Theor Appl Genet (1989) 78: 233–242

Some aspects of autotetraploid population dynamics

D. R. Glendinning

45 Morningside Drive, Edinburgh, UK

Received October 10, 1988; Accepted February 23, 1989
Communicated by A. R. Hallauer

Summary. Self-compatible autotetraploids are likely to set much of their seed by selfing. Formulae are presented for the frequencies in any generation of states of loci, from homozygous to quadri-allelic, considering the frequencies of selfing and of double reduction but not allowing for the effects of inbreeding depression on population structure. The changing structure of populations over generations is also explored by computer simulation, incorporating selection against inbreds but ignoring double reduction. The findings are discussed in relation to mass-selection programmes.

Key words: Mass selection – Inbreeding depression – Computer simulation

Introduction

The author, while mass-selecting an autotetraploid potato population, found that about 80% of the naturally set seed was selfed (Glendinning 1976) and subsequently derived the algebraic formulae given below for the states of an individual locus, incorporating the frequency of double reduction. Various publications later came to attention but none covers all the aspects dealt with here. Demarly (1961) deals with inbreeding, not considering double reduction. Bennett (1968) deals with mixed selfing and crossing, incorporating double reduction, but he considers only two alleles and his findings do not extend to the frequencies of loci with more than two alleles present. (There are misprints in Bennett's formulae; see Appendix 1). Gallais (1968) deals with inbreeding and outbreeding separately, not incorporating double reduction.

In developing the formulae it was, of necessity, assumed that genotypes having the locus in all its possible states, from quadri-allelic to homozygous, contribute as parents to the next generation in proportion to their frequencies. But where both selfing and crossing are taking place there could, in theory, be a continuous range from fully inbred to fully outbred genotypes in the population, with the less heterozygous states of the locus occurring mainly in the more inbred genotypes. In practice, 'inbreeding depression' will reduce the contributions of inbred genotypes as parents of future generations, altering the structure of the population in favour of outbreeding. As regards any particular locus, this will reduce the frequencies of the less heterozygous states.

An algebraic approach combining consideration of the status of the whole genotype together with that of the individual locus, if in fact possible, is beyond the author's competence. However, a system of computer simulation has been devised. This is based on the frequencies of the various states of loci in each individual genotype in the population. The probabilities of a random locus in a random genotype being in each of the possible states (equivalent to the algebraic coefficients) are represented by population mean values. The inbred status of each individual genotype (considering all its loci) being known, selection against inbreds can be simulated, leading to perhaps more realistic estimates of coefficients than are obtained using the algebraic formulae.

But of necessity, the 'rules of inheritance' must be considered to be the same for all loci and it is not possible to take double reduction, varying between loci, into account. Use of a mean value has been rejected, as it would not give findings relevant to any particular locus; a mean value would in any case generally be a small one and would have little effect on the outcome. (For potatoes, small-chromosome species in which only about a third of the chromosomes are usually in quadrivalents and in which there may rarely be more than one chiasma in a chromosome arm, the mean may be 1/36 or less, though for individual loci at the ends of arms of chromosomes always involved in closed quadrivalents, the frequency could be as high as 1/6.)

In both studies, of necessity, the possibility of aneuploidy is ignored.

Nomenclatures for genotypes

The nomenclature for genotypes as regards individual alleles is well established: homozygous, triplex, duplex, simplex and nulliplex, their frequencies coded here A, B, C, D and E, respectively. However, the author knows of no established English nomenclature as regards states of loci and he finds problems with a nomenclature used by Demarly (1961), who terms a genotype triplex for one allele and simplex for another 'digéniques simplex'; 'digéniques triplex' would have seemed preferable, facilitating summation of states over alleles, as below. For present purposes loci are referred to as homozygous, triplex (actually triplex for one allele, simplex for another), double-duplex (duplex for each of two alleles), tri-allelic (duplex for one allele, simplex for each of two others) and quadri-allelic (simplex for each of four alleles), their frequencies indicated by A', B', CC', T' and Q', respectively. Summing the frequencies of homozygous states of individual alleles gives the total frequency of homozygous states of that locus in the population ($\Sigma A = A'$), similarly for triplex states ($\Sigma B = B'$), but the sum of the frequencies of duplex states of alleles is equal to twice the frequency of double-duplex plus the frequency of tri-allelic states of the locus ($\Sigma C = 2CC' + T'$), and the sum of the frequencies of simplex states of alleles is equal to the frequency of triplex plus twice the frequency of tri-allelic plus four times the frequency of quadri-allelic states of the locus ($\Sigma D = B' + 2T' + 4Q'$).

Algebraic formulae for the states of a locus

The following coefficients were derived assuming the population to be of infinite size and to have, at the locus, an infinite number of alleles, each of infinitesimal frequency. One effect of these assumptions is that the probability of any two genotypes involved in a cross having alleles in common is zero. The coefficients thus derived are prefixed F (or G) to indicate that, in common with the traditional inbreeding coefficient F in diploids, they are independent of allelic frequencies and are, in effect, inbreeding/outbreeding coefficients. For notes on their derivation, see Appendix 2.

The four coefficients which can be calculated directly are: Ga, the probability of a random pair from the four alleles at a locus in a random genotype not being identical, $= (3FB' + 4FCC' + 5FT' + 6FQ')/6$;
Gb, the frequency of duplex associations of alleles in the population, $= 2FCC' + FT'$;
Gc, the frequency with which alleles are simplex while not being in a genotype which is triplex for another allele, $= 2FT' + 4FQ'$; and FQ' itself.

Then:

$$FT' = Gc/2 - 2FQ'$$

$$FCC' = (Gb - FT')/2 = Gb/2 - Gc/4 + FQ'$$

$$\begin{aligned} FB' &= 2Ga - (4FCC' + 5FT' + 6FQ')/3 \\ &= 2Ga - 2Gb/3 - Gc/2 \end{aligned}$$

$$\begin{aligned} FA' &= 1 - FB' - FCC' - FT' - FQ' \\ &= 1 - 2Ga + Gb/6 + Gc/4 \end{aligned}$$

The inbreeding coefficient F as traditionally applied to autotetraploids, the probability of a random pair from the four alleles at a locus being identical by descent, is $1 - Ga$. It is referred to below as the generalised inbreeding coefficient, coded Fg.

