Theor Appl Genet (1990) 79: 241–250

THEORETICAL AND APPLIED GENETICS

The effect of parental divergence on F_2 heterosis in winter wheat crosses*

T.S. Cox[1] **and J.P. Murphy**[2]

[1] USDA-ARS, Agronomy Department, Kansas State University, Manhattan, KS 66506, USA
[2] Crop Science Department, North Carolina State University, Raleigh, NC 27695-7629, USA

Received May 25, 1989; Accepted October 20, 1989
Communicated by A.R. Hallauer

Summary. In winter wheat (*Triticum aestivum* L.), the development of a methodology to estimate genetic divergence between parental lines, when combined with knowledge of parental performance, could be beneficial in the prediction of bulk progeny performance. The objective of this study was to relate F_2 heterosis for grain yield and its components in 116 crosses to two independent estimates of genetic divergence among 28 parental genotypes of diverse origins. Genetic divergence between parents was estimated from (a) pedigree relationships (coefficients of kinship) determined without experimentation, and (b) quantitative traits measured in two years of field experimentation in Kansas and North Carolina, USA. These distances, designated $(1-r)$ and G, respectively, provided ample differentiation among the parents. The 116 F_2 bulks were evaluated at four locations in Kansas and North Carolina in one year. Significant rank correlations of 0.46 ($P=0.01$) and 0.44 ($P=0.01$) were observed between G and grain yield and kernel number heterosis, respectively. Although $(1-r)$ was poorly associated with grain yield heterosis, G and midparent performance combined to account for 50% of the variation in F_2 yields among crosses when $(1-r)$ was above the median value, whereas they accounted for only 9% of the variation among crosses when $(1-r)$ was below the median. Midparent and $(1-r)$ had equal effects on F_2 grain yield ($R^2=0.40$) when G was greater than the median value. A breeding strategy is proposed whereby parents are first selected on the basis of performance per se and, subsequently, crosses are made between genetically divergent parents that have both large quantitative (G) and pedigree divergence $(1-r)$.

Key words: *Triticum aestivum* L. – Genetic diversity – Coefficient of kinship – Genetic distance

Introduction

Heterosis, defined herein as the deviation of the mean of the progeny of a cross from the parental mean, is a function of the number of loci at which the parents carry different alleles and the magnitude and net direction of the nonadditive effects within or between those loci in hybrid combinations (Cress 1966; Jinks 1983). In the absence of information on the combining ability of genotypes and because of the vagaries inherent in the expression of nonadditive gene effects, prediction of heterosis is limited to estimation of potential allelic differences by genetic divergence analysis.

In winter wheat (*Triticum aestivum* L.), no established patterns of genetic divergence [such as "heterotic groups" in maize (*Zea mays* L.)] are recognized. Therefore, the development of a methodology to predict heterosis from estimates of divergence, when combined with knowledge of mean genotype performance, could be beneficial for two purposes: (a) to predict F_1 hybrid or synthetic cultivar performance, and (b) to predict the mean of F_1 or later-generation bulk progeny as an aid in selection of crosses for inbred cultivar development (Busch et al. 1974; Cregan and Busch 1977, 1978).

We investigated two methods of estimating genetic divergence in winter wheat: (a) pedigree relationships, e.g., kinship coefficients (Malecot 1948) determined or postulated without experimentation, and (b) genotypic

---

* Paper No. 12162 of the Journal Series of the North Carolina Agricultural Research Service, Raleigh, NC 27695-7643, and Contribution No. 89-396-J of the Kansas Agricultural Experiment Station, Manhattan, KS 66506

differences for quantitative traits, determined from replicated experiments. Explicit use of kinship coefficients to predict heterosis in autogamous crops has been rare. In U.S. spring oat (*Avena sativa* L.) cultivars, Cowen and Frey (1987a) reported a significant correlation of genealogical distance between parents and genetic variance among progeny, but they found no correlation between heterosis and genealogical distance. In other studies, combinations of kinship and quantitative trait information were of more predictive value. Souza (1988) combined morphological and kinship information to account for a significant proportion of variation in specific combining ability for grain yield in oats. Lefort-Buson et al. (1986) combined either parental performance data or Mahalanobis (1936) generalized distances (D^2) between parental lines with kinship coefficients, to account for yield heterosis variation in rapeseed (*Brassica napus* L.). In a second study utilizing a larger data set, Lefort-Buson et al. (1987) reported good agreement between kinship coefficients and heterosis for several traits in one of two years.

Significant correlations have been reported between divergence estimates based on quantitative traits and yield heterosis in rapeseed (Lefort-Buson et al. 1986), wheat (Shamsuddin 1985), dry bean (*Phaseolus vulgaris* L.) (Ghaderi et al. 1984), peanut (*Arachis hypogaea* L.) (Arunachalam et al. 1984; Isleib and Wynne 1983), and tomato (*Lycopersicon esculentum* Mill.) (Maluf et al. 1983). Although no correlation coefficients were reported, Jatasra and Paroda (1983) and Ramanujam et al. (1974) also found associations between divergence and heterosis in wheat and mung bean (*Phaseolus aureus* Roxb.), respectively. No significant positive correlations were found between parental divergence and yield heterosis in studies on faba bean (*Vicia faba* L.) (Ghaderi et al. 1984) or oats (Cowen and Frey 1987b). Although the majority of these studies estimated genetic divergence between parents on the basis of Mahalonobis' distance (Mahalanobis 1936; Goodman 1972), euclidian distances based upon principal components were utilized by Isleib and Wynne (1983) and Cowen and Frey (1987b). Two studies (Lefort-Buson et al. 1986; Cowen and Frey 1987b) also calculated distances based upon combining ability in a diallel analysis after Hanson and Casas (1968) and Cervantes et al. (1978).

In this study, our objectives were (1) to relate F_2 heterosis for grain yield and its components in 116 winter wheat crosses to two independent estimates of divergence among 28 parental genotypes, and (2) to use divergence and midparent performance estimates to predict F_2 bulk grain yields. Coefficients of kinship were estimated from pedigree data and utilized to calculate pedigree divergence estimates, and quantitative traits were measured on parental lines in a series of experiments over two years and were utilized to calculate divergence estimates.

