

Parental proportions maximizing the mean value of a parameter in a panmictic population can be useful in plant breeding

L Marchais *

Centre Orstom, 911, avenue Agropolis, BP 5045, F-34032 Montpellier cedex 1, France

(Received 23 January 1996; accepted 5 April 1996)

Summary — Many breeding experiments deal with the mean value of a parameter in a set of genotypes (or accessions). The parameter may relate either to the set itself (genetic distances between the genotypes, genetic diversity, multiline resistance) or to its offspring after intermating (yield of a synthetic population, heterosis). Most often, the mean value is computed with equal frequencies of the genotypes. This paper aims to determine the proportions of each genotype that maximize the mean value in the set, in order to increase the efficiency of breeding procedures. The interest of this procedure of maximized mean values is illustrated by an application in the management of genetic resources and by simulations in marker-assisted recurrent selection. It has drastically reduced redundancies in the core collection example while increasing the diversity by 81%, and it has reduced by 38% the time and money needed by the classical marker-assisted selection method in the simulated selection example. It points the way towards formulating many breeding problems in terms of maximizing quadratic forms of frequencies.

core collection / genetic resource / recurrent selection / marker assisted selection / QTL

Résumé — La détermination des proportions parentales qui maximisent la valeur moyenne d'un paramètre dans une population panmictique peut être utile en amélioration des plantes. De nombreux travaux de sélection portent sur la valeur moyenne d'un paramètre dans un ensemble de génotypes (ou d'échantillons). Ce paramètre peut caractériser les génotypes eux-mêmes (distances entre génotypes, diversité génétique de l'ensemble, résistance d'un mélange de lignées à un ensemble de pathogènes) ou la descendance de ces génotypes par panmixie (rendement d'une variété synthétique, hétérosis moyen). La plupart du temps, la valeur moyenne est calculée avec des fréquences égales des génotypes. Le présent article propose de déterminer les proportions des génotypes qui maximisent la valeur moyenne dans le mélange panmictique de façon à augmenter l'efficacité des méthodes de sélection. L'intérêt de cette procédure est illustré par une application en gestion des ressources génétiques et par des simulations en sélection récurrente assistée par marqueurs. La procédure a permis de supprimer toute redondance dans une collection restreinte (core collection) en augmentant la diversité moyenne de 81 % et de réduire de 38 % le temps et l'argent requis dans une sélection récurrente conventionnelle assistée par marqueurs. Elle invite à formuler de nombreux problèmes de sélection en termes de maximisation de formes quadratiques.

collection restreinte / ressources génétiques / sélection récurrente / sélection assistée par marqueurs / QTL

* Present address: 175, chemin des Écureuils, 34170 Castelnaud-le-Lez, France.

INTRODUCTION

In crop breeding, synthetic cultivars are usually produced by intermating a set of selected parental genotypes in equal proportions (Allard, 1960). It could be rewarding to initiate a synthetic by determining the relative contributions of each genotype that maximize the mean value of the synthetic for a given parameter. The parameter in this case could be, for instance, grain yield of the parents and of their hybrids.

Multiline varieties are traditionally developed as a mixture in equal proportions in order to stabilize their performances over a range of environments or against a range of pathogen races (Marshall and Brown, 1973). The variance in yield (V_m) of the mixture over the environments is taken as an inverse measure of its stability: $V_m = \sum_i \sum_j p_i p_j C_{ij}$, where p_i is the proportion of line i in the mixture; C_{ij} is the covariance in yield of the i th and j th lines over the environments; C_{ii} is the variance of the i th line. The best mixture would be obtained with a vector \mathbf{p} that minimizes V_m or maximizes $\sum_i \sum_j p_i p_j (V_{\max} - C_{ij})$, where V_{\max} is the highest variance. It is hoped that V_m is less than the variance of the line with the smallest variance.

These two examples show that important breeding problems can be defined in terms of maximizing mean values which are quadratic forms of frequencies ($Q(\mathbf{p}) = \sum_i \sum_j p_i p_j d_{ij} = \mathbf{p}' \cdot \mathbf{D} \cdot \mathbf{p}$, where \mathbf{p} is a vector of frequencies summing up to 1. The vector \mathbf{p} refers to a set of genotypes and \mathbf{D} is a symmetric matrix of which the elements refer to pairs of genotypes in the set.