Using 'S' to represent the frequency of selfing in the population and 'α' to represent the frequency of double reduction at the locus, the vectors which control the rate of approach of the coefficients to equilibrium are:

$$\beta = (2(1 - \alpha) + 3S)/6$$

$$\gamma = S(1 - \alpha)^2/6, \quad \text{and}$$

$$\delta = S(1 - \alpha)/2.$$

The equilibrium values of the directly calculable coefficients are:

$$Ga, \infty = 2(1 - S)/3(1 - \beta)$$

$$Gb, \infty = JGa, \infty/(1 - \gamma) \quad \text{where} \quad J = 3\alpha + 9\gamma$$

$$Gc, \infty = KGa, \infty/(1 - \delta) \quad \text{where} \quad K = 4(1 - S)(1 - \alpha)$$

$$FQ', \infty = LGa, \infty^2/(1 - \gamma) \quad \text{where} \quad L = (1 - S)(1 - \alpha)^2.$$

If the deviations from equilibrium values were initially: $Da, o = Ga, o - Ga, \infty$; $Db, o = Gb, o - Gb, \infty$ etc., in the 'nth' generation:

$$Da, n = \beta^n Da, o$$

$$Db, n = \gamma^n Db, o + (J - 3(1 - \beta))\, Da, o(\beta^n - \gamma^n)/(\beta - \gamma)$$

$$Dc, n = \delta^n Dc, o + KDa, o(\beta^n - \delta^n)/(\beta - \delta)$$

$$\begin{aligned} DFQ', n = {}& \gamma^n DFQ', o + 2LDa, o\, Ga, \infty(\beta^n - \gamma^n)/(\beta - \gamma) \\ & + LDa, o^2(\beta^{2n} - \gamma^n)/(\beta^2 - \gamma) \end{aligned}$$

then $Ga, n = Ga, \infty + Da, n$; $Gb, n = Gb, \infty + Db, n$, etc.

That concludes the presentation of the coefficients. Using them, one may obtain the probabilities of the locus being in the states from A (homozygous) to E (nulliplex) as regards allele 'i' with frequency 'p':

$$A(i) = pFA' + p^2[FB' + FCC'] + p^3FT' + p^4FQ'$$

$$B(i) = p(1 - p)FB' + 2p^2(1 - p)FT' + 4p^3(1 - p)FQ'$$

$$C(i) = 2p(1 - p)FCC' + p(1 - p)FT' + 6p^2(1 - p)^2FQ'$$

$$D(i) = p(1 - p)FB' + 2p(1 - p)^2FT' + 4p(1 - p)^3FQ'$$

$$\begin{aligned} E(i) = {}& (1 - p)FA' + (1 - p)^2[FB' + FCC'] + (1 - p)^3FT' \\ & + (1 - p)^4FQ' \end{aligned}$$

Considering all alleles at the locus, allowing for their having different individual frequencies (values of 'p'), the actual probabilities of loci being in the various states from homozygous to quadri-allelic are:

$$A' = FA' + \Sigma p^2(FB' + FCC') + \Sigma p^3FT' + \Sigma p^4FQ'$$

$$\begin{aligned} B' = {}& (1 - \Sigma p^2)FB' + 2(\Sigma p^2 - \Sigma p^3)FT' \\ & + 4(\Sigma p^3 - \Sigma p^4)FQ' \end{aligned}$$

$$\begin{aligned} CC' = {}& (1 - \Sigma p^2)FCC' + (\Sigma p^2 - \Sigma p^3)FT' \\ & + 3([\Sigma p^2]^2 - \Sigma p^4)FQ' \end{aligned}$$

$$\begin{aligned} T' = {}& (1 - 3\Sigma p^2 + 2\Sigma p^3)FT' \\ & + 6(\Sigma p^2 - 2\Sigma p^3 + 2\Sigma p^4 - [\Sigma p^2]^2)FQ' \end{aligned}$$

$$Q' = (1 - 6\Sigma p^2 + 8\Sigma p^3 - 6\Sigma p^4 + 3[\Sigma p^2]^2)FQ'$$

If all alleles at the locus in the population are of equal frequency, one may simplify; if there are 'm' such alleles, each of frequency $1/m$, probabilities are:

$$A' = FA' + (FB' + FCC')/m + FT'/m^2 + FQ'/m^3$$

$$B' = (m-1)(FB'/m + 2FT'/m^2 + 4FQ'/m^3)$$

$$CC' = (m-1)(FCC'/m + FT'/m^2 + 3FQ'/m^3)$$

$$T' = (m-1)(m-2)(FT'/m^2 + 6FQ'/m^3)$$

$$Q' = (m-1)(m-2)(m-3)FQ'/m^3$$

Computer simulation

Using a 32K microcomputer, two numeric arrays are established, each equivalent to a table of 500 rows and four columns, the rows to represent genotypes and the columns to contain the probabilities (coefficients) of random loci of the genotype being homozygous (FA'), triplex (FB'), double-duplex (FCC'), or tri-allelic (FT'). The probability (FQ') of a locus being quadri-allelic is obtained by subtraction from unity. (A larger computer memory would permit a larger population to be simulated, but 500 is a reasonable number.) On commencement, all rows in the first array are set to contain identical values (usually equivalent either to $FQ' = 1$ or to $FA' = 1$). 'S', the frequency of selfing, is entered.

Then, $500S$ random selections are made among the 500 'parental' genotypes in the first array and processed according to the 'rules' for selfing (see below), the resultant probability estimates being entered into the first $500S$ rows in the second array. Each of the remaining $500(1-S)$ 'places' in the second array is filled by making two random selections among the 500 genotypes in the first array and processing their values according to the 'rules' for crossing. The values in the second array, representing the first derived generation, are then transferred into the first array and the process is repeated to obtain a second derived generation, and so on.