Materials and methods

One hundred sixteen two-way crosses were made using 28 hard red winter (HRW), soft red winter (SRW), and "international" (INT) wheat parents (Table 1). The first two genetically divergent classes (Murphy et al. 1986) comprised winter wheat cultivars grown commercially in the central and eastern United States, excepting KS831957, a HRW germ plasm release. The INT parents comprised cultivars and breeding lines selected from the 1984 International Winter Wheat Performance Nursery to represent a geographically and genetically diverse sample of germ plasm. Although only HRW and SRW are classes of wheat as defined by the U.S. Federal Grain Inspection Service, we will refer to all three parent categories as "classes."

Genetic distance estimates between parents

Coefficients of kinship (r) (Malecot 1948) for the 116 parental combinations were generated from complete pedigrees as described by Cox et al. (1986). Genetic distances were computed as $(1-r)$.

For computation of genetic distances based upon quantitative traits, the 28 parents were evaluated in harvest years 1985 and 1986 near Raleigh/NC and near Manhattan (1985) and Hutchinson (1986)/KS, USA. Kansas and North Carolina represent environments where hard and soft red winter wheats, respectively, are grown commercially. Experiments were arranged in six randomized complete blocks at both locations in 1985, and six and three replicates at Kansas and North Carolina, respectively, in 1986. At Raleigh, plots consisted of a single 2.5-m row with 25 cm between plots in 1985, and four rows, each 3.05 m long, with 25 cm between rows in 1986. At the Kansas locations, plots consisted of three rows, each 1 m long, with 18 cm between rows in plots and 25 cm between plots. Recommended fertility and cultural practices at each location were followed and, in addition, two applications (GS-37 and 50) (Zadocks et al. 1974) of Bayleton 1.8 EC (triadimefon, 125 g active ingredient per hour), a broad spectrum fungicide, were administered to three replicates in both years in North Carolina. Unsprayed replicates in North Carolina in 1985 were utilized for disease assessment only.

Data on decimal growth-stage codes (Zadoks et al. 1974) and ten morphological traits were recorded for all 28 entries. The trait-environment combinations used in the analysis may be seen in Table 2. Growth-stage codes were recorded weekly between resumption of growth in the spring and harvest. In autumn, when plots were near GS 22, entries were assigned growth-habit scores of 1 (prostrate), 2 (intermediate), or 3 (upright). At GS 83, plant height was recorded as the distance between the ground surface and the tip of the heads. Then ten flag leaves were selected at random, and the mean distance between the ligules and leaf tips (flag leaf length) and width at the widest point (flag leaf width) was recorded.

Six heads (without peduncles) were pulled from each plot just prior to harvest; the number of spikelets, number and weight of kernels per head, and weight per thousand kernels were recorded. Kernels were ground, and protein concentration of the whole flour was determined with a Technicon near-infrared analyzer, calibrated with checks of known Kjeldahl N concentration every 25 samples.

Plants were cut at ground level when all plots had reached full harvest maturity. Grain weight per plot was determined, and harvest index was calculated from a 0.3-m subsample as grain weight divided by total plant weight. Number of kernels per square meter (in thousands) was estimated as grain weight per square meter divided by weight per thousand kernels.

Table 1. Parent class, geographic origin, number of crosses in which utilized, and agronomic characteristics recorded in Kansas and North Carolina in the 1986–87 seasons

Class[a]	Parent	Geographic origin	No. crosses	Grain yield ($g\ m^{-2}$)	1,000 kernel wt (g)
HRW	Arkan	Kansas	11	369	26.9
	Chisholm	Oklahoma	9	525	31.8
	Hawk	Private	8	400	31.2
	KS 831957	Kansas	8	443	30.4
	Larned	Kansas	7	362	29.3
	Newton	Kansas	7	399	27.5
	Rose	S. Dakota	7	339	24.5
	Roughrider	N. Dakota	6	293	27.2
	TAM 107	Texas	10	416	33.0
	TAM 108	Texas	11	403	27.7
SRW	Arthur 71	Indiana	10	414	30.7
	Adena	Ohio	12	409	24.8
	Auburn	Indiana	9	406	24.6
	Blueboy	N. Carolina	8	427	29.2
	Compton	Indiana	7	457	31.1
	Florida 302	Florida	10	489	33.9
	Hart	Missouri	7	438	32.4
	Pike	Missouri	10	450	29.8
	Saluda	Virginia	4	499	28.7
	Scotty	Illinois	9	442	29.1
INT	Bezostaya 1	USSR	8	329	33.2
	CA 8055	China	9	443	33.4
	Feng Kang 15	China	8	494	38.6
	Fundulea 262	Romania	8	328	29.5
	Katya A1	Bulgaria	8	479	29.8
	Kosutka	Czechoslovakia	7	407	28.0
	Norman	G. Britain	4	172	25.1
	TAW-12399-75	FRG	10	337	29.6
Mean				403	29.8
LSD (0.05)				64	1.9

[a] Hard red winter (HRW) and soft red winter (SRW) are classifications used by the U.S. Federal Grain Inspection Service (FGIS). All other parents were classified as international (INT), which is not an FGIS class

Five additional traits were evaluated at Manhattan in 1985: resistance to soil-borne mosaic virus (scale of 1 to 3), resistance to *Septoria* leaf blotch (causal agent *Septoria tritici*) (scale 1 to 9), incidence of leaf rust (causal agent *Puccinia recondita*) (scale of 1 to 9), size of leaf rust pustules (scale of 1 to 9), and winter hardiness (scale of 1 to 3). Resistance to *Septoria* glume blotch (causal agent *Leptosphaeria nodorum* Muller) was evaluated in North Carolina in 1985 on the three replications receiving no fungicide application. In all cases, lower scores indicated greater resistance or hardiness.