This paper will illustrate the interest of determining the relative contributions of parents that maximize some mean value in a panmictic population by means of two applications: (1) the management of genetic resources and creation of a core collection (Frankel and Brown, 1984); and (2) plant breeding, at a time when it is becoming possible to predict the value of a plant from the knowledge of its genotype at a marker locus tightly linked to a quantitative trait locus (QTL) (Edwards et al, 1992), recurrent selection for the fixation of a set of particular genes.

METHOD

The problem is to find the frequencies p_i maximizing the quadratic form:

$$Q(\mathbf{p}) = \sum_i \sum_j p_i p_j d_{ij} = \mathbf{p}' \cdot \mathbf{D} \cdot \mathbf{p}$$

where in matrix notation $\mathbf{p}' = (p_1 \dots p_n)$ is the vector of parental contributions to the bulk or to the synthetic population. Each p_i element is subject to the restrictions:

$$0 \leq p_i \leq 1 \text{ and } \sum_i p_i = 1$$

The matrix $\mathbf{D} = (d_{ij})$ is real symmetric. Its elements are the values attached to the couples (i, j) of genotypes.

Many consistent iterative algorithms are available in statistical packages for obtaining the maximum of any function. They have not been tried and compared in the present work. Instead an algorithm found in population genetics has been used that has given good results, although it has not been proven so far that this is the best one for maximizing quadratic forms of frequencies.

With genotypes considered as multiple alleles of a locus, the method employs the reasoning elaborated in the theory of natural selection acting on a panmictic population with distinct generations and constant selective values. Population genetics manuals (eg, Crow and Kimura, 1970) establish the following relations between the frequencies $p_{i,n+1}$ of A_i alleles in generation $(n+1)$ and frequencies $p_{i,n}$ in generation (n) :

$$p_{i,n+1} = p_{i,n} \sum_j p_{j,n} d_{ij} / (\sum_i \sum_j p_{i,n} p_{j,n} d_{ij}) \quad (1)$$

d_{ij} (often noted w_{ij} in the manuals) is the selective value of the $A_i A_j$ genotype. In this context, all the d_{ij} are positive or null.

It has been shown that, with these frequency changes, the evolution of $Q(\mathbf{p})$ is such that $\text{Maximum}(d_{ij}) \geq Q(\mathbf{p}_{n+1}) \geq Q(\mathbf{p}_n)$ and that $Q(\mathbf{p}_n)$ tends towards a limit which is a local maximum. The exact demonstration is probably due to Mulholland and Smith (1959) and to Kingman (1961). An approximate demonstration is given in Crow and Kimura (1970, 274).

The method consists simply of repeated application of the algorithm (1), starting with the initial vector $p_i = 1/n$ (n is the number of genotypes in the set), until an equilibrium is reached which gives the maximum of $Q(\mathbf{p})$. In theory, this maximum may depend on the initial vector \mathbf{p} , but the vector $(p_i = 1/n)$ is of the most practical concern. Anyway, with high numbers of genotypes it would be completely unfeasible to iterate the maximization process with several sets of initial \mathbf{p} values.

Two other methods may be mentioned. Cox (1988) proposed maximizing $Q(\mathbf{p})$ by means of a method based on the first eigenvector of the \mathbf{D} matrix: this is not mathematically justified and is inefficient. A mathematical solution is proposed in population genetics manuals (eg, Elandt-Johnson, 1971) consisting in the search for the extremum of $Q(\mathbf{p})$: in practice, this does not work, because it gives solutions outside the $(0,1)$ interval that are not frequencies. An example of these two methods will be given in the application on core collections.

APPLICATION 1. DEVELOPMENT OF A CORE COLLECTION

Managers of genetic resources are interested in the creating working collections comprising maximum diversity in a minimum size: the so-called core collections. The goal of these core collections is to facilitate access to and use of existing diversity in integral collections to scientists and breeders (Frankel and Brown, 1984).

