Smith (1989) found no significant association between pedigree and morphological distance in maize. Most evidence suggests that marker-based estimates and pedigree-based estimates are in closer agreement with each other than with morphological-based estimates, and that analysis of quantitative morphological data may not provide a dependable or consistent measure of genetic divergence in the absence of extensive pedigree records (Smith et al. 1991).

Correlation of FD with heterosis and genetic variance

The only significant associations between *FD* and population genetic variances were small positive correlations between *FD* and the genetic variance for straw yield in 1983, the generalized genetic variance (GGV) in 1983, and the variance for plant height in the combined analysis (Table 4). When the two closely-related crosses were excluded from the analysis, only the genetic variance for plant height was significantly correlated with *FD* ($r = 0.35$ and 0.61, respectively). *FD* was not significantly correlated with heterosis in either 1983 or 1984. Thus, *FD* has limited utility for predicting genetic variance or heterosis among these lines. Likewise, *GD** and *DI* showed little or no association with genetic variance or heterosis in this set of materials (Cowen and Frey 1987a, b).

Several authors have suggested combining different types of distance estimates to predict heterosis or genetic variance (Cox et al. 1985; LeFort-Buson et al. 1986; Cox and Murphy 1990; Souza and Sorrells 1991c). Several different combinations of *FD* and *GD**

Table 4. Correlations of RFLP distance (*FD*), genealogical distance (*GD**), and multivariate distance (*DI*) with genetic variance (σ_G^2) for six traits, generalized genetic variance (*GGV*), and grain-yield heterosis

Estimate	Trait[a]	*FD*			*GD**			*DI*		
		1983[b]	1984	Comb.	1983	1984	Comb.	1983	1984	Comb.
σ_G^2	GY	0.05	0.00	−0.05	−0.07	−0.03	−0.11	−0.16	−0.20	−0.18
	BY	0.11	0.01	0.02	0.10	0.14	0.11	−0.11	−0.20	−0.17
	SY	0.32*	0.06	0.19	0.46**	0.25	0.36*	−0.20	−0.13	−0.18
	HI	0.08	0.25	0.07	−0.08	−0.02	−0.08	−0.20	0.08	−0.19
	HT	–	–	0.35**	–	–	0.40**	–	–	−0.14
	HD	–	–	0.02	–	–	0.22	–	–	−0.14
GGV		0.26	0.13	–	0.42**	−0.02	–	−0.36**	0.06	–
Heterosis	GY	−0.30	0.08	–	−0.33	−0.11	–	0.34	−0.08	–

*,** Correlation significant at $P = 0.10$ and 0.05, respectively

[a] GY = grain yield; BY = above-ground biological yield; SY = straw yield; HI = harvest index; HT = height; HD = heading date

[b] 1983 = experiment 3 for genetic variances and experiment 1 for heterosis; 1984 = experiment 4 for genetic variances and experiment 2 for heterosis; Comb. = combined analysis of variance from 1983 and 1984

were calculated and evaluated, but combined distance estimates were not associated with either genetic variance or heterosis any more than were *FD* or *GD** (data not shown).

Strong associations between RFLP or molecular-marker based genetic distance and heterosis or genetic variance will only occur under certain conditions. Several factors could influence this relation, including the number of probes or marker loci used in the analysis, linkage between RFLP loci and loci that govern trait expression (QTLs), the range of pedigree relationships among the parents, and differences in gene expression among populations or crosses.

The first attempts to predict heterosis from molecular data used a few, randomly-dispersed isozyme markers to estimate the percentage of heterozygous loci in hybrid progenies (Hunter and Kannenberg 1971; Heidrich-Sobrinho and Cordeiro 1975; Stuber 1989; Ehoibu et al. 1990). These studies observed only small positive correlations between heterozygosity at marker loci and heterosis or specific combining ability. One hypothesis regarding the lack of a strong association was that there were too few isozyme markers to sufficiently sample the entire genome; it was thought that the more numerous and polymorphic RFLP loci would give better results (Hunter and Kennenberg 1971; Lamkey et al. 1987). Several reports, however, have shown that increasing the number of arbitrarily-selected (dispersed) loci through the use of RFLP markers did not improve the association between marker-based distances and heterosis (Melchinger et al. 1990a, b).