Traits were divided into four sets: stage of growth (Z), vegetative (V), reproductive (R), and resistance (P) (Table 2). Set names were not completely descriptive because of interrelationships between vegetative and reproductive growth (e.g., in determining height or harvest index). Significance of differences among parents for trait sets within environments was tested by multivariate analyses of variance. Principal component analysis, based on the correlation matrix for entry means, was conducted for each set of traits as evaluated in Kansas and North Carolina in each year. Distances between all pairs of parents, i, and j, were computed according to Goodman (1972) for each trait set by environment combination as

$$g_{su}=\left[\sum_{k=1}^{q}(y_{ik}-y_{ik})^2\,\lambda_k{}^{-1}\right]^{1/2}$$

where $g=Z$, V, R, or P from growth stage, vegetative, reproductive, or resistance trait sets, respectively; $s=N$ or K for North Carolina or Kansas; $u=1$ or 2 for 1985 or 1986; y_{ik} and y_{jk} were the k^{th} principal component scores for entries i and j; λ_k was the k^{th} eigenvalue; and q was the number of eigenvalues equal to or greater than 1.0.

Distances for the two states (N and K) combined over trait sets and years and for trait sets (Z, V, and R) combined over states and years as well as a distance combined over all traits, states, and years (G) were computed as euclidian distances from component distances:

$N=[2Z^2{}_{N1}+2V_{N1}^2+R_{N1}^2+R_{N2}^2]^{1/2}$;
$K=[Z_{K1}^2+Z_{K2}^2+V_{K1}^2+V_{K2}^2+R_{K1}^2+R_{K2}^2]^{1/2}$;
$Z=[2Z_{N1}^2+Z_{K1}^2+Z_{K2}^2]^{1/2}$;
$V=[2V_{N1}^2+V_{K1}^2+V_{K2}^2]^{1/2}$;
$R=[R_{N1}^2+R_{N2}^2+R_{K1}^2+R_{K2}^2]^{1/2}$; and
$G=[2Z_{N1}^2+2V_{N1}^2+R_{N1}^2+R_{N2}^2+Z_{K1}^2+Z_{K2}^2+V_{K1}^2+V_{K2}^2+$
$+R_{K1}^2+R_{K2}^2+P_{K1}^2]^{1/2}$;

where subscripts $N1$, $N2$, $K1$, and $K2$ refer to North Carolina in 1985 and 1986 and Kansas in 1985 and 1986, respectively. These formulas gave each state or trait group equal weighting within a given distance measure. Because of seed mixtures in some plots of "Adena", "Katya A1", "TAW-12399-75", and "Norman" in

Table 2. Eigenvectors with variable weightings from principal component analysis for four trait groups. Data collected from 28 wheat parents in Kansas and North Carolina in harvest years 1985 and 1986

Trait set	Trait	N. Carolina (N)				Kansas (K)				
		PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2	PC3
Growth stage (Z)		Z_{N1}				Z_{K1}		Z_{K2}		
	Date 1[a]	0.35	−0.24			0.02	−0.26	0.02	−0.15	0.96
	2	0.35	−0.22			0.28	−0.25	0.22	−0.67	−0.04
	3	0.37	−0.02			0.30	−0.27	0.32	−0.43	−0.07
	4	0.37	−0.05			0.27	−0.33	0.34	−0.25	−0.09
	5	0.37	−0.07			0.30	−0.28	0.35	−0.06	−0.16
	6	0.35	−0.08			0.32	−0.20	0.35	0.22	0.07
	7	0.32	0.23			0.33	−0.14	0.35	0.17	0.16
	8	0.34	0.24			0.33	0.03	0.35	0.19	−0.08
	9	0.10	0.88			0.32	0.28	0.34	0.31	0.07
	10					0.27	0.43	0.35	0.26	0.05
	11					0.29	0.36			
	12					0.28	0.39			
		(0.74)[b]	(0.13)			(0.67)	(0.13)	(0.72)	(0.12)	(0.10)
Vegetative (V)		V_{N1}				V_{K1}		V_{K2}		
	Growth habit	0.55	0.11			0.66	−0.14			
	Height	0.70	−0.04			0.36	0.63	0.74	0.25	
	Flag length	0.30	0.70			0.34	0.51	0.12	0.86	
	Flag width	−0.35	0.70			−0.56	0.56	−0.67	0.44	
		(0.35)	(0.32)			(0.39)	(0.25)	(0.44)	(0.38)	
Reproductive (R)		R_{N1}		R_{N2}		R_{K1}		R_{K2}		
	Head weight			−0.21	0.73			0.61	0.25	
	Spikelets/head	0.02	0.82	0.54	0.22	0.52	0.30			
	Kernel weight	0.60	0.42	0.37	0.59	−0.65	0.21	0.22	0.76	
	Kernel no.	0.46	−0.36	−0.58	0.14	−0.41	0.56	0.53	−0.36	
	Protein conc.					0.08	−0.74	0.20	−0.47	
	Harvest index	0.65	−0.10	0.44	0.24	0.37	0.00	0.50	−0.07	
		(0.47)	(0.32)	(0.51)	(0.30)	(0.38)	(0.29)	(0.44)	(0.28)	
Disease resistance (P)						P_{K1}				
	Soil-borne mosaic					0.22	0.78			
	Leaf rust (pustule no.)					0.57	−0.29			
	Leaf rust (pustule size)					0.60	−0.31			
	Septoria tritici					0.51	0.35			
	S. nodorum					0.03	0.29			
						(0.46)	(0.29)			

[a] Dates range from mid-March to mid-June in both North Carolina and Kansas

[b] Proportion of total variation accounted for by each component

1985–86, quantitative distances involving those cultivars were not computed.