Brown (1989) has proposed a two-step procedure for developing a core collection:

- determining groups of accessions relatively homogeneous for available taxonomic, geographic, enzymic, morphological, and agronomical information;
- sampling of about 10% of the integral collection by random selection in each group of a number of accessions proportional to the logarithm of the group size.

When intra-group information is available, instead of a random selection of accessions assumed to comprise overall diversity, the method proposed here is well adapted to the exact identification of the accessions, and their corresponding proportions, to be included in the core collection.

An example is given here of 95 wild millet samples (*Pennisetum glaucum* subsp *monodii*) collected in a vast area of the West African Sahel extending between Adrar des Iforas, the Aïr mountains and the agricultural part of République du Niger, and described for eight enzymic systems with 12 loci and 46 alleles (Tostain, 1992). These millets comprise a relatively homogeneous

group in terms of their enzymic diversity, in comparison with other geographical groups. Let us determine the selection of accessions using maximum Nei's diversity (Nei, 1973).

Let x_{ial} be the frequency of allele a of locus l in accession i . The whole Nei's diversity H_t of the weighted bulk is:

$$H_t = \sum_i \sum_j p_i p_j \sum_l (1 - \sum_a x_{ial} x_{jal})$$

The maximization of H_t is obtained by applying the algorithm (1) to the matrix with general term

$$d_{ij} = \sum_l (1 - \sum_a x_{ial} x_{jal})$$

A state of equilibrium is practically reached after 700 iterations: the five most significant figures of $Q(\mathbf{p})$ and the three most significant figures of p_i do not change in the following 200 iterations. Table I shows that the initial diversity of the complete collection before weighting ($p_i = 1/95$) is 3.58. After maximization, the diversity is 6.48 and is obtained with only eight accessions. The most important accession represents 27% of the genetic material in the core collection of Niger wild millets. The current procedure is very efficient at increasing the diversity and radically eliminating redundancy in this group. The same procedure could be applied again to the mean allelic frequencies of each group included in the core collection to determine the proportions of each group that maximize the whole core collection diversity.

The diagonalization method (Cox, 1988) applied to the same matrix gave a mean diversity of 4.31 and selected all the accessions with

Table I. Composition of the core collection extracted from the 95 accessions of wild pearl millets collected in Niger by maximization of mean Nei's diversity.

Label of selected accession	Origin	Nei's diversity	Proportion
B5	Gourma	2.476	0.27
H11	Niger agricole	3.323	0.20
B1	Gourma	2.738	0.17
K9	Aïr	1.726	0.09
D16	Menaka	3.456	0.09
H20	Niger agricole	2.833	0.08
F1	Burkina Faso	2.449	0.06
E9	Adrar des Iforas	3.043	0.04
Total			1.000

Diversity of the integral equal frequencies collection = 3.58; diversity of the weighted core collection = 6.48.

weights between 0.009 and 0.023. This method did not find the maximum diversity and did not eliminate redundancies. The other method searching for an extremum also failed: the extremum obtained, 0.315, is a minimum and the weights are between -5.69 and $+6.17$. These are not frequencies. Clearly, both methods are inappropriate.

But the procedure proposed here is useful only if the core collection diversity is not altered too much by modifications of the **D** matrix due to errors in estimating **D** elements. Ten simulations were performed on **D** matrices randomly modified: with equal probabilities, each term was either increased or decreased by 10%. Table II shows that the core collection determined with the initial **D** matrix keeps a high diversity when the **D** matrix is modified (on average 92% of the maximum diversities computed on modified matrices). In practice, then, in the case of strong redundancies, the value of a core collection determined by the present procedure is not very sensitive to errors in estimating **D** elements.

The present procedure to create core collections could be applied simultaneously to qualitative and quantitative characters, once each quantitative character has been split into classes. But this raises the problem of determining the number of classes attributed to each character.

The main objection to the present procedure is that, in most cases, accessions are not analysed before creating a core collection. But this procedure could be considered, after analysis of a pre-

liminary selection of accessions made according to the principles of Brown (1989).