More important than the absolute number of loci selected for analysis is the linkage relationships between marker loci and the QTLs of interest. As Melchinger et al. (1990b) and Smith et al. (1990) argue, an arbitrarily selected set of markers that covers the entire genome will not accurately predict heterosis if the QTLs are located only in certain regions of the genome. A more precise approach might be to preselect markers based on linkage relations with known QTLs. Charcosset et al. (1991) have shown theoretically that marker loci not linked to QTLs (i.e., "nonmarking" markers) and QTLs not marked by marker loci (i.e., "unmarked" loci) will play symmetrical roles and considerably reduce the correlation between heterosis and heterozygosity at marker loci. Bernardo (1992) arrived at similar conclusions from theoretical models and computer simulations, although his results show that the percentage of nonmarking markers will have a greater effect on the relationship than the percentage of unmarked loci. The inferences in our study were based on DNA polymorphism detected by 26 probes homologous to sequences presumably dispersed throughout the genome. Linkage relations between these sequences and QTLs are unknown, but the possible lack of linkage between them may explain the low correlation between *FD* and heterosis and genetic variance observed in this study.

However, even if marker loci are selected for their close proximity to a QTL, the relationship between heterozygosity at the molecular-marker loci (i.e., RFLP-based genetic distance) and heterosis or genetic variance will also depend on the extent of linkage disequilibrium between the marker loci and QTLs in the germplasm under consideration (Charcosset et al. 1991; Dekkers and Dentine 1991). To accuratetly predict heterosis or genetic variance from molecular data, each allele at a QTL must be linked to a unique marker allele. The validity of this assumption will vary with the reference population or the source of germplasm of the parents under consideration.

Pedigree relationships among a set of parents can affect the association between marker distance and heterosis or genetic variance in two ways. First, the pedigree of a set of parents can indicate whether or not the RFLP "alleles" are identical by descent (*ibd*). If they are *ibd*, the markers should be informative about that locus and linked regions. Among unrelated parents, the marker alleles may only be alike in state and tell little about their neighbors. Second, marker-based genetic distance estimates and heterosis are both positively correlated with genetic divergence within a set of related parents (Melchinger 1993). Thus, parents closely related by pedigree will produce marker-based genetic distance estimates that are small and will also produce progeny with little heterosis or genetic variance even if no marker loci are linked to loci governing the trait of interest. Melchinger (1993) compared several empirical studies in maize and noted that a close relationship between marker-based estimates of genetic distance and heterosis, SCA, or F_1-hybrid performance, was observed in a group of parents only if they showed some relationship by pedigree; the correlations were either low or nonexistent for unrelated parents ($f_{ij} < 0.1$).

Most of the crosses we examined (in this study) were among relatively-unrelated parents (f_{ij} between 0.05 and 0.25), although we did include two crosses between highly-related lines (f_{ij} = approximately 0.60). Crosses with a wider range and better distribution of pedigree relationships would be required to clearly establish the relationship between *FD*, *GD*, and heterosis or genetic variance in oats.

Finally, the poor association between marker-based estimates of genetic distances and heterosis or genetic variance in this study might be due to differences in the average degree of dominance in the F_1 hybrids or the average amount of variance contributed by each locus in the inbred populations (Melchinger et al. 1990b). Variation in gene expression among progeny could be due to different loci affecting yield in

different crosses, to the presence of multiple alleles per locus, or to epistasis.

Summary

FD was positively associated with *GD** in this set of nine oat lines. RFLP markers may be useful for providing an estimate of the genetic distance among breeding materials when pedigree information is not available. An example of the utility of this approach can be seen in this study, becuse it is impossible to trace the parentage of B605-1085. B605-1085 was a single-plant selection from a bulk of several crosses. Reliable RFLP data can also be used to establish cultivar identity and estimate minimum genetic distance for legal protection (Duesing and Raeber 1989; Smith and Smith 1989).