Estimation of F_2 heterosis

The 116 F_2 bulks (crosses) and their 28 parents were evaluated in 1987 near Raleigh and Plymouth/NC and Manhattan and Hutchinson/KS. Each of the four experiments comprised 28 blocks, each block containing three plots of one parental line and one plot each of the crosses in which that line was a parent. Thus, each parent and its crosses were included in exactly one block per experiment. Plots were managed as in the 1985 and 1986 experiments. Plots comprised two rows, each 1.23 m long, with 0.3 m between rows in North Carolina, and three rows, each 1.5 m long, with 18 cm between rows and 25 cm between plots in Kansas. Grain yields in grams per square meter and kernel weight (determined for 200 kernels and expressed as grams per thousand kernels) were measured for each plot; number of kernels per square meter was calculated as in 1985 and 1986.

This experimental design was chosen to provide precise estimates of midparent heterosis values and to allow comparison of heterosis and F_2 mean values with external divergence estimates. It was not designed for comparison of different cross means or combining ability effects. In this design, the value, x_{ijk}, of a variable measured on a plot representing the hybrid between parents i and j in the block containing parent i was expressed as a linear function:

$$x_{ijk} = m + p_i + p_j + h_{ij} + b_i + e_{ik}.$$

The value for the plot of the same cross lying in the block with parent j was expressed as:

$$x_{jil} = m + p_i + p_j + h_{ij} + b_j + e_{jl},$$

where m was the experiment mean, p_i and p_j were general effects of parents i and j, h_{ij} was heterosis between i and j, b_j was the environmental effect of block j, and e_{jl} was the environmental error effect of plot l within block j. Likewise, the mean of the three plots of parent i in an experiment was:

$$\bar{x}_{ii} = m + 2p_i + b_i + \bar{e}_i,$$

where $\bar{e}_i$ was the mean over the three within-block errors e_{ik} for the parent plots.

Heterosis was estimated as:

$$\hat{h}_{ij} = (x_{ij} + x_{ji})/2 - (\bar{x}_{ii} + \bar{x}_{jj})/2 = h_{ij} + (e_{ik} + e_{jl} - \bar{e}_i - \bar{e}_j)/2,$$

where $\bar{x}_{ii}$ and $\bar{x}_{jj}$ were means over the three plots of parents i and j, respectively, and $\bar{e}_i$ and $\bar{e}_j$ were their respective mean within-plot errors. Therefore, heterosis effects were not confounded with block effects.

Assuming that within-block error deviations had the same normal distributions within all blocks and that error effects on parental and F_2 means were similar, the variance of a heterosis estimate within an experiment was computed as:

$$\sigma_h^2 = (2\sigma_e^2 + 2\sigma_e^2/3)/4 = 2\sigma_e^2/3,$$

where σ_e^2 was pooled within-parent variance, with 56 *df* in each experiment.

An analysis of variance including all experiments was performed on heterosis values for each trait. The significance of experiment × cross interaction was determined with the F-statistic using σ_h^2, pooled over experiments, in the denominator. The variance of a mean heterosis value over the four experiments was computed as the experiment × cross mean square multiplied by $(2/3 \times 1/4) = 1/6$. This interaction mean square was used even when not significant, because it was computed directly from heterosis values, unlike σ_h^2, which was computed from parent means.

Mean heterosis values were computed for groups of crosses based on parent groups (all combinations of HRW, SRW, and INT) and on high or low distance [$(1-r)$ and G]. Spearman rank correlations (Snedecor and Cochran 1980) were computed among distance estimates and between distance and heterosis estimates. Standardized partial regression coefficients (Snedecor and Cochran 1980) were computed from multiple regression of 1987 cross means on 1985–86 parent means and G or parent means and $(1-r)$. Regression and correlation analyses involving quantitative (Goodman-euclidian) distances included only 85 of the 116 crosses, because such estimates were not available for crosses involving Adena, Katya A1, TAW-12399-75, or Norman.

Results

Genetic divergence between parents

Genetic distance based upon pedigrees $(1-r)$ between parental lines varied from 0.25 for the HRW cultivars "Larned" and "TAM 107" to 1.0 for a majority of the distances between the British cultivar "Norman" and the rest of the parents. Indeed, few of the distances between INT parents and U.S. parents were less than 0.9; distances within the INT group also tended to be of this magnitude. It is noteworthy that the $(1-r)$ distance estimate extended over three-quarters of the total possible range for this variable.

Multivariate analyses of variance indicated that genotype effects among parents for all trait groups (i.e., Z, V, R, and P) were significant at all locations in both 1985 and 1986. The first two principal components accounted for more than 79% of the variation among parents in each of the three growth-stage trait groups (Z) (Table 2). Overall similarity was evident in both the North Carolina and Kansas data sets. Weightings on the first principal components were similar over most dates, differentiating earlier- and later-developing genotypes. Weightings on the second principal component described differences in rates of development.

Eigenvalues summed over the first and second principal components for vegetative trait groups (V) were proportionally less than for Z groups, yet over 60% of variation was distributed along these axes. Plant height, growth habit, and flag leaf width were accorded the highest weightings on the first component axes. Plant height was accorded heavier weightings in Kansas than in North Carolina on the second component axes, whereas flag leaf measurements were most important in both states.

Eigenvalues summed over the first two principal components for reproductive trait groups from North Carolina (R_{N1} and R_{N2}) accounted for more than 78% of the variation in those traits, whereas those from Kansas (R_{K1} and R_{K2}) were greater than 66%. Some lack of balance in data sets from each state contributed to the greater contrasts in relative variable weightings between the two states for this trait group (R) than for the previous trait groups (Z or V). However, the relative importance of spikelets per head, kernel weight, and harvest index was noticeable in parent differentiation based on R_{N1} and R_{K1}. Protein concentration was an important variable in both R_{K1} and R_{K2}.

Seventy-five percent of the variation among parents was accounted for by the first two principal components in the resistance trait group (P). Differentiation occurred on the basis of the three leaf rust and *Septoria tritici* resistance measurements on the first principal component axis and of the soil-borne mosaic virus resistance on the second component axis. In Kansas, where disease reactions and other traits were measured for the same plots, levels of disease were not high enough to have significant effects on other traits, except that a few entries highly susceptible to soil-borne mosaic and winter injury may have had their development slowed in 1985 by the virus.