A treatment of the problem created by a lack of data on some accessions remains to be studied when using the present procedure, especially in terms of robustness.

APPLICATION 2. RECURRENT SELECTION FOR A PARTICULAR GENOTYPE

The identification of loci acting on quantitative traits (QTL) is now possible thanks to the establishment of plant genetic maps with molecular markers. It is possible to determine the genotype of a plant at a set of marker loci tightly linked to QTLs controlling the expression of different characters. The phenotypes of a set of plants and of their crosses can therefore be predicted (Edwards et al, 1992; Goldman et al, 1993).

Marker assisted selection raises a lot of hopes but has not yet produced real results (Dudley, 1993). Beforehand, a detailed and expensive analysis of the material to be improved is required: to identify the different QTLs of interest, especially when they are borne on the same segment of a chromosome, to find a marker tightly linked to each QTL, to evaluate the effects of epistasis and genotype x environment interactions (Gallais and Rives, 1993).

Table II. Comparison of the diversities of the core collection determined on the original **D** matrix with the maximum diversities computed with ten randomly-modified matrices.

<i>Matrix D</i>		<i>Mean diversity</i>			
		<i>Initial collection</i>	<i>Maximum</i>	<i>Core collection</i>	<i>Core/maximum</i>
Original		3.58	6.48	6.48	1.0
Modified	1	3.58	6.88	6.37	0.92
	2	3.58	6.92	6.14	0.88
	3	3.58	6.94	6.45	0.92
	4	3.57	6.80	6.30	0.92
	5	3.59	6.98	6.29	0.90
	6	3.58	6.96	6.62	0.95
	7	3.58	7.00	6.40	0.91
	8	3.59	6.96	6.47	0.92
	9	3.59	7.00	6.61	0.94
	10	3.58	6.91	6.44	0.93
Mean					0.919

We shall suppose here that all these difficulties have been resolved, and that the breeder works on an F2 generation from two parents each possessing some favourable Q_1 and some unfavourable Q_2 alleles at QTLs tightly linked to segregating molecular marker alleles, respectively M_1 and M_2 .

Recurrent selection has been simulated in order to bring together in a single plant all the favourable marker alleles (designated M_1) at 20 unlinked loci. Unfavourable alleles are designated (M_2). The initial population comprises 100 individuals each created by random sampling of 40 alleles from a pool of equiproportional alleles ($1/2M_1, 1/2M_2$).

Genotypes are described with the Θ variable of Hayman (1954):

$$\Theta = 1 \text{ for } M_1M_1 \quad \Theta = 0 \text{ for } M_1M_2 \quad \Theta = -1 \text{ for } M_2M_2$$

The variable Θ for the offspring of the cross between two individuals i and j is:

$$\Theta_{ij} = (\Theta_i + \Theta_j)/2$$

The allele M_1 frequency in this offspring is $1/2 + (\Theta_i + \Theta_j)/4$.

In the simulation, the selective value d_{ij} of the cross $i \times j$ is chosen such that:

$$d_{ij} = \left(\frac{1}{2} + \frac{\Theta_{i1} + \Theta_{j1}}{4}\right) * \left(\frac{1}{2} + \frac{\Theta_{i2} + \Theta_{j2}}{4}\right) * \dots * \left(\frac{1}{2} + \frac{\Theta_{i20} + \Theta_{j20}}{4}\right)$$

It is the product of favourable allelic frequencies at the 20 loci. Thus, in the case of two individuals having the genotype (M_2M_2) at the same locus, their selective values and that of their cross are null. Only one maximum is possible, ie, unity. All

the favourable alleles are constrained to reach fixation.

