Theor Appl Genet (1988) 75:617–620

TAG THEORETICAL AND APPLIED GENETICS

Adjustment of parental frequencies to maximize the diversity of a synthetic population*

T. S. Cox

U.S. Department of Agriculture, Agricultural Research Service, Department of Agronomy, Kansas State University,
Manhattan, KS 66506, USA

Received June 15, 1987; Accepted September 7, 1987
Communicated by A. R. Hallauer

Summary. A method is outlined for maximizing the mean genetic distance among plants in a synthetic population by adjusting the relative contributions of the population's parents. The largest latent vector of the genetic distance matrix is used to find relative parental frequencies. The largest increases in the diversity of a synthetic will be achieved when there are different-sized clusters of parents, with considerably larger distances between than within clusters. The method may have application in maximizing the yields of synthetic cultivars or the resistance of multi-line cultivars.

Key words: Genetic distance – Random-mating – Heterosis – Multiline

Introduction

In crop breeding, a synthetic cultivar is produced by intermating a set of selected parental genotypes (Allard 1960). Similarly, base populations for recurrent selection programs are often synthetics formed from known parents (Hallauer and Miranda 1981). The productivity of a synthetic cultivar depends on both the productivity of the parents and the mean heterosis expressed by their progeny. For the purposes of recurrent selection, the expected genetic gain in a synthetic population depends, in part, on the amount of additive genetic variance. The heterosis or genetic variance in a population, in turn, depends in part on the number of segregating loci and the frequencies and effects of alleles at those loci.

Given a set of plant genotypes chosen on the basis of their own performance (per se or in hybrid combination), the objective was to determine how their contributions as parents to a synthetic population could be adjusted to maximize the mean genetic distance (univariate or multivariate) among the plants in a synthetic, thereby increasing its genetic variability or mean heterosis.

For a set of n selected parents, one may compute a $n \ x \ n$ genetic distance or similarity matrix, the elements of which could be statistics computed from any of several sources of data: metric traits (Martinez et al. 1983; Lee and Kaltsikes 1973), genetic markers (Cox et al. 1985; Smith et al. 1985), pedigrees (Cox et al. 1985; Murphy et al. 1986; St. Martin 1982), nonadditive genetic effects (Hanson and Johnson 1981), or some combination of data sources. Given a distance matrix $\boldsymbol{D}_{n \times n}$ (computed either directly or by converting a similarity matrix) for a set of n parents and a vector $\boldsymbol{p}_{n \times 1}$ of expected frequencies of parental genes in a synthetic population, the expected mean genetic distance among individuals in the population after one or more cycles of intermating is $\boldsymbol{p}'\boldsymbol{D}\,\boldsymbol{p}$.

The usual practice in initiating a synthetic is to allow equal contribution of all parents, although wind, insect, or even hand pollination will often favor certain parents. Equal parental contributions maximize $\boldsymbol{p}'\boldsymbol{D}\boldsymbol{p}$ if the parents are unrelated or are all related to the same extent. But degrees of relationship are rarely equal among all parents, and extreme situations may occur. For example, parents that are very closely related but differ for one or more desirable traits may be intercrossed with several other less closely related parents.

* Joint contribution of the USDA-ARS and Department of Agronomy, Kansas Agricultural Experiment Station. Contribution No. 87-435-J

In situations such as this, equal parental frequencies will not result in maximum mean genetic distance.

A vector of parental frequencies that maximizes $\boldsymbol{p}'\boldsymbol{D}\,\boldsymbol{p}$ can be found through calculations similar to those used in principal component analysis (Karson 1982), though here the operations are done on a matrix of distances between individuals (a Q-mode matrix). In principal component analysis, a scalar $\boldsymbol{a}'\boldsymbol{S}\,\boldsymbol{a}$, is maximized where $\boldsymbol{S}$ is a variance-covariance or correlation (R-mode) matrix defining relationships among variables. This is accomplished by setting $\boldsymbol{a}$ equal to the latent vector (or eigenvector) associated with the largest latent root (or eigen-value, λ_1) of $\boldsymbol{S}$, normalized so that $\boldsymbol{a}'\boldsymbol{a} = 1$. In fact, $\boldsymbol{a}'\,\boldsymbol{S}\boldsymbol{a} = \lambda_1$.