Both the kinship coefficient and the empirical, quantitative distances resulted in ample differentiation among the 28 parents in the study. Kinship methods agreed with intuitive observations concerning relationships through a single value, $(1-r)$, thought to reflect diversity at the genomic level. For quantitative methods, the sums of eigenvalues over the first two or three principal components in all trait groups suggested effective descrimination among parents; these four sets of diverse traits should also reflect the activity of loci distributed over much of the genome. Overall, there was concordance among the discriminative variables in both states over both years.

F_2 heterosis

Mean 1987 parental performances in Kansas and North Carolina for three agronomic traits are presented in Table 1. The highest yielding parent overall was "Chisholm" (HRW), followed closely by "Saluda" (SRW), "Feng Kang 15" (INT), and "FL 302" (SRW). Feng Kang 15, "CA 8055" (INT), and "Bezostaya 1" (INT), respectively, together with FL 302 and "TAM 107" (HRW), had the largest kernel weights. The late-maturing British parent, Norman (INT), had the lowest grain yield and kernel weight in the study; it was the only parent that could be classified as unadapted to the test environments on the basis of maturity and performance. The 116 F_2 bulk entries represented all six possible combinations among SRW, HRW, and INT genotypes (Table 3). Significant variation in heterosis for grain yield and kernel number was observed among the six groups (HRW × HRW, SRW × SRW, INT × INT, HRW × SRW, HRW × INT, and SRW × INT) and among crosses within each of the six groups. Less significant variation in heterosis for kernel weight was observed. First-order interaction means involving location effects, with one exception, were not significant; this was an unexpected result considering the geographical distances among the four test locations. This stability may reflect high levels of buffering within each F_2 bulk entry; if so, buffering effects were expressed equally in nearly all groups and over all locations.

Mean intragroup genetic distance estimated from pedigrees $(1-r)$ was 0.93 (Table 4); the HRW × HRW and SRW × SRW groups had the lowest $(1-r)$. The overall mean of the Goodman-euclidian genetic distance estimated G was 6.72; the SRW × SRW group had the lowest intragroup diversity, whereas the HRW × INT and SRW × INT groups had the highest.

For grain yield, the INT × INT and HRW × INT groups exhibited the greatest frequencies of significantly positive heterosis (Table 4). The presence or absence of significant heterosis for yield was not related to large differences in mean midparent values of the six groups, which had a range of only 8% of the overall mean midparent.

Table 3. Combined analyses of variance for grain yield heterosis (HGY), kernel number heterosis (HKN), and kernel weight heterosis (HKW) among 116 F_2 bulk populations grown at four locations in harvest year 1987

Source	*df*	Mean squares		
		HGY	HKN	HKW
		$\times 10^{-3}$	$\times 10^{-3}$	
Locations	3	1,733**	11,191**	0.28
Groups	5	1,069*	17,816**	6.04
Entries within groups	110	567**	6,495**	7.39**
HRW × HRW	11	715**	7,664**	3.03
SRW × SRW	11	518**	3,991**	19.20**
INT × INT	10	828**	9,136**	12.40**
HRW × SRW	37	480**	6,942**	2.72
HRW × INT	21	550**	6,549**	11.05**
SRW × INT	20	563**	5,027**	4.96*
Location × groups	15	240	1,832	2.33
Location × entries within groups	330	237**	1,968	2.60
Pooled error	224	177	1,772	1.51

*,** Significant at $P=0.05$ and $P=0.01$, respectively

Mean heterosis over all groups and locations was 2.7% for grain yield (Table 4). No positive heterosis was observed for any group at Raleigh, but the SRW × SRW group, with low parental diversity, displayed significantly negative heterosis. At Plymouth, where overall heterosis was 1.8%, this group also displayed the highest negative heterosis level. Significant positive heterosis at Plymouth was observed in INT × INT and HRW × INT, two groups containing hybrids developed from diverse germ plasm not grown commercially in North Carolina. The largest mean heterosis for grain yield (6.1% overall) was observed at Manhattan, where the three groups having INT parentage as well as the largest intragroup genetic diversity [based upon both $(1-r)$ and G] averaged between 8% and 15% heterosis for yield. At Hutchinson, where overall heterosis was 3.7%, the greatest contrast occurred between the most genetically narrow group, SRW × SRW, with heterosis of −2.4%, and the highly diverse INT × INT group, with 11.4% heterosis.

The expression of grain yield heterosis in the F_2 bulks was highly influenced by location effects, but groups of crosses tended to rank similarly whatever the mean heterosis level. Most frequently, significant mean heterosis for groups was positive; the only significantly negative value occurred at Raleigh, where no significantly positive heterosis was recorded. The highest frequency of significantly positive heterosis for grain yield occurred at Manhattan, where intragroup genetic diversity and heterosis also were most closely allied.

Mean heterosis over all groups and locations was −4.2% and 6.6% for kernel number and weight, respectively (Table 4). Variation among group means for kernel weight heterosis was not significant, yet the overall means suggest that yield component compensation occurred between number and weight of kernels. The largest negative heterosis for kernel number again occurred in the

Table 4. Group mean heterosis for grain yield (HGY) at four locations and combined over locations, and combined heterosis for kernel number (HKN) and kernel weight (HKW); mean intragroup genetic distances [kinship $(1-r)$ and Goodman-Euclidian (G)] and rank correlations of distances with HGY