Accurate prediction of heterosis or genetic variance in oats by using arbitrarily-selected RFLP markers does not seem likely, based on the results of this study and on recent results presented in the literature. Even preselected markers (i.e., those in close proximity to QTLs) may not be asociated with heterosis or genetic variance unless the reference population is in linkage disequilibrium and gene expression is constant in all populations. Perhaps, the inferences should be limited to specific gene pools or germplasm groups. Further data in oats are needed regarding: (1) the inheritance of the variation in RFLPs detected by these probes, (2) linkage relationships among the markers and QTLs, and (3) the expression of QTLs in different oat populations.

The results of this study, along with other empirical and theoretical studies, provide guidance for conducting further research. The relationship between molecular marker-based distance and genetic divergence is best estimated from a large group of parents with a wide and continuous range of known pedigree relationships, breeding records, and performance data. Many dispersed probes would provide the best sampling of the entire genome and provide the most precise estimates. Conversely, estimating the relationship between RFLP-based distances and heterosis or genetic variance requires a set of unrelated parents, or a mating design that can account for pedigree relationships (Frei et al. 1986), and a set of probes selected for their linkage relationships with QTLs governing the traits of interest.

References

Atchley WR, Newmann S, Cowley DE (1988) Genetic divergence in mandible form in relation to molecular divergence in inbred mouse strains. Genetics 120:239–253

Bernardo R (1992) Relationship between single-cross performance and molecular marker heterozygosity. Theor Appl Genet 83:628–634

Bulmer MG (1957) Approximate confidence limits for components of variance. Biometrika 44:159–167

Charcosset A, Lefort-Buson M, Gallais A (1991) Relationship between heterosis and heterozygosity at marker loci: a theoretical computation. Theor Appl Genet 81:571–575

Cowen NM, Frey KJ (1987a) Relationship between genealogical distance and breeding behavior in oats (*Avena sativa* L.) Euphytica 36:413–424

Cowen NM, Frey KJ (1987b) Relationships between three measures of genetic distance and breeding behavior in oats (*Avena sativa* L.) Genome 29:97–106

Cox TS, Murphy JP (1990) The effect of parental divergence on F_2 heterosis in winter wheat crosses. Theor Appl Genet 79:241–250

Cox TS, Kiang YT, Gorman MB, Rodgers DM (1985) Relationship between coefficient of parentage and genetic similarity indices in the soybean. Crop Sci 25:529–532

Cruden D (1949) The computation of inbreeding coefficients for closed populations. J Hered 40:248–251

Dekkers JCM, Dentine MR (1991) Quantitative genetic variance associated with chromosomal markers in segregating populations. Theor Appl Genet 81:212–220

Duesing JH, Raeber JG (1989) Requirements of industrial firms for intellectual property. Votr Pflanzenzuchtg 16:253–269

Efron B (1982) The jackknife, the bootstrap and other resampling plans. Society for Industrial and Applied Mathematics, Philadelphia, USA

Ehoibu NG, Goddard ME, Taylor JF (1990) Prediction of heterosis in crosses between inbred lines of *Drosophila melanogaster*. Theor Appl Genet 80:321–325

Emik LO, Terrill CE (1949) Systematic procedures for calculating inbreeding coefficients. J Hered 40:51–55

Feinberg AP, Vogelstein B (1983) A technique for radiolabelling DNA restriction fragment length polymorphisms to high specific activity. Anal Biochem 132:6–13

Frei OM, Stuber CW, Goodman MM (1986) Use of allozymes as genetic markers for predicting performance in maize single-cross hybrids. Crop Sci 26:37–42

Graner A, Siedler H, Jahoor A, Herrmann RG, Wenzel G (1990) Assessment of the degree and type of restriction fragment length polymorphism in barley (*Hordeum vulgare* L). Theor Appl Genet 80:826–832

Heidrich-Sobrinho E, Cordeiro AR (1975) Codominant isoenzymic alleles as markers of genetic diversity correlated with heterosis in maize (*Zea mays* L.). Theor Appl Genet 46:197–199

Helentjaris T, King G, Slocum M, Siedenstrang C, Wegman S (1985) Restriction fragment length polymorphisms as probes for plant diversity and their development as tools for applied plant breeding. Plant Mol Biol 5:109–118

Hunter RB, Kannenberg LW (1971) Isozyme characterization of corn (*Zea mays* L.) inbreds and its relation to single-cross hybrid performance. Can J Genet Cytol 13:649–655