Each cycle of selection comprises three steps in one generation: (i) for the 100 individuals included in the generation, estimating the selective values of their offspring in crosses (d_{ij}) and selfings (d_{ii}); (ii) determining the proportions p_i maximizing the mean selective value of the panmictic offspring by iterative application of algorithm (1); and (iii) random sampling of 100 individuals generated by weighted panmixia: each individual is generated by random sampling of two parents with probabilities p_i . To select a parent, the interval (0, 1) is segmented in successive intervals of lengths $p_1, p_2 \dots p_{100}$. A pseudo-random number with a uniform distribution between 0 and 1 is drawn. An individual is chosen as parent if the random value falls in its segment p_i . In both parents, at each locus there is random sampling of an allele with probability $(1 + \Theta)/2$ for M_1 and $(1 - \Theta)/2$ for M_2 . Thus, the 20 loci are considered independent, but any linkage disequilibrium in the parents is taken into account.

In the 32 simulations performed, several genotypes homozygous for M_1M_1 at all 20 loci were obtained after two, three or four cycles of selection. The examples shown in table III prompt the following practical comments.

In each generation, the number of parents selected was low: from one to three, rarely four. Consequently, it would be easy to make all the required crosses manually and to compose the weighted bulk: the proportions p_i^2 for selfs and $2p_i p_j$ for the crosses, although this low number can cause some undesired effects of inbreeding in allogamous crops. In such a case, the procedure should be convenient for preparing inbred lines in view of breeding hybrids. Considering

Table III. Examples of simulations MASMAX showing the number and proportions of the parents selected in each cycle; the number of individuals fixed for the M1 alleles obtained at last cycle of selection is indicated.

Simulation MASMAX no	Parental proportions p_i				Number of fixed individuals
	Cycle 1	Cycle 2	Cycle 3	Cycle 4	
1	0.50 0.50	1.00			1
2	0.43 0.43 0.14	1.00	1.00		23
3	0.50 0.50	1.00	1.00		6
4	0.50 0.50	0.89 0.11	1.00		4
5	0.47 0.45 0.08	0.46 0.34 0.20	0.50 0.50	0.50 0.50	28
6	0.50 0.50	0.68 0.32	0.40 0.30 0.30	1.00	5
7	0.41 0.35 0.24	0.37 0.37 0.26	0.73 0.27	0.27 0.25 0.25 0.23	9

that at each generation 100 individuals are generated from a low number of different crosses, it can be presumed that the realized mean value in the offspring should not be too far from the expected one. Thus, selection should not be disturbed by random drift. However this point has not been checked. The choice of the progenitors and the determination of their relative proportions would not be easy to make by hand, especially in case of unequal proportions.

This method of selection, which will be labelled MASMAX, has been compared to the classical marker assisted recurrent selection method, labelled MAS, such as set out for instance by Edwards and Page (1994), as follows. In each generation, the 10 individuals with the highest score $d_i = \sum_l \Theta_{ij}$ (summing over the 20 loci) are chosen among 100 individuals as parents of the next generation and intercrossed with equal probabilities. With MAS method, the fixation of M_1 alleles is obtained after four or five cycles. As an average, MASMAX needs 3.41 cycles for fixing M_1 alleles instead of 4.72 cycles with MAS (table IV). That means that MASMAX is 38% faster than MAS and consequently 38% cheaper in expenses of field cultures and laboratory analyses. Another slight advantage of MASMAX over MAS is a higher number of fixed lines when both methods need four cycles for fixation, which can facilitate further breeding work.

The stringency of the selective formula chosen in MASMAX (defining the d_{ij}) was such that with a probability of $[1-(3/4)^{20}]^{100} = 0.728$ the 100 initial F2 genotypes had a null selective value ($d_{ij} = 0$), but this could not hamper the selection process because the expected number of non-null d_{ij} values is $100.99./2.(15/16)^{20} = 1\ 361$ (a count in a simulation gave 1 314 non-null d_{ij}) and the risk that all d_{ij} are null is $(1-(15/16)^{20})^{4950} = 0$. So, MASMAX exerts an intensity of selection adjusted

at each cycle for the maximum efficiency but always close to the maximum (from one to three parents) based mainly on the offspring values when MAS exerts a constant lower intensity of selection (ten parents) based only on parental values. If MAS exerted a more severe intensity of selection, the risk of losing some M_1 alleles would become high. Furthermore, MAS has no means to choose among parents having an equal score. Finally, the selective formula chosen for MASMAX could perhaps be improved: provided that M_2 has not reached fixation in the selected parents, it might be useful to keep as a selected parent an individual with M_1M_1 genotypes at every other loci, even if it is M_2M_2 at some locus.