Group	No. of crosses	Grain yield heterosis (HGY)					Kernel no. heterosis	Kernel weight heterosis	Genetic distance			
		Raleigh	Plymouth	Manhattan	Hutchinson	All loc.	All locations		Mean		Rank correlation with HGY	
									$(1-r)$	G	$(1-r)$	G
		$g\ m^{-2}$					kernels m^{-2}	g				
HRW × HRW	12	0.2	18.9	2.7	14.5	9.1	−595**	1.9**	0.82	6.25	−0.06	0.43
		(0)[a]	(4.7)	(0.7)	(4.0)	(2.3)	(−4.7)	(6.6)				
SRW × SRW	12	−27.5*	−18.5	9.5	−2.7	−9.8	−1,564**	2.4**	0.86	5.74	0.31	−0.24
		(−7.2)	(−4.3)	(2.1)	(−0.7)	(−2.4)	(−12.6)	(8.4)				
INT × INT	11	−6.3	42.9**	58.6**	39.0**	33.5**	340	2.0**	0.96	6.86	0.15	0.10
		(−1.7)	(9.3)	(15.2)	(11.4)	(8.6)	(2.7)	(6.4)				
HRW × SRW	38	1.4	−3.0	20.8**	9.6	7.2*	−557**	1.7**	0.93	6.42	−0.24	0.50**
		(0.3)	(−0.7)	(4.8)	(2.4)	(1.7)	(−4.2)	(5.8)				
HRW × INT	22	−2.5	32.3**	31.2**	18.8**	20.0**	−315**	2.3**	0.98	7.32	0.21	0.44
		(−0.7)	(7.3)	(7.7)	(5.3)	(5.0)	(−2.5)	(7.6)				
SRW × INT	21	−4.0	−8.1	31.3**	10.0	7.9	−566**	1.8**	0.98	7.41	0.11	0.63*
		(−1.1)	(−1.8)	(7.6)	(2.6)	(1.9)	(−4.5)	(5.9)				
All groups	116	−4.1	3.8	25.4**	13.6**	9.7**	−536**	2.0**	0.93	6.72	0.09	0.46**
		(−1.1)	(1.8)	(6.1)	(3.7)	(2.7)	(−4.2)	(6.6)				
		F_2 mean										
All groups	116	393.5	454.2	443.1	394.4	421.3	13,529	31.7				

[a] Heterosis as percent of 1987 midparent in parentheses

*,** Significant at $P=0.05$ and $P=0.01$, respectively

SRW × SRW group, whereas the INT × INT group was unique in combining positive heterosis for kernel weight with non-negative kernel number heterosis to produce the highest grain yield heterosis of any group.

Associations between heterosis and genetic divergence estimates were evaluated by rank correlations (Table 4) because, even though all heterosis values were normally distributed, $(1-r)$ and G were not. A significant rank correlation of 0.46 ($P=0.01$) was observed between G and grain yield heterosis over 85 crosses. All correlations between grain yield heterosis and less comprehensive distance measures – N (0.30), K (0.50), Z (0.41), V (0.34), and R (0.25) – were significant, but the correlation with P (0.03) was not. All correlations between kernel number heterosis and G (0.44), N (0.29), K (0.52), Z (0.41), V (0.35), and R (0.21) were also significant. Conversely, kernel weight heterosis had a significant correlation only with the K genetic distance measurement, and it was small and negative (−0.25). Heterosis for grain yield or kernel number was uncorrelated with the $(1-r)$ distance measure, but a small, highly significant correlation of 0.18 was recorded between kernel weight heterosis and $(1-r)$. All correlations of $(1-r)$ with G, N, K, Z, V, R, and P were low and nonsignificant.

The association between G and heterosis was not entirely due to differences between groups; rank correlations within groups were similar to the overall value, except in the SRW × SRW and INT × INT groups (Table 4). Correlations between G and grain yield heterosis within the HRW × SRW and SRW × INT groups were significant and positive. Two other groups had nonsignificant but relatively large positive correlations; the SRW × SRW group had the only negative estimate. None of the within-group correlations between heterosis and $(1-r)$ were significant.

Of the 27 F_2 bulks that had positive mean heterosis for grain yield over the four test locations (Table 5), only five had G-distances below the mean G for their respective groups (Table 4). There were 15 crosses with significantly negative heterosis (data not shown). Of the nine negatively heterotic crosses for which we had G-distances, seven were either intraclass crosses or had G-distances less than 4.3.

Both positively heterotic HRW × HRW crosses involved TAM 107 (Table 5). "TAW-12399-75" from the Federal Republic of Germany was the most frequent INT parent forming heterotic combinations with the HRW class. The SRW parents involved in heterotic crosses con-

Table 5. Goodman-euclidian distance (G) between parents and heterosis values for the 27 F_2 bulks that displayed significant positive heterosis over four locations in 1987

Group	G	Grain yield heterosis	
		g m^{-2}	% of midparent
HRW × HRW			
TAM 107/Rose	9.3	106	28
TAM 107/TAM 108	7.4	59	16
SRW × SRW			
Scotty/Auburn	5.6	66	15
Pike/Blueboy	6.0	43	10
INT × INT			
Feng Kang 15/Norman	–	108	32
Fundulea 262/Feng Kang 15	7.5	82	20
Fundulea 262/CA 8055	6.3	63	16
Fundulea 262/Kosutka	6.9	63	17
Bezostaya 1/CA 8055	6.6	43	11
HRW × SRW			
Pike/TAM 107	9.0	93	22
TAM 108/FL 302	5.2	68	15
Rose/Arthur 71	7.3	62	16
Roughrider/Blueboy	8.3	62	17
TAM 108/Arthur 71	6.0	46	11
Pike/Arkan	7.4	41	10
HRW × INT			
Rose/TAW 12399-75	–	89	26
TAM 108/TAW 12399-75	–	74	20
TAM 107/Bezostaya 1	8.2	68	18
Roughrider/TAW 12399-75	–	52	17
Kosutka/Rose	7.5	46	12
Newton/TAW 12399-75	–	41	11
SRW × INT			
Auburn/CA 8055	8.5	84	20
Blueboy/CA 8055	9.0	63	15
Auburn/Feng Kang 15	11.3	52	12
TAW 12399-75/Hart	–	50	13
TAW 12399-75/Pike	–	40	10
Fundulea 262/Hart	6.4	37	10

Table 6. Mean grain yield, mean heterosis for grain yield, and standard partial regression coefficients (b) and coefficients of determination (R^2) from regression of cross mean grain yield on G and 1985–86 midparent yield, or on $(1-r)$ and 1985–86 midparent yield