For a Q-mode distance matrix $\boldsymbol{D}$ one may also find λ_1 and $\boldsymbol{a}$ that would maximize $\boldsymbol{a}'\boldsymbol{D}\,\boldsymbol{a}$, but the elements a_i of $\boldsymbol{a}$ would not sum to 1; that is, they would not be frequencies. They do, however, give the optimum relative contributions of the parents, so $\boldsymbol{p} = \left(\sum_{i=1}^{n} a_i\right)^{-1} \boldsymbol{a}$ would maximize $\boldsymbol{p}'\boldsymbol{D}\boldsymbol{p}$. Therefore, the mean genetic distance between progeny produced by intermating n parents will be a maximum when parent i is represented with a frequency p_i that is proportional to the corresponding element a_i of the first latent vector of D. This method is a valid use of a Q-mode matrix in multivariate analysis (Gower 1966), because the objective is simply to find a maximum for $\boldsymbol{p}'\boldsymbol{D}\,\boldsymbol{p}$.

Examples

Suppose a set of 6 parents (A-F) for Population 1 and another set of 10 parents (G-P) for Population 2 are selected for intermating. Suppose further that parental distance matrices (Tables 1 and 2), which could be based on any appropriate data, have been computed. These distances happen to vary between zero and one, with the distance between each parent and itself equalling zero. This might not be the case in every parental set and is not essential to the analysis. The parental-frequency vector that results in a maximum mean genetic distance of 0.57 for Population 2 has elements ranging from 0.13 to 0.23 (Table 3); these frequencies result in a 14% larger mean genetic distance than when all frequencies are equal to 0.17 (rounded off). The optimum frequency vector for Population 2 has elements ranging from 0.09 to 0.12, with an increase of only 3% in mean genetic distance over equal parental frequencies.

As expected, closely related parents A and B in Population 1 and G and H in Population 2 are assigned the lowest frequencies in both populations, but their combined contribution is still significantly larger than that of any individual parent. Conversely, the parents in Population 1 not closely related to the others (E and F) are assigned much higher frequencies, but parents O and P in Population 2 are only slightly increased by adjustment.

In general, there must be some clustering of parents, and the clusters themselves must vary in diversity and/or size if adjustment is to have a significant effect

Table 1. Genetic distances between parents in example Population 1, with 6 parents

Parent						
A	B	C	D	E	F	Parent
0.00	0.02	0.25	0.20	0.75	0.95	A
	0.00	0.26	0.21	0.76	0.95	B
		0.00	0.20	0.90	0.96	C
			0.00	0.85	0.94	D
				0.00	0.02	E
					0.00	F

Table 2. Genetic distances between parents in example Population 2, with 10 parents

Parent										
G	H	I	J	K	L	M	N	O	P	Parent
0.00	0.02	0.25	0.87	0.49	0.50	0.46	0.78	0.97	0.96	G
	0.00	0.73	0.82	0.50	0.51	0.47	0.79	0.95	0.94	H
		0.00	0.90	0.62	0.62	0.59	0.84	0.97	0.96	I
			0.00	0.74	0.92	0.85	0.91	0.96	0.95	J
				0.00	0.74	0.72	0.87	0.96	0.95	K
					0.00	0.72	0.87	0.97	0.96	L
						0.00	0.86	0.96	0.95	M
							0.00	0.95	0.94	N
								0.00	0.98	O
									0.00	P

Table 3. Unadjusted and adjusted parental frequencies and resulting mean genetic distances for example Populations 1 and 2

Pop. 1 Parent	Frequency		Pop. 2 Parent	Frequency	
	Unadjusted	Adjusted		Unadjusted	Adjusted
A	0.17	0.13	G	0.10	0.08
B	0.17	0.13	H	0.10	0.08
C	0.17	0.15	I	0.10	0.09
D	0.17	0.15	J	0.10	0.11
E	0.17	0.21	K	0.10	0.10
F	0.17	0.23	L	0.10	0.10
			M	0.10	0.09
			N	0.10	0.11
			O	0.10	0.12
			P	0.10	0.12
	Mean genetic distance				
	0.51	0.57		0.70	0.72

on the diversity of a population. This can most easily be illustrated by looking at adjustment in only one dimension. Suppose that, instead of simultaneously adjusting all frequencies, we are able to divide a set of parents into two groups that have small mean within-group distances and a large mean distance between them. For simplicity, suppose that these are "true" distances so that the distance between a parent and itself is zero.