For reasons given previously, double reduction is ignored. On this simplified basis:
On selfing:

$$FA', n+1 = (FA' + FB'/4 + FCC'/18 + FT'/36), n$$

$$FB', n+1 = (FB'/2 + 4FCC'/9 + 2FT'/9), n$$

$$FCC', n+1 = (FB'/4 + FCC'/2 + FT'/4 + FQ'/6), n$$

$$FT', n+1 = (FT'/2 + 2FQ'/3), n$$

$$FQ', n+1 = FQ', n/6$$

On crossing: One considers the probabilities of a gamete from a parent being homozygous or heterozygous at a locus. (Assuming, for estimation of coefficients, that each allele is of infinitesimal frequency, the probability of the two parents being crossed having alleles in common is zero.) Where $\alpha = 0$ the probability of a gamete being homozygous is Fg as given above, and the probability of its being heterozygous is Ga. Then, using 'f' to indicate the female and 'm' the male parent:

$$FCC', n+1 = (Fg, \mathrm{f} \times Fg, \mathrm{m}), n$$

$$FT', n+1 = [(Fg, \mathrm{f} \times Ga, \mathrm{m}) + (Fg, \mathrm{m} \times Ga, \mathrm{f})], n$$

$$FQ', n+1 = (Ga, \mathrm{f} \times Ga, \mathrm{m}), n$$

($FA', n+1$ and $FB', n+1$ are zero).

Simulation of selection against inbreds is effected by setting a 'rejection' value of FA', or of Fg or whatever indicator of inbreeding is under study, at and above which it is assumed that a genotype in natural conditions will not survive to reproduce, or in a selection programme will not be kept as a parent, or will be completely sterile. As each random choice of a genotype is made from the parental array, a test value is obtained by multiplying the set 'rejection' value by a random value in the range from zero to unity, and if the value for the potential parent exceeds the test value, that parent is rejected and a further random choice is tested against another, similarly obtained test value. Thus, if selection based on FA' is being studied and the set rejection value is 1, the actual probability of a random selection being rejected is directly proportional to its value of FA'. If the set value is 0.5, the actual probability is $FA'/0.5$, thus double the value of FA'; for example, a selection with $FA' = 0.2$ will have a 40% probability of being rejected (60% of acceptance) while one with $FA' = 0.5$ or more will always be rejected. Mild selection can be simulated by setting a rejection value in excess of unity; thus if the set value is $FA' = 2$, the actual probability of rejection is $FA'/2$. The absence of selection against inbreds can be simulated by making the set rejection value a very high one, e.g. $FA' = 100$.

Computations and comments

Tables 1–4 are based on the algebraic formulae and do not take into account selection against inbreds.

Table 1 shows equilibrium values of the coefficients FA' (homozygous), FB' (triplex), FCC' (double-duplex), FT' (tri-allelic) and FQ' (quadri-allelic), together with Fg, the generalised inbreeding coefficient (see above), for a range of frequencies of selfing ('S') and for zero and for 15% double reduction ('α') at the locus. With no double reduction, FA' is effectively zero with less than 30% selfing, and it is only with more than 90% selfing that it exceeds 0.5; FQ' exceeds 0.5 with up to about 40% selfing, and it is only with more than 90% selfing that it becomes effectively zero. Of the coefficients for lesser degrees of heterozygosity, FT' attains a maximum of about 0.37 with 50% selfing and FCC' and FB' (both bi-allelic states) maxima of about 0.24 and 0.21, respectively, with 80% selfing. The generalised inbreeding coefficient Fg is trivial where the frequency of selfing is low, and it

Table 1. Equilibrium frequencies of coefficients[a], and generations to effective equilibrium[b], for a locus in an autotetraploid population where 'S' is the frequency of selfing and 'α' the frequency of double reduction, assuming there is no selection against inbreds. From algebraic formulae

S	FA' (%)	FB' (%)	FCC' (%)	FT' (%)	FQ' (%)	Fg	Generations to equilibrium from	
							FA', o = 1	FQ', o = 1
With $\alpha = 0$								
0.00	0	0	0	0	100	0.00	5	–
0.10	0	0	2	11	87	0.03	6	2
0.20	0	1	4	21	73	0.06	6	3
0.30	1	3	7	29	60	0.10	7	4
0.40	2	6	10	34	47	0.14	7	5
0.50	4	10	14	37	35	0.20	8	5
0.60	8	14	18	36	24	0.27	8	6
0.70	15	18	22	31	14	0.37	10	8
0.80	28	21	24	22	6	0.50	12	11
0.90	52	18	20	9	1	0.69	13	15
0.95	72	12	13	3	0	0.83	12	20
1.00	100	0	0	0	0	1.00	–	20
With $\alpha = 0.15$								
0.00	0	0	4	33	63	0.07	5	2
0.10	0	2	7	37	53	0.10	5	3
0.20	1	5	10	40	44	0.14	5	3
0.30	2	8	13	41	36	0.18	6	4
0.40	5	12	16	41	27	0.23	6	4
0.50	8	16	19	38	20	0.29	6	4
0.60	14	19	22	33	13	0.36	6	6
0.70	23	21	23	26	7	0.45	7	8
0.80	37	22	23	16	3	0.58	9	10
0.90	60	16	17	6	1	0.75	10	13
0.95	77	10	11	2	0	0.86	9	15
1.00	100	0	0	0	0	1.00	–	18

[a] Coefficients: Probabilities, as regards alleles distinct in origin, of random loci being: FA'–Homozygous (One allele, quadruplex); FB'–Triplex (Two alleles, triplex and simplex); FCC'–Double-duplex (Two alleles, both duplex); FT'–Tri-allelic (Three alleles, one duplex, two simplex); FQ'–Quadri-allelic (Four alleles, each simplex). Also Fg–Probability of two alleles taken randomly from a random locus being identical

[b] Equilibrium assumed when each coefficient is within 0.02 of its final value

is only with more than 80% selfing that it exceeds 0.5. Double reduction at the level investigated (15%, near to its theoretical maximum) has a broadly similar effect to an increase in the frequency of selfing of about 30% where the actual level of selfing is low, falling to about 10% when actual selfing is 50%–70% and less at higher rates of selfing; the main exception to this generalisation is that double reduction alone, or with very low rates of selfing, does not give rise to significant values of FA'. The approach to equilibrium is fairly rapid where S is low, but where S is high it is very slow; in theory, equilibrium is never quite attained but the table shows the number of generations required for all coefficients to be within 0.02 of their final values. Double reduction somewhat increases the rate of approach.