Statistic	For crosses having		
	$(1-r)$ <0.96	$(1-r)$ >0.96	All $(1-r)$
Mean grain yield ($\bar{Y}$)			
for $G<6.6$	419	417	418
for $G>6.6$	428	452	445
Mean heterosis for Y:			
for $G<6.6$	−1	5*	2
for $G>6.6$	25**	26**	26**
$b_{Y \cdot G}$[a]	0.17	0.67**	0.44**
$b_{Y \cdot MP}$[b]	0.11	0.40**	0.31**
R^2	0.09	0.50	0.23
	$G<6.6$	$G>6.6$	All G
$b_{Y \cdot (1-r)}$[c]	0.06	0.46**	0.23**
$b_{Y \cdot MP}$[d]	0.10	0.51**	0.22**
R^2	0.01	0.40	0.09
No. crosses	42	43	85

[a, b] Standard partial regression coefficients for grain yield on G and 1985–86 midparent yield in a model containing only those two independent variables

[c, d] Standard partial regression coefficients for grain yield on $(1-r)$ and 1985–86 midparent yield in a model containing only those two independent variables

stituted germ plasm from five different breeding programs distributed throughout the SRW production area. There were only two heterotic SRW × SRW crosses. Blueboy, a parent one of them, has 50% HRW parentage; it grouped with HRW wheats in a cluster analysis based on $(1-r)$ (McCammon and Wenninger 1970; data not shown). The frequency of the Chinese parents "Feng Kang 15" and "CA 8055" in heterotic combinations was notable in the SRW × INT group. Among the INT parents, Chinese parents appeared frequently in heterotic combinations, as did the Romanian entry "Fundulea 262".

Five of the 11 INT × INT crosses, or 45%, had positive heterosis, whereas the proportion of heterotic crosses in other groups ranged from 16% to 29%. Heterosis expressed as a percentage of 1987 midparent yield was within a similar range for all groups, the seven highest values being 20% or greater.

Prediction of F_2 mean yield

Although $(1-r)$ was a poor predictor of heterosis, estimates of $(1-r)$, G, and 1985–86 midparent yields were all important in predicting 1987 F_2 mean yields. Examination of crosses that fell above or below the median values of $(1-r)$ and G showed that heterosis and F_2 means were consistently higher for crosses with $G>6.6$ (Table 6). However, the standard partial regression coefficients (b) and coefficient of determination (R^2) for regression of mean F_2 grain yield on the two independent variables G and 1985–86 midparent value were larger in crosses between parents with higher $(1-r)$ (Table 6). The G-distance between parents had a larger direct effect on F_2 yield than did the parents' mean yield in these more diverse crosses.

For crosses with G-distances greater than the median value, $(1-r)$ and midparent had similar effects on mean grain yield, with an R^2 of 0.40 (Table 6). However, the R^2 value was close to zero for crosses with G-distances less than 6.6. In multiple regression of mean yield on the three independent variables $(1-r)$, G, and midparent, $(1-r)$ did not have a significant effect (data not shown).

Discussion

An unique feature of this study, in comparison with other studies of the relationship between heterosis and genetic distance, is that morphological distances and heterosis were estimated in different years. Furthermore, our test locations were diverse in climate and soil type. For example, mean 1985–87 rainfall amounts at Plymouth, Raleigh, Manhattan, and Hutchinson were 1,170, 800, 990, and 940 mm, respectively. Mean January temperatures were 4.0°, 2.4°, −1.0°, and −1.0 °C, while mean April temperatures were 15.2°, 15.7°, 14.5°, and 13.7 °C, respectively. Soils at experimental sites ranged from Aquults and Udults, highly weathered soil from humid environments typical of eastern North Carolina, to Ustolls, dry, slightly weathered prairie soils, typical of the eastern Great Plains in Kansas. Therefore, the relationships we found between G and heterosis are more likely to have been reduced than inflated by genotype × environment interaction.

In computing distances from morphological data, we arranged correlated traits into groups, where a large portion of the variation was accounted for by a few principal component axes. This is in agreement with the conditions outlined by Goodman (1972) for the proper use of his diversity estimate. Furthermore, correlations between principal component scores for different trait groups were low and generally nonsignificant (data not shown); this justified our computation of the distances N, K, and G from their components as in euclidian distances, ignoring correlations among components.

We estimated heterosis in a single year over four locations. Even though the locations covered an extensive geographical range, genotype × location interaction for heterosis was small, slightly exceeding the 5% significance level only for grain yield (Table 3). Despite the large differences in mean heterosis, the relationship between G and heterosis was very consistent between states. The rank correlations between G and grain yield heterosis, for example, were 0.37 ** and 0.36 ** for North Carolina and Kansas locations, respectively, compared with the overall correlation of 0.46 **.

Our results suggest that G or any of its components N, K, Z, V, or R are superior to the pedigree-based estimate $(1-r)$ as estimators of F_2 heterotic potential in these winter wheat genotypes. However, the predictive value of morphological data depended on the environment in which they were obtained. Distance data collected on parental lines over two years in Kansas were highly significantly correlated with heterosis for grain yield (0.50) and kernel number (0.52); comparable associations from North Carolina were also highly significant, although lower (0.30 and 0.29, respectively). Among the distances based on individual trait groups, Z, which was computed from growth-stage data, was most closely associated with heterosis. The only positive correlation between $(1-r)$ and heterosis for any of the three traits involved kernel weight, a trait, incidentally, with which G had little association. Thus, $(1-r)$ and G were measuring different facets of heterotic potential, and G was much more effective in the process. Lefort-Buson et al. (1986) and Cowen and Frey (1987b) also reported low correlations among various distance estimates that varied in their own associations with trait heterosis.

In these crosses, there was a general, positive, heterotic effect on kernel weight, unrelated to G, and a more specific, usually negative effect on kernel number. In more diverse crosses (based on G), the kernel-number effect was less negative or even positive and did not nullify the kernel-weight effect through compensation.