The mean genetic distance in a population developed from randomly intermating parents from two groups is

$$\left(1-\frac{1}{k_1}\right) P_1^2 \bar{d}_{11} + 2P_1(1-P_1)\bar{d}_{12} + \left(1-\frac{1}{k_2}\right)(1-P_1)^2 \bar{d}_{22}$$

where P_1 is the total of frequencies of parents in group 1; k_1 and k_2 are the numbers of parents in groups 1 and 2 respectively; $\bar{d}_{11}$ and $\bar{d}_{22}$ are the mean distances between different parents within groups 1 and 2, respectively; and $\bar{d}_{12}$ is the mean distance between parents in group 1 and parents in group 2. Setting the first derivative of this expression equal to zero, we find the value of P_1 that maximizes the mean distance in the population (given these particular groups of parents):

$$P_1' = \frac{\bar{d}_{12} - \left(1-\frac{1}{k_2}\right)\bar{d}_{22}}{2\bar{d}_{12} - \left(1-\frac{1}{k_2}\right)\bar{d}_{22} - \left(1-\frac{1}{k_1}\right)\bar{d}_{11}}.$$

This illustrates three points. (1) If between-group ($\bar{d}_{12}$) and within-group ($\bar{d}_{11}$ and $\bar{d}_{22}$) mean distances are similar, P_1' approaches $k_1/(k_1+k_2)$, which is its value when there is no adjustment. (2) If group sizes (k_1 and k_2) are similar, and within-group mean distances are similar, then P_1' again approaches $k_1/(k_1+k_2)$. (3) As within-group mean distances approach zero, P_1' approaches 1/2.

The effects of one-dimensional adjustment on the diversity of the population, given a range of within- to between-group mean distances and group sizes, are shown in Table 4. Adjustment has its greatest effect when there is one group with one or few members and another group with many members and low diversity, relative to its distance from the first group (small $\bar{d}_{22}/\bar{d}_{12}$). When the goups are of equal size, adjustment has little effect, no matter how great the distance between them or how low the within-group diversity.

The effect of simultaneous adjustment of all parental frequencies, as outlined in the previous section, would depend in the same way on the degree of clustering in the set of parents and the variation in the size of clusters. Populations 1 and 2 (Tables 1–3) do not have an unusual degree or pattern of clustering; the magnitude of increase in diversity resulting from adjustment in those populations is probably typical.

Table 4. Increase in mean genetic distance in a population (expressed as percentage of the mean genetic distance produced by equal parental frequencies), when relative frequencies of two parental groups are optimized. ($\bar{d}_{12}$ is the mean distance between parents in different groups; $\bar{d}_{11}$ and $\bar{d}_{22}$ are mean distances within groups 1 and 2 respectively, not considering distances between parents and themselves, and k_1 and k_2 are numbers of parents in groups 1 and 2, respectively. The ratio $\bar{d}_{12}/\bar{d}_{11}$ is not involved in the computation when $k_1 = 1$)

Ratios of within- to between-group distance		No. of parents in each group				
$\bar{d}_{22}/\bar{d}_{12}$	$\bar{d}_{11}/\bar{d}_{12}$	$k_1=1$, $k_2=4$	$k_1=1$, $k_2=19$	$k_1=4$, $k_2=16$	$k_1=10$ $k_2=10$	$k_1=16$ $k_2=4$
0.80	0.80	2	3	3	0	3
0.67	0.80	4	9	8	2	1
0.67	0.67	4	9	6	0	6
0.50	0.80	9	26	20	3	0
0.50	0.67	9	26	16	1	4
0.50	0.50	9	26	12	0	12
0.25	0.50	25	83	37	1	9
0.25	0.33	25	83	29	0	19
0.25	0.25	25	83	26	0	24

Practical considerations

Relatively small adjustments in parental contributions are difficult or impossible to make when intermating is being accomplished by hand-pollination. To do so would require using each parent in many more crosses than are normally made. For wind or insect pollination in the field, the proportions of parental seed mixed into a large bulk could easily be adjusted. Whereas deviations in parental contributions could still be caused by differences in emergence, vigor, pollination ability, or crossed seed production (Carlson 1971; Knowles 1969), an adjusted population would nevertheless have a greater likelihood to achieve maximum diversity than would an unadjusted population.