Jacquard A (1974) The genetic structure of populations. Translated by D. and B. Charlesworth. Springer-Verlag, New York

Lamkey KR, Hallauer AR, Kahler AL (1987) Allelic differences at enzyme loci and hybrid performance in maize. J Hered 78:231–234

Lee M, Godshalk EB, Lamkey KR, Woodman WW (1989) Association of restriction fragment length polymorphisms among maize inbreds with agronomic performance of their crosses. Crop Sci 29:1067–1071

Lefort-Buson M, Guillot-Lemoine B, Dattee Y (1986) Heterosis and genetic distance in rapeseed (*Brassica napus* L.). Use of different indicators of genetic divergence in a 7 × 7 diallel. Agronomie 6:839–844

Melchinger AE (1993) Use of RFLP markers for analysis of genetic relationships among breeding materials and prediction of hybrid performance. Int Crop Sci I (in press)

Melchinger AE, Lee M, Lamkey KR, Hallauer AR, Woodman WL. (1990 a) Genetic diversity for restriction fragment polymorphisms and heterosis for two diallel sets of maize inbreds. Theor Appl Genet 80:488–496

Melchinger AE, Lee M, Lamkey KR, Woodman WL (1990b) Genetic diversity for restriction fragment length polymorphisms: relation to estimated genetic effects in maize inbreds. Crop Sci 30:1033–1040

Melchinger AE, Messmer MM, Lee M, Woodman WL, Lamkey KR (1991) Diversity and relationships among U.S. maize inbreds revealed by restriction fragment length polymorphism. Crop Sci 31:669–678

Messmer MM, Melchinger AE, Lee M, Woodman WL, Lee EA, Lamkey KR (1991) Genetic diversity among progenitors and elite lines from the Iowa Stiff Stalk Synthetic (BSSS) maize population: comparison of allozyme and RFLP data. Theor Appl Genet 83:97–107

Nei M, Li W (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci USA 76:5269–5273

Saghai-Marroof MA, Soliman KM, Jorgenson R, Allard RA (1984) Ribosomal DNA spacer length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. Proc Natl Acad Sci USA 81:8014–8018

SAS (1988) SAS introductory guide for personal computers. SAS Institute, Inc., Cary, North Carolina, USA

Smith JSC, Smith OS (1989) The description and assessment of distances between inbred lines. II. The utility of morphological, biochemical, and genetic descriptors and a scheme for the testing of distinctiveness between lines. Maydica 34:151–161

Smith JSC, Smith OS, Bowen SL, Tenborg RA, Wall SJ (1991) The description and assessment of distances between inbred lines of maize. III. A revised scheme for the testing of distinctiveness between inbred lines utilizing DNA RFLPs. Maydica 36:213–226

Smith OS, Smith JSC, Bowen SL, Tenborg RA, Wall SJ (1990) Similarities among a group of elite maize inbreds as measured by pedigree, F_1 grain yield, heterosis, and RFLPs. Theor Appl Genet 80:833–840

Smith OS, Smith JSC, Bowen SL, Tenborg RA (1992) Numbers of RFLP probes necessary to show associations between lines. Maize Genet Coop Newslett 66:66

Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 98:503–517

Souza E, Sorrells ME (1989) Pedigree analysis of North American oat cultivars released from 1951 to 1985. Crop Sci 29: 595–601

Souza E, Sorrells ME (1991a) Relationships among 70 North American oat germplasms. I. Cluster analysis using quantitative clusters. Crop Sci 31:599–605

Souza E, and Sorrells ME (1991b) Relationships among 70 North American oat germplasms. II. Cluster analysis using qualitative characters. Crop Sci 31:605–612

Souza E, and Sorrells ME (1991c) Prediction of progeny variation in oat from parental genetic relationships. Theor Appl Genet 82:233–241

St. Martin SK (1982) Effective population size for the soybean improvement program in maturity groups 00 to IV. Crop Sci 22:151–152

Stuber CW (1989) Marker-based selection for quantitative traits. Votr Pflanzenzuchtg 16:31–49

Ward JH (1963) Hierarchical grouping to optimize an objective function. J Am Stat Assoc 58:236–244