The combination of midparent yield and G had predictive value in estimating F_2 mean yields only when parental germ plasm was sufficiently diverse, as indicated by $(1-r)$ (Table 5). We emphasize that $(1-r)$ was not used in this case simply to increase the frequency of low-yielding, unadapted crosses. As evidence, note that the mean yield of 43 crosses with high $(1-r)$ was actually slightly higher than the mean of 42 crosses with low $(1-r)$, with equal heterosis. However, INT parents were involved in 67% and 7% of the crosses with high and low $(1-r)$, respectively.

It is noteworthy that in predicting F_2 bulk grain yield from *previous years*' evaluations of the parental lines, G accounted for a larger proportion of the variation than did midparent values. Therefore, combinations of G and midparent together accounted for 50% of the variation among F_2 yields in crosses with high $(1-r)$. Conversely, they accounted for a nonsignificant 9% of the variation among F_2 yields in crosses with low $(1-r)$.

Likewise, $(1-r)$ predicted grain yield in combination with midparent only for crosses with greater G-distances (Table 6). Therefore, classification by either type of distance enhanced the predictive value of the other type. Whereas classification by $(1-r)$ caused considerable separation of crosses based on INT parentage, classification by G did not. Crosses with a least one INT parent accounted for 29% and 49% of all crosses with low and high G, respectively. Souza (1988) found that a method of "double classification" similar to ours aided in predicting specific combining ability in oats from pedigree and morphological data.

Assuming that F_2 yield is associated either with F_1 yield or with the frequency of superior inbred progeny, it could be concluded on the basis of our results that selection of parental combinations could be done effectively in three stages. First, superior parents would be chosen based on their own performance. Next, more detailed data would be collected from these parents for computation of distances similar to G or its components. Finally, crosses would be made between pairs of parents with both large phenotypic divergence and large $(1-r)$ value.

Based on this study, a broad germ plasm base is necessary to obtain a high frequency of positive F_2 heterosis, as well as to predict cross mean yields from distance estimates. The problem facing breeders, of course, is balancing the need for locally desirable agronomic and quality traits with the need for diversity.

Acknowledgements. We gratefully acknowledge the assistance of L. Harrell and R. Navarro in collecting field data for genetic divergence computations. The late Dr. D. M. Rodgers wrote the program for computing $(1-r)$ values. We acknowledge Dr. D. Wetzel, Grain Science Department, Kansas State University, for the protein analyses and Dr. B. Heer, agronomist-in-charge, South Central Experimental Field, Hutchinson/KS for his help.

References

Arunachalam V, Bandyopadhyay A, Nigam SN, Gibbons RW (1984) Heterosis in relation to genetic divergence and specific combining ability in groundnut (*Arachis hypogaea* L.). Euphytica 3:33–39

Busch RH, Janke JC, Frohberg RC (1974) Evaluation of crosses among high and low yielding parents of spring wheat (*Triticum aestivum* L.) and bulk prediction of line performance. Crop Sci 14:47–50

Cervantes T, Goodman MM, Casas E, Rawlings JO (1978) Use of genetic effects and genotype by environmental interactions for the classification of Mexican races of maize. Genetics 90:339–384

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

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

Cox TS, Murphy JP, Rodgers DM (1986) Changes in genetic diversity in the red winter wheat regions of the United States. Proc Natl Acad Sci USA 83:5583–5586

Cregan PB, Busch RH (1977) Early generation bulk hybrid yield testing of adapted hard red spring wheat crosses. Crop Sci 17:887–891

Cregan PB, Bush RH (1978) Heterosis, inbreeding, and line performance in crosses of adapted spring wheats. Crop Sci 18:247–251

Cress CE (1966) Heterosis of the hybrid related to gene frequency differences between two populations. Genetics 53:269–274

Ghaderi A, Adams MW, Nassib AM (1984) Relationship between genetic distance and heterosis for yield and morphological traits in dry edible bean and faba bean. Crop Sci 24:37–42

Goodman MM (1972) Distance analysis in biology. Syst Zool 21:174–186

Hanson WD, Casas E (1968) Spatial relationship among eight populations *Zea mays* L. utilizing information from a diallel mating design. Biometrics 24:867–880

Isleib TG, Wynne JC (1983) Heterosis in testcrosses of 27 exotic peanut cultivars. Crop Sci 23:832–841

Jatasra DS, Paroda RS (1983) Genetic divergence in wheat. Indian J Genet 43:63–67

Jinks JL (1983) Biometrical genetics of heterosis. In: Frankel R (ed) Heterosis. Reappraisal of theory and practice. Springer, Heidelberg Berlin New York pp 1–46 (Monographs on theoretical and applied genetics 6)

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

Lefort-Buson M, Dattee Y, Guillot-Lemoine B (1987) Heterosis and genetic distance in rapeseed (*Brassica napus* L.): Use of kinship coefficient. Genome 29:11–18

Mahalanobis PC (1936) On the generalized distance in statistics. Proc Natl Inst Sci India 2:49–55

Malecot G (1948) Les mathématiques de l'hérédité. Masson, Paris

Maluf WR, Ferreira PE, Miranda JEC (1983) Genetic divergence in tomatoes and its relationship with heterosis for yield in F_1 hybrids. Rev Brasil Genet 6:453–460

McCammon RB, Wenninger G (1970) The dendograph. State Geological Survey, University of Kansas Computer Contr. No. 38, Lawrence

Murphy JP, Cox TS, Rodgers DM (1986) Cluster analysis of red winter wheat cultivars based upon coefficents of parentage. Crop Sci 26:672–676

Ramanujam S, Tiwari AS, Mehra RB (1974) Genetic divergence and hybrid performance in mung bean. Theor Appl Genet 45:211–214

Shamsuddin AKM (1985) Genetic diversity in relation to heterosis and combining ability in spring wheat. Theor Appl Genet 70:306–308

Snedecor GW, Cochran WG (1980) Statistical methods. The Iowa State Univ Press, Ames

Souza EJ (1988) Measures and applications of genetic relationships in oats. PhD dissertation, Cornell University, Ithaca/NY

Zadoks JC, Chang TT, Konzak CF (1974) A decimal code for the growth stages of cereals. Weed Res 14:415–421