In a synthetic destined for use as a base population in a recurrent selection program, it is a simple matter to estimate the effect that adjusting parental frequencies to maximize genetic diversity will have on the population mean for a given trait (disregarding heterotic effects) by computing a weighted mean of parent values. One might find that the cost in mean performance would outweigh the enhanced diversity, especially considering that: (1) means are subject to much less error than genetic distance estimates, and (2) distance estimates are only as good as the quantity and quality of data used to compute them. Furthermore, any beneficial effect would be seen only in the first few selection cycles, because selection and perhaps genetic drift would change allelic frequencies in their own ways. Adjustment of frequencies would have the greatest potential when a set of highly productive parents with a distance matrix meeting the conditions described herein is chosen to produce a synthetic.

If the synthetic is to be used as a cultivar, and yield data for all hybrid combinations of parents are available, parental frequencies could be computed using the matrix of hybrid yields (with parental yields on the diagonal) as $\boldsymbol{D}$ in the foregoing analysis. Although $\boldsymbol{D}$ in this instance would not be a distance matrix, adjustment of parental frequencies would directly maximize the expected yield of the synthetic cultivar. Each hybrid combination, of the cross of parent i and parent j, would occur with a frequency of $2p_ip_j$, the same frequency as the union of the ith and jth gametes in a randomly-mated synthetic. Such adjustment would, of course, only be useful if the open-pollinated synthetic itself is to be used as a cultivar. The least-squares method of Pederson (1981) gives the best parental proportions for a population that is to be selfed to near-homozygosity, or which displays no dominance for the traits of interest. Under those conditions, the least-squares method, which involves only parental measurements, would be the best choice for simultaneously optimizing the population mean for several traits.

Though the method outlined herein may not provide large increases in diversity or productivity in a majority of situations, it could have applications in a number of specific cases, some of which have been mentioned. In a multi-line cultivar (Browning and Frey 1969), for example, diversity is an end in itself. A similarity index could be computed for all components of a multi-line, based on their reactions to pathogen genotypes occurring in the field and the frequency of those genotypes. Then, component frequencies could be adjusted to maximize the probability that two random plants would react differently to a given pathogen genotype.

References

Allard RW (1960) Principles of plant breeding. Wiley and Sons, New York, 483 pp

Browning JA, Frey KJ (1969) Multiline cultivars as a means of disease control. Annu Rev Phytopathol 7:355–382

Carlson IT (1971) Randomness of mating in a polycross of orchardgrass *Dactylis glomerata* L. Crop Sci 11:499–502

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

Gower JC (1966) Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53:325–338

Hallauer AR, Miranda FO (1981) Quantitative genetics in maize breeding. Iowa State University Press, Ames, Iowa, 467 pp

Hanson WD, Johnson EC (1981) Evaluation of an exotic maize population adapted to a locality. Theor Appl Genet 60:55–63

Karson MJ (1982) Multivariate statistical methods. Iowa State University Press, Ames, Iowa, 307 pp

Knowles RP (1969) Nonrandom pollination in polycrosses of smooth bromegrass *Bromus inermis* Leyss. Crop Sci 9: 58–61

Lee J, Kaltsikes PJ (1973) The application of Mahalanobis's generalized distance to measure genetic divergence in durum wheat. Euphytica 22:124–131

Martinez OJ, Goodman MM, Timothy DH (1983) Measuring racial differentiation in maize using multivariate distance measures standardized by variation in F_2 populations. Crop Sci 23:775–781

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

Pederson DG (1981) A least-squares method for choosing the best relative proportions when intercrossing cultivars. Euphytica 30:153–160

Smith JSC, Goodman MM, Stuber CW (1985) Genetic variability within U.S. maize germplasm. II. Widely-used inbred lines 1970 to 1979. Crop Sci 25:681–685

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