Table 2 shows the approach to equilibrium under random mating commencing from a fully inbred population. FA' and FB' do not occur and, both with zero and with 15% double reduction, equilibrium is virtually achieved by the fifth generation.

The need for five generations of random mating to achieve virtual equilibrium implies that to obtain an almost fully outbred genotype by controlled breeding would require commencement from 32 ancestors totally unrelated to each other (crossed in 16 pairs, then to give 8 combinations of 4, etc.); it would be difficult, in practice, to obtain such a number of appropriate ancestors, as most cultivars are somewhat interrelated. Even then, many loci affected by double reduction would not be quadri-allelic. The fully heterozygous state in autotetraploids thus appears to be a theoretical abstraction.

Table 3 shows the approach with continuous inbreeding commencing from a fully outbred population. FQ' is effectively eliminated by the third and FT' by the 9th generation but equilibrium is not achieved within 20 generations, though with 15% double reduction it is virtually

Table 2. The approach towards equilibrium with fully random mating ($S = 0$) commencing from complete homozygosity ($FA'_0 = 1$). From algebraic formulae. No selection against inbreds

Generations	FA' (%)	FB' (%)	FCC' (%)	FT' (%)	FQ' (%)	Fg
With $\alpha = 0$						
1	0	0	100	0	0	0.33
2	0	0	11	44	44	0.11
3	0	0	1	20	79	0.04
4	0	0	0	7	93	0.01
5	0	0	0	2	98	0.00
At equilibrium:	0	0	0	0	100	0.00
With $\alpha = 0.15$						
1	0	0	100	0	0	0.33
2	0	0	19	49	32	0.14
3	0	0	7	40	53	0.09
4	0	0	5	35	60	0.07
5	0	0	5	34	62	0.07
At equilibrium:	0	0	4	33	63	0.07

Table 3. The approach towards equilibrium with full inbreeding ($S = 1$) commencing with all genotypes fully outbred ($FQ'_0 = 1$). From algebraic formulae. No selection against inbreds

Generations	FA' (%)	FB' (%)	FCC' (%)	FT' (%)	FQ' (%)	Fg
With $\alpha = 0$						
1	0	0	17	67	17	0.17
2	3	22	28	44	3	0.31
3	11	33	31	24	0	0.42
4	22	36	30	12	0	0.52
5	33	34	27	6	0	0.60
6	43	30	24	3	0	0.67
7	52	26	20	2	0	0.72
8	60	22	17	1	0	0.77
9	66	19	14	0	0	0.81
10	72	16	12	0	0	0.84
11	76	13	10	0	0	0.87
12	80	11	8	0	0	0.89
13	84	9	7	0	0	0.91
14	86	8	6	0	0	0.92
15	89	6	5	0	0	0.94
16	91	5	4	0	0	0.95
17	92	5	3	0	0	0.95
18	93	4	3	0	0	0.96
19	95	3	2	0	0	0.97
20	95	3	2	0	0	0.97
At equilibrium:	100	0	0	0	0	1.00
With $\alpha = 0.15$						
1	1	13	14	61	12	0.22
2	9	33	23	33	1	0.39
3	22	38	24	15	0	0.52
4	36	35	22	6	0	0.62
5	49	30	18	3	0	0.71
6	60	24	15	1	0	0.77
7	68	20	12	1	0	0.82
8	75	16	9	0	0	0.86
9	80	12	7	0	0	0.89
10	85	10	6	0	0	0.91
11	88	8	5	0	0	0.93
12	91	6	4	0	0	0.95
13	93	5	3	0	0	0.96
14	94	4	2	0	0	0.97
15	95	3	2	0	0	0.97
16	96	2	1	0	0	0.98
17	97	2	1	0	0	0.98
18	98	1	1	0	0	0.99
19	98	1	1	0	0	0.99
20	99	1	1	0	0	0.99
At equilibrium:	100	0	0	0	0	1.00

achieved by the 18th. To inbreed to virtual homozygosity, even if not prevented by infertility, would take a great many years and the fully homozygous state in autotetraploids also appears to be a theoretical abstraction.

It appears, therefore, that any autotetraploid genotype obtained by normal means will be partially, though not completely, inbred. A complete inbred could perhaps be obtained by twice doubling the chromosomes of a monoploid genotype, and a fully heterozygous genotype by somatic fusion of two unrelated and fully heterozygous diploid genotypes, but not naturally or by normal breeding procedures. The 'commencements' in the tables are thus somewhat artificial and more realistic commencements would be genotypes which are already partially inbred.

Before further discussion of the tables, some comments on the coefficient FA', which resembles F in diploids in being associated with homozygosity, are desirable. It is calculated by subtraction and one may write $FA' = 2Fg - 1 + Gb/6 + Gc/4$. The approaches of Gb and Gc to equilibrium, determined by γ and δ, are more rapid than that of Fg, determined by β (always a larger value) and, after Gb and Gc stabilise, one may write $FA'(n) = 2Fg(n)$ minus a constant, whence $DFA'(n+1) = \beta DFA'(n)$. Prior to Gb and Gc stabilising, β will be the major determinant and the latter equation will be approximately correct beyond the earliest generations from maximum heterozygosity, and probably immediately from a more realistic commencement.

With continuous inbreeding ($S = 1$) the equilibrium values of Gb and Gc are zero and, once they stabilise, $FA'(n)$ will be $2Fg(n) - 1$. In Table 3, commencing from maximum heterozygosity, this is approximately correct from the 9th generation with $\alpha = 0$ and the 7th where $\alpha = 0.15$.