The efficiencies of the MASMAX and MAS methods were compared when some degree of recombination occurs between QTLs and markers. The *Appendix* shows how to compute the probability of drawing a gamete M_1Q_1 at the end of the selection procedure for each pair of loci (M,Q), if parental genotypes at marker loci are known in each generation. The expected number of favourable alleles Q_1 , and its standard deviation, can then be estimated.

Calculations have shown that the expected probability, m , of a non recombinant gamete M_1Q_1 in an individual fixed for the 40 M_1 alleles is practically the same in both methods. For instance, with a 10% recombination, a MASMAX simulation achieved in three cycles gave $m = 0.833$ instead of 0.824 with a MAS simulation achieved in five cycles. As a matter of fact, in both methods the initial probability of a recombination is the same, ie, 10%, and cannot be reduced.

According to the simulations of Edwards and Page (1994) with additive characters, when recombination between markers and QTLs is above 10%, flanking markers become necessary to keep the efficiency of MAS (and MASMAX)

Table IV. Percentage of the simulations MAS and MASMAX fixing M1 alleles after two, three, four and five cycles of selection. The number of fixed genotypes is indicated in brackets.

Method	Total number of simulations	Number of cycles needed for fixation			
		2	3	4	5
MAS	32			28% (1.3)	72% (7.0)
MASMAX	32	6% (1)	47% (8.1)	47% (20.2)	

Mean time of fixation for MAS = 4.72 cycles, for MASMAX = 3.41 cycles.

superior to that of conventional phenotypic recurrent selection (PRS) at least for a rapid response early in the selection process. When all markers are fixed, PRS remains necessary to achieve further responses and reach the maximum potential.

But MASMAX and MAS seem preferable, even with some M-Q recombination, in two situations where PRS is not efficient: (1) when the tests of selective values are difficult to operate at each generation (eg, drought or disease resistance ratings) and (2) when creating genotypically contrasted inbred lines in view of breeding hybrid varieties, in the case of superdominant, epistatic characters.

CONCLUSION

The results obtained here are encouraging. The proposed breeding procedure should be usefully employed in the management of genetic resources to create core collections. The method seems to be adequate in recurrent selection in dealing rationally with the information provided by molecular markers on QTLs. It should increase the rapidity of recurrent selection, reduce its cost and advance the time of earnings.

The two applications shown in the present article have endeavoured to point the way to formulating many breeding problems in terms of maximizing quadratic forms of frequencies. Apart from the attempt of Cox (1988), breeders have so far curiously not considered that possibility.

ACKNOWLEDGMENTS

I thank S Tostain for the use of his enzymic data on wild pearl millets from Niger, and S Tostain, X Perrier and the reviewers of agronomie for their comments on an earlier version of the manuscript.

APPENDIX

Effect of recombinations between a QTL Q and its marker M on the number of favourable alleles Q_1 obtained in the recurrent selection procedure reported here.

Let us call:

r the recombination rate between a QTL and its marker locus

a_g, b_g, c_g, d_g the probabilities of gametes $M_1Q_1, M_1Q_2, M_2Q_1, M_2Q_2$ produced by parents selected

in generation $g-1$ to make by panmixia the generation g

$m_{1g} = a_g + b_g$ is the observed frequency of allele M_1 in parents selected in generation $g-1$ and also the probability of M_1 in generation g before selection

$m_{2g} = c_g + d_g$ has the same meaning as m_{1g} but for allele M_2

$(M_1M_1)_g, (M_1M_2)_g, (M_2M_2)_g$ the observed frequencies of the respective genotypes in the parents selected in generation g .