Another relationship apparent in Table 3, continuous inbreeding from maximum heterozygosity, is that, from the 5th generation onwards, FA' in one generation is similar to Fg three generations previously. In these circumstances $Fg, n = 1 - \beta^n$ and $\beta = (5 - 2\alpha)/6$, effectively

5/6. Thus where n exceeds 4 from maximum heterozygosity, $FA', n \approx 1 - (5/6)^{n-3}$. It follows that, where n exceeds 4 from maximum heterozygosity, $FA'(n) \approx 1.73\,Fg(n) - 0.73$; from a realistic commencement this may be so at once. This, however, is a chance finding, not one derived algebraically and, where S is less than unity, no simple relationship between FA', n and Fg, n (prior to Gb and Gc stabilising) has been detected.

In diploids, F is 0.5 in the first selfed generation, 0.75 in the second, 0.875 in the third, 0.94 in the fourth; in autotetraploids (with $\alpha = 0$), commencing from maximum theoretical heterozygosity, FA' approximates to these values in the 7th, 11th, 15th and 19th generations (Table 3). Thus, from a realistic commencement, four generations of selfing of an autotetraploid are required to increase homozygosity to the same extent as is effected by a single generation of selfing of a diploid.

Table 4. Equilibrium frequencies of actual values of A' (homozygotes) to Q' (quadri-allelic states) for a locus at which there are 40 alleles of equal frequencies. From algebraic formulae. No double reduction. No selection against inbreds

S	A' (%)	B(%)	CC'(%)	T' (%)	Q' (%)
0	0	0	0	14	86
0.1	0	1	2	22	74
0.2	0	3	5	29	63
0.3	1	5	8	35	52
0.4	2	8	11	38	40
0.5	5	11	15	39	30
0.6	9	15	19	37	20
0.7	16	19	22	31	12
0.8	29	21	24	21	5
0.9	53	18	20	8	1
0.95	72	12	13	3	0
1.00	100	0	0	0	0

Table 4 illustrates the effect on equilibrium population structures (with $\alpha = 0$) of the actual number of alleles at the locus, in this case taken as numbering 40, each with equal frequency, 'p' = 0.025. Relative to Table 1, the main effect is a reduction in the equilibrium frequency of quadri-allelic genotypes, these mostly being replaced by tri-allelics where S is low, though these are also reduced where S exceeds 0.7. The bi-allelic classes CC' and B' are also increased, most significantly in the range $S = 0.2$ to 0.6, but effects on A' are slight. The frequency of quadri-allelic genotypes is 86% of FQ'; it is apparent from the equations that it would be about 93% of FQ' if there were 80 alleles and 96% if there were 160, but that the equilibrium frequency of homozygous loci would be increased substantially only when the number of alleles is very low. Inequalities in allelic frequencies would accentuate these effects.

Tables 5–7 are based on 'runs' of the computer simulation programme, in all cases ignoring double reduction. Due to the random process involved, no two runs with the same commencements and conditions are likely to give identical results; however, those which have been made with no selection against inbreds have given results which match closely those, with $\alpha = 0$, obtained using the algebraic formulae. Even beyond 'equilibrium', slight fluctuations occur from generation to generation.

Table 5a illustrates the distribution of inbreeding levels, as shown by the values of FA' of the genotypes, in a population at equilibrium with 80% selfing, following various levels of selection against inbreds (presumed to be equivalent to selection for vigour and fertility). Selection is based on the values of FA' of the potential parents and population structures are determined by averaging the frequencies over advanced generations. In the absence of selection, 31% of the genotypes have values of FA' over

Table 5. From computer simulation: with $S = 0.8$ ($\alpha = 0$), distribution of inbreeding levels of genotypes in the equilibrium population with no, mild, moderate or severe selection against inbreeding

Selection pressure[a]	Frequencies (%) of genotypes with **a** FA' or **b** Fg				
	0.8 or more	0.6–0.8	0.4–0.6	0.2–0.4	Under 0.2
a					
None	9	11	11	18	51
Mild	0	6	14	18	62
Moderate	0	2	11	21	66
Severe	0	0	3	21	76
b					
None	18	19	24	25	15
Mild	4	21	28	29	19
Moderate	1	19	30	30	20
Severe	0	9	34	36	21

[a] Probabilities of a genotype not functioning as a parent proportional to:
Mild selection: Half its value of FA'
Moderate selection: Its value of FA'
Severe selection: Twice its value of FA'

0.4, including 9% with values over 0.8. Even mild selection against inbred parents changes this pattern, and with severe selection only 3% of genotypes have values over 0.4. The proportions of relatively outbred genotypes, with values of FA' under 0.2, rise from 51% with no selection to 76% with severe selection against inbreds. Table 5b shows the distribution of inbreeding levels in the same populations (the same computer runs, selection still being based on FA'), assessed using the generalised inbreeding coefficient Fg. The total frequency of genotypes with values of Fg exceeding 0.4 falls from 61% with no selection to 43% with severe selection (of genotypes with Fg exceeding 0.6, from 37% to 9%), while the proportion appearing relatively highly outbred, with Fg under 0.2, rises from 15% to 21%.

Table 6 shows the evolution of the above populations, commencing with all genotypes fully outbred, as regards population mean values (coefficients). In the computer run with no selection, equilibrium values as in Table 1 were virtually attained in 13 generations. With mild selection the equilibrium value of FA' was reduced, as was Fg, while the values of FT' and FQ' were raised, and while due to continuing fluctuations it is difficult to specify the exact stage at which effective equilibrium was reached, this was clearly earlier (about generation 7). With moderate selection these trends were more marked and with severe selection even more so; FA' was substantially reduced, as was Fg, and effective equilibrium was attained in about four generations.

Table 6. From computer simulation: the approach to equilibrium with $S = 0.8$ ($\alpha = 0$) with selection against inbreds, commencing with all genotypes fully outbred

Generation	FA' (%)	FB' (%)	FCC' (%)	FT' (%)	FQ' (%)	Fg
No selection						
1	0	0	13	53	33	0.13
2	2	15	21	43	19	0.23
3	6	21	23	35	15	0.30
4	11	23	24	31	12	0.35
5	15	22	24	28	11	0.39
6	17	23	24	27	9	0.41
7	19	22	24	26	9	0.43
8	22	22	24	25	7	0.45
9	23	22	24	24	7	0.46
10	24	22	24	23	7	0.47
11	24	22	24	23	7	0.47
12	25	22	24	22	6	0.48
13	27	21	24	22	6	0.49
14	27	20	24	23	6	0.49
15	28	21	24	22	6	0.50
Mild selection, proportional to half FA'						
1	0	0	13	53	33	0.13
2	2	14	20	44	20	0.23
3	6	20	23	36	15	0.29
4	10	23	24	30	12	0.35
5	15	23	24	27	10	0.39
6	18	23	24	27	9	0.42
7	19	22	24	26	8	0.43
8	20	22	24	26	8	0.43
9	20	22	24	25	9	0.43
10	21	22	23	25	9	0.44
11	21	22	24	26	8	0.44
Moderate selection, proportional to FA'						
1	0	0	13	53	33	0.13
2	2	14	21	44	20	0.23
3	6	21	24	35	14	0.30
4	11	23	24	29	12	0.36
5	14	23	24	28	10	0.39
6	16	22	24	28	9	0.40
7	17	22	24	27	9	0.41
8	16	22	24	28	10	0.40
Severe selection, proportional to twice FA'						
1	0	0	13	53	33	0.13
2	2	14	21	44	20	0.23
3	6	21	23	35	15	0.30
4	10	22	24	31	12	0.34
5	11	21	23	33	12	0.35
6	12	22	24	30	12	0.36
7	12	22	24	31	12	0.36

Similar studies were made with 50%, 70% and 90% selfing; findings are summarised in Table 7. Selection against FA' has little effect where the selfing frequency is low and few genotypes would attain high levels of inbreeding even in the absence of selection. At selfing frequencies up to 0.8, selection tends to increase FQ' and to a lesser extent FT', and, where $S = 0.9$, FB' and FCC' are also increased. Severe selection against inbreds leads to fairly rapid attainment of equilibrium even at high selfing frequencies.

Studies have also been made with selection operated against the value of Fg of the parents (not tabulated). This is more severe in its effects than selection against FA'. With selection based on twice the value of Fg of the parents, with $S = 0.5$ the equilibrium frequencies of FA' and FQ' were 1% and 46%, respectively, attained in three generations; with $S = 0.8$ they were 5% and 17%, also attained in three generations; and with $S = 0.9$, 8% and 9%, attained in four generations (compare with Table 7).

While only simple selection applied uniformly to all parents is considered above, a full simulation of a population with mixed selfing and intercrossing would require at least two assessments, with differing selection intensities, of each potential parent. Inbreds are likely to be smaller plants bearing fewer flowers and yielding less seed; in natural situations they may be less likely to survive. A strong initial selection against inbreeding levels may be appropriate to simulate these aspects. However if, in a mass-selection programme, there are several selection criteria among which yield is not pre-eminent, and equal quantities of seed are kept from each chosen parent, a less severe selection pressure would seem appropriate as

Table 7. From computer simulation: equilibrium frequencies under selection against *FA'*

Selfing frequency	Selection pressure[a]	*FA'* (%)	*FB'* (%)	*FCC'* (%)	*FT'* (%)	*FQ'* (%)	*Fg*	Generations to effective equilibrium
0.5	None	5	10	14	36	35	0.20	5
	Mild	4	9	14	37	37	0.19	5
	Moderate	4	9	14	37	37	0.19	5
	Severe	3	9	14	37	38	0.18	4
0.7	None	14	18	21	32	15	0.37	8
	Mild	12	18	22	33	16	0.33	6
	Moderate	10	18	21	34	17	0.32	6
	Severe	8	17	20	35	20	0.29	5
0.8	None	28	21	24	22	6	0.50	13
	Mild	20	22	24	26	8	0.43	7
	Moderate	17	22	24	27	9	0.41	6
	Severe	12	22	24	30	12	0.36	4
0.9	None	51	18	20	10	1	0.69	16
	Mild	36	24	23	14	3	0.58	8
	Moderate	26	27	26	18	4	0.52	6
	Severe	19	27	27	23	5	0.45	5

[a] Selection pressures as in Table 5

regards choice of parents. But in both cases, a strong simulated selection may be appropriate as regards the probability of such seed as is obtained from inbreds, which may be less pollen fertile, being selfed. Even stronger simulated selection may be appropriate as regards the male parents of crosses to take account, jointly, of the probabilities that free-flowering outbreds will contribute disproportionately to the pollen transported within the population and that their pollen will be more fertile. Simulations of this nature are not reported here, as the possible permutations of parameters are innumerable.

Discussion: mass selection

In a population in which both selfing and intercrossing are taking place, there will almost inevitably be some selection against inbreds, leading to a more outbred population structure than would be expected on the basis of the selfing frequency. But even strong selection for vigour cannot prevent the population becoming somewhat inbred (Tables 5–7).

The population will consist, in any generation, of a mixture of inbred lineages and of crosses between them, the latter potentially the origins of future lineages. The effect of selection for vigour will be to limit the duration of lineages. Comparison of Tables 3 and 5 suggests that at equilibrium with 80% selfing and severe selection, only a few (3%) of the genotypes will be inbred to an extent equivalent to six generations from maximum heterozygosity (*FA'* 0.4 or more); even with only mild selection, few will be eight generations inbred. As the crosses between inbreds from which new lineages originate are not fully heterozygous, the actual duration of lineages will usually be less than this.

Mass selection programmes are likely to be concerned with individual characters as much as or more than with vigour. Selection can only be operated on readily discernible characters but it may be important that other characters, such as disease resistances or quality factors, detectable only by tests which cannot be applied on a population basis, should remain present. Selection on individual characters is not considered above but one may draw inferences.