Let us note $D_g = a_gd_g - b_gc_g$

The following recursion formulas can be established:

$$a_{g+1} = \frac{(M_1M_1)_g a_g / m_{1g} + (M_1M_2)_g (a_g m_{2g} - r D_g)}{(2m_{1g} m_{2g})}$$

$$b_{g+1} = \frac{(M_1M_1)_g b_g / m_{1g} + (M_1M_2)_g (b_g m_{2g} + r D_g)}{(2m_{1g} m_{2g})}$$

$$c_{g+1} = \frac{(M_2M_2)_g c_g / m_{2g} + (M_1M_2)_g (c_g m_{1g} + r D_g)}{(2m_{1g} m_{2g})}$$

$$d_{g+1} = \frac{(M_2M_2)_g d_g / m_{2g} + (M_1M_2)_g (d_g m_{1g} - r D_g)}{(2m_{1g} m_{2g})}$$

These formulas are reduced to:

$$a_{g+1} = a_g m_{1,g+1} / m_{1g} - (M_1M_2)_g r D_g / (2m_{1g} m_{2g})$$

$$D_{g+1} = D_g (2m_{1,g+1} m_{2,g+1} - r (M_1M_2)_g) / (2m_{1g} m_{2g})$$

The final value of a_g is computed recursively, knowing the parental genotypes at each generation and the initial values $a_1 = (1 - r)/2$, $D_1 = (1 - 2r)/4$, $m_{11} = 0.5$. With selection, m_{1g} tends towards 1 and D_1 towards 0.

The 20 final a_g values computed in a particular simulation can be used to estimate the expected value, m , and the variance, v , of a_g at loci (M,Q) in all possible selection experiments. All the a_g values are thus considered as independent random variables with the same expectations and variances.

Consequently, the number n of favourable alleles Q_1 obtained at the end of a selection procedure has an expected value:

$$E(n) = 40m$$

and a variance:

$$V(n) = 40(v + m(1-m))$$

REFERENCES

- Allard RW (1960) *Principles of Plant Breeding*. Wiley and Sons, New York, 483 p
- Brown AHD (1989) Core collections: a practical approach to genetic resources management. *Genome* 31, 818-824
- Cox TS (1988) Adjustment of parental frequencies to maximize the diversity of a synthetic population. *Theor Appl Genet* 75, 617-620
- Crow JF, Kimura M (1970) *An Introduction to Population Genetics Theory*. Harper and Row, New York, 591 p
- Dudley JW (1993) Molecular markers in plant improvement: manipulation of genes affecting quantitative traits. *Crop Sci* 33, 660-668
- Edwards MD, Helentjaris T, Wright S, Stuber CW (1992) Molecular marker facilitated investigations of quantitative trait loci in maize. *Theor Appl Genet* 83, 765-774
- Edwards MD, Page NJ (1994) Evaluation of marker-assisted selection through computer simulation. *Theor Appl Genet* 88, 376-382
- Elandt-Johnson R (1971) *Probability models and statistical methods in genetics*. John Wiley, New York, 592 p
- Frankel OH, Brown AHD (1984) Current plant genetic resources – a critical appraisal. In: *Genetics: New Frontiers* 4, p 3-13. Proc XV Int Congress of Genetics, Oxford & IBH Publishing Co
- Gallais A, Rives M (1993) Detection, number and effects of QTLs for a complex character. *agronomie* 13, 723-738
- Goldman IL, Rocheford TR, Dudley JW (1993) Quantitative trait loci influencing protein and starch concentration in the Illinois long-term selection maize strains. *Theor Appl Genet* 87, 217-224
- Hayman BI (1954) The theory and analysis of diallel crosses. *Genetics* 39, 789-809
- Kingman JF (1961) On an inequality in partial averages. *Quarter J Math* 12, 78-80
- Marshall DR, Brown AHD (1973) Stability of performance of mixtures and multilines. *Euphytica* 22, 405-412
- Mulholland HP, Smith CA (1959) An inequality arising in genetical theory. *Amer Math Monthly* 66, 673-683
- Nei M (1973) Analysis of gene diversity in subdivided populations. *Proc Nat Acad Sci USA* 70, 3321-3323
- Tostain S (1992) Enzyme diversity in pearl millet (*Pennisetum glaucum* L) 3. Wild millet. *Theor Appl Genet* 83, 733-742