On random mating, the expected frequency of homozygotes for an allele is the fourth power of its frequency; e.g., a recessive with frequency 0.1 will be exposed in only 1 seedling in 10,000. Assuming that a reasonable sample of the population is kept as parents of the next generation (to maintain genetic diversity), inclusion of homozygotes of desirable but scarce recessives in that sample will be virtually without effect on the frequencies of the alleles. A measure of inbreeding will increase the exposure of recessive alleles and greatly improve the efficiency of selection for desirable recessives, thus against undesirable dominants. It may also facilitate selection on quantitatively inherited characters.

Where inbreeding is taking place, alleles will not be randomly distributed. Each will occur in one or more lineages and be absent from others; each will tend to be associated with those others which happen to occur in the same lineages.

A lineage having desirable characters may increase in the population initially and, through crosses with others similary favoured, become a parent of new lineages before itself declining as inbreeding becomes more advanced.

This may permit rapid improvement of the population as regards the characters under selection.

However, the expansion of some lineages implies the reduction and possible elimination of others. The persistence of alleles at loci controlling characters not under selection depends on their occurrence in lineages favoured for other reasons. The risk of loss of alleles may be greatest in the earliest generations when lineages may trace to individual founder-members and be the only ones bearing alleles unique to those ancestors, but there will be a continuing risk of loss at subsequent stages. To maintain genetic diversity it may, therefore, be advisable: (1) to intercross manually to disrupt character associations before resorting to natural pollination and commencing selection; (2) to apply relatively gentle selection pressures and, in particular, to avoid strong selection for one or a few sharply defined phenotypes which may represent individual lineages; (3) to discontinue mass selection with reliance on natural pollination when clear advances from generation to generation cease to be apparent.

Should the aim be to issue the population as a cultivar propagated by true seed, it may be desirable to maximise heterozygosity and reduce overt variation by intercrossing manually for a generation or two before release, using pollen collected in bulk, and to maintain the stock to permit repeated re-issue. If this is the aim, maintenance of genetic diversity may appear unneccessary or even undesirable, but one should note that if diversity has been reduced excessively, the heterozygosity attainable will be limited. Selection could continue during the intercrossing phase, though essentially only to increase or decrease the frequencies of scarce dominants.

If the aim is the provision of parents for crossing with unrelated stocks, the maintenance of diversity, to provide parents for various breeding objectives not necessarily specified in advance, will probably be a major objective. Selection may be aimed at elimination of undesirable characters, but retention of as much acceptable variation in observable characters as possible; this would increase the probability of other unobservable variation remaining present. One may place little emphasis on vigour and depend on restoration of heterozygosity in the subsequent crosses with unrelated parents to ensure adequate vigour and yield in the resultant cultivars. However, crosses between even totally unrelated inbreds will not be of maximum heterozygosity; they should have no homozygous or triplex loci but will have some double-duplex and tri-allelic loci. Whether the presence of the latter causes a reduction in performance relative to that attainable with all loci quadri-allelic is uncertain, but manual crossing within the population, to provide more outbred parents, might be advisable. Some controlled breeding may, in any case, be desirable to combine or enhance characters not considered in the mass-selection process, detected by screening selections from the advanced population.

Acknowledgements. Thanks are due to J. H. Bennett for constructive discussions, to F. J. W. England and G. C. C. Tai for encouragement, and to P. D. S. Caligari for comments on a draft.

Appendix 1: misprints in Bennett (1968)

In the expression for the equilibrium frequency of the duplex genotype (p. 495), a bracket has been misplaced. For convenience, multiplying both numerator and denominator by four, it should read:

$$\frac{8(1-s)\,p(A)\,p(a) + 6(1-s)\,p\infty^2(Aa) + [6s - 4 + 3s\alpha^2/(1-\alpha)]\,p\infty(Aa)}{4 - [4s(1-\alpha)^2/6]}$$

In Bennett's term 'C', $p(A)\,p(a)$ should be replaced by $p\infty(Aa)$ so that it reads:

$$(1-s)\,[p(a) - p(A)]\,p\infty(Aa)/(2 - s + s\alpha)$$

Bennett (personal communication) states that these misprints do not affect anything later in the paper (e.g. tables) where correct values were used. He adds, however, that $\alpha = 0.1$ was omitted from the heading to Table 6 and that, near the top of p. 497, R(n) and S(n) should be R′(n) and S′(n).

Appendix 2: derivation of formulae

Bennett (1968) explored the dynamics of a population assuming there were two alleles at the locus. His findings were thus, in effect, for a single allele, the other being equivalent to all other alleles pooled. One may extrapolate from his findings to determine the total frequencies of homozygous and triplex loci, but not to distinguish between double-duplex, tri-allelic and quadri-allelic loci. The present author's contribution to the algebraic approach is thus in such extrapolation, and in the derivation of Q' and FQ' which, in turn, permit determination of the other parameters. Although they are presented in a different form, many of the values given above can be verified against those given by Bennett and lengthy presentations of their derivations therefore seem unecessary; however, derivation of Q' will be given and it serves to demonstrate the general method employed.

If the four alleles at a locus in a quadri-allelic genotype are coded a, b, c and d, one finds that six gametes are formed by 'normal' means, each with frequency $(1-\alpha)/6$:

ab, ac, ad, bc, bd, cd

and four by 'double reduction', each with frequency $\alpha/4$:

aa, bb, cc, dd

Laying out the 10 × 10 grid, one finds that the selfed progeny of a quadri-allelic genotype contains:

$(1-\alpha)^2/6$	quadri-allelic,
$4(1-\alpha)^2/6+\alpha(1-\alpha)$	tri-allelic,
$(1-\alpha)^2/6+3\alpha^2/4$	double-duplex,
$\alpha(1-\alpha)$	triplex, and
$\alpha^2/4$	homozygous genotypes.

Thus, on selfing, $FQ'_1 = FQ'_0(1-\alpha)^2/6$

On random mating, quadri-allelic genotypes are formed only from combinations of heterozygous gametes. In the absence of double reduction the frequency of such gametes (assuming each allele to have infinitesimal frequency) is Ga: thus, allowing for double reduction, it is $Ga(1-\alpha)$ and they will combine with frequency $Ga^2(1-\alpha)^2$.

Thus, combining selfing and random mating:

$$\begin{aligned} FQ'_1 &= SFQ'_0(1-\alpha)^2/6+(1-S)\,Ga_0^2(1-\alpha)^2 \\ &= \gamma FQ'_0+(1-S)\,Ga_0^2(1-\alpha)^2 \end{aligned}$$

and:

$$FQ'_2 = \gamma^2 FQ'_0+\gamma(1-S)\,Ga_0^2(1-\alpha)^2+(1-S)\,Ga_1^2(1-\alpha)^2$$

$$\begin{aligned} FQ'_3 = {} & \gamma^3 FQ'_0+\gamma^2(1-S)\,Ga_0^2(1-\alpha)^2+\gamma(1-S)\,Ga_1^2(1-\alpha)^2 \\ & +(1-S)\,Ga_2^2(1-\alpha)^2 \end{aligned}$$

Substituting $Ga_\infty + Da_0$ for Ga_0, one has: $Ga_1 = Ga_\infty + \beta\,Da_0$, $Ga_2 = Ga_\infty + \beta^2\,Da_0$, etc.

$$\begin{aligned} FQ'_3 &= \gamma^3 FQ'_0+(1-S)(1-\alpha)^2\,[\gamma^2(Ga_\infty+Da_0)^2 \\ &\quad +\gamma(Ga_\infty+\beta Da_0)^2+(Ga_\infty+\beta^2 Da_0)^2] \\ &= \gamma^3 FQ'+(1-S)(1-\alpha)^2\,Ga_\infty^2(1+\gamma+\gamma^2) \\ &\quad +(1-S)(1-\alpha)^2\,2\,Ga_\infty\,Da_0(\beta^2+\beta\gamma+\gamma^2) \\ &\quad +(1-S)(1-\alpha)^2\,Da_0^2(\beta^4+\beta^2\gamma+\gamma^2) \end{aligned}$$

Using L for $(1-S)(1-\alpha)^2$ and summing the geometric progressions:

$$\begin{aligned} FQ'_n = {} & \gamma^n FQ'_0+L\,[Ga_\infty^2(1-\gamma^n)/(1-\gamma) \\ & +2\,Ga_\infty Da_0(\beta^n-\gamma^n)/(\beta-\gamma)+Da_0^2(\beta^{2n}-\gamma^n)/(\beta^2-\gamma)] \end{aligned}$$

One may extract the equilibrium value and the deviation as given previously.

When alleles of non-trivial frequency are taken into account, the effects of selfing remain as above but, as regards random mating, two further considerations apply. Firstly, the frequency of heterozygous gametes containing an allele with frequency p is $2p(1-p)(1-Fg)(1-\alpha)$ whence, summing over alleles and substituting Ga for $1-Fg$ (remembering that each such gamete contains two alleles and that $\Sigma p = 1$), the total frequency of such gametes is $Ga(1-\alpha)(1-\Sigma p^2)$. Thus, the frequency with which heterozygous gametes will combine is $Ga^2(1-\alpha)^2(1-\Sigma p^2)^2$. Secondly, heterozygous gametes may have alleles in common and, if so, they will not give quadri-allelic genotypes on combining.

The gamete containing alleles 1 and 2 has frequency $2\,Ga(1-\alpha)^2 p_1 p_2$ and the double-duplex genotype containing these alleles will occur with the square of that frequency; the same applies to alleles 1 and 3, 1 and 4 etc. Totalling, one finds that double-duplex genotypes containing allele 1 will be formed with frequency $4\,Ga^2(1-\alpha)^2 p_1^2(\Sigma p^2 - p_1^2)$; similarly, those containing allele 2 will occur with frequency $4\,Ga^2(1-\alpha)^2 p_2^2(\Sigma p^2 - p_2^2)$. Totalling, as each double-duplex genotype contains two alleles it will be counted twice; thus, the frequency with which double-duplex genotypes will be formed is $2\,Ga^2(1-\alpha)^2((\Sigma p^2)^2 - \Sigma p^4)$.

Heterozygous gametes containing allele 1 have frequency $2\,Ga(1-\alpha)p_1(1-p_1)$ and genotypes duplex for that allele will be formed with the square of that frequency; the same is true for each other allele. Totalling, the frequency of duplex loci in the derived random-mated population is $4\,Ga^2(1-\alpha)^2(\Sigma p^2 - 2\Sigma p^3 + \Sigma p^4)$. Deducting twice the freqency of double -duplex loci, one finds that the frequency with which tri-allelic loci is formed is $4\,Ga^2(1-\alpha)^2(\Sigma p^2 - 2\Sigma p^3 + 2\Sigma p^4 - (\Sigma p^2)^2)$.

Deducting the frequency of double-duplex and of tri-allelic genotypes from the total of offspring from combinations of heterozygous gametes, one obtains the total frequency of quadri-allelic loci from random mating as being $Ga^2(1-\alpha)^2[1-6\Sigma p^2 + 8\Sigma p^3 - 6\Sigma p^4 + 3(\Sigma p^2)^2]$.

References

Bennett JH (1968) Mixed self- and cross-fertilization in a tetrasomic species. Biometrics 24:485–500

Demarly Y (1961) In: Valdeyron G. Genètique et amélioration des plantes. Bailliere, Paris, pp 318–335

Gallais A (1968) Interactions between alleles and their variability in autotetraploid cross-fertilised plants – consequences for selection. 5th Conf European Association Res Plant Breed, Milan, pp 312–323

Glendinning DR (1976) Neo-Tuberosum: New Potato breeding material. 4. The breeding system of Neo-Tuberosum, and the structure and composition of the Neo-Tuberosum gene-pool. Potato Res 19:27–36