Theor. Appl. Genet. 59, 73-77 (1981)

Methods for the Production of Multi-marker Strains

J. Weller and M. Soller
Department of Genetics, The Hebrew University, Jerusalem (Israel)

Summary. The production of a multi-marker strain given a series of strains each carrying a single marker will require a considerable investment in time and resources when based on a series of formal crosses. Such strains can be produced rapidly and with only minimal resources by recurrent cycles of random mating and selection, with selection based on the number of desired alleles carried.

Key words: Linkage – Quantitative traits – Multi-marker strain

Introduction

In this paper we describe methods for the production of a multi-marker strain incorporating a large number of Mendelian markers, given a series of strains each carrying one of the markers singly. Although the individual markers could be of direct economic importance (e.g., resistance genes) our interest in such lines lies in their potential usefulness in experiments aimed at identifying linkage between marker loci and loci affecting quantitative traits (Thoday 1961; Patterson et al. 1968; Chai 1975; Zhuchenko et al. 1979). In such experiments the F-2 or backcross progeny of crosses between inbred lines differing at one or more marker loci are analyzed to identify marker-linked chromosomal neighborhoods affecting specific quantitative traits (Soller et al. 1976). A single cross of this sort can serve to simultaneously evaluate the chromosomal neighborhoods of all markers differentiating the lines (Spickett and Thoday 1966). Thus, depending on the effect of the quantitative loci and on their specific location with respect to the markers (Soller et al. 1979), crosses between a line to be evaluated and a suitable multi-marker tester strain could provide the general chromosomal location of all or most loci differentiating tester and line with respect to any given set of quantitative traits.

Theory

In a multi-marker strain intended for use as a tester in linkage studies, there should be at least two well-separated markers present on each chromosome. In the calculation to follow therefore, it is assumed that all markers are unlinked. Weak linkage will not affect the results appreciably.

Given n markers, M_1 ..., M_n, each present in a different parental strain, two alternative strategies can be envisioned for the production of a single multi-marker strain, homozygous for all n markers. These are, (1) a formal series of crosses, in which specific pre-planned genotypes are produced and crossed each generation, and (2) successive cycles of random mating and selection, in which those individuals carrying the greatest number of marker alleles are selected and mated at random each generation. Within each strategy the specific tactics employed will depend to a certain extent on the dominance relationships of the marker alleles: recessive, dominant, codominant, as the case might be.

Formal Crosses

The basic approach in this case is to build up, by successive cycles of crossing, sets of homozygous marker strains that have an increasing number of markers in common, while maintaining coverage of the entire spectrum of markers between them. Each cycle of crosses will involve two generations: (1) An F-1 generation, produced by crosses between pairs of strains sharing some markers and different in others. Each such cross produces F-1 indivi-

0040/5752/81/0059/0073/$ 1.00

Table 1. Formal crossing sequence for the production of a 10-marker tester strain

Cycle and (Generation)	Activity	Crosses
1(0)	Cross parentals	$\frac{M_1}{M_1} \times \frac{M_2}{M_2}$ $\quad \frac{M_3}{M_3} \times \frac{M_4}{M_4}$ $\quad \frac{M_5}{M_5} \times \frac{M_6}{M_6}$ $\quad \frac{M_7}{M_7} \times \frac{M_8}{M_8}$ $\quad \frac{M_9}{M_9} \times \frac{M_{10}}{M_{10}}$
1(1)	Obtain 2-marker heterozygotes and self	$\frac{M_1\ M_2}{+\ +}$ $\quad \frac{M_3\ M_4}{+\ +}$ $\quad \frac{M_5\ M_6}{+\ +}$ $\quad \frac{M_7\ M_8}{+\ +}$ $\quad \frac{M_9\ M_{10}}{+\ +}$
2(2)	Obtain 2-marker homozygotes and cross	$\frac{M_1\ M_2}{M_1\ M_2} \times \frac{M_3\ M_4}{M_3\ M_4} \times \frac{M_5\ M_6}{M_5\ M_6} \times \frac{M_7\ M_8}{M_7\ M_8} \times \frac{M_9\ M_{10}}{M_9\ M_{10}}$
2(3)	Obtain 4-marker heterozygotes and self	$\frac{M_1\ M_2\ M_3\ M_4}{+\ +\ +\ +}$ $\quad \frac{M_3\ M_4\ M_5\ M_6}{+\ +\ +\ +}$ $\quad \frac{M_5\ M_6\ M_7\ M_8}{+\ +\ +\ +}$ $\quad \frac{M_7\ M_8\ M_9\ M_{10}}{+\ +\ +\ +}$
3(4)	Obtain 4-marker homozygotes and cross	$\frac{M_1\ M_2\ M_3\ M_4}{M_1\ M_2\ M_3\ M_4} \times \frac{M_3\ M_4\ M_5\ M_6}{M_3\ M_4\ M_5\ M_6} \times \frac{M_5\ M_6\ M_7\ M_8}{M_5\ M_6\ M_7\ M_8} \times \frac{M_7\ M_8\ M_9\ M_{10}}{M_7\ M_8\ M_9\ M_{10}}$
3(5)	Obtain 4-marker heterozygotes and self	$\frac{M_1\ M_2\ M_3\ M_4\ M_5\ M_6}{+\ +\ M_3\ M_4\ +\ +}$ $\quad \frac{M_3\ M_4\ M_5\ M_6\ M_7\ M_8}{+\ +\ M_5\ M_6\ +\ +}$ $\quad \frac{M_5\ M_6\ M_7\ M_8\ M_9\ M_{10}}{+\ +\ M_7\ M_8\ +\ +}$
4(6)	Obtain 6-marker homozygotes and cross	$\frac{M_1\ M_2\ M_3\ M_4\ M_5\ M_6}{M_1\ M_2\ M_3\ M_4\ M_5\ M_6} \times \frac{M_3\ M_4\ M_5\ M_6\ M_7\ M_8}{M_3\ M_4\ M_5\ M_6\ M_7\ M_8} \times \frac{M_5\ M_6\ M_7\ M_8\ M_9\ M_{10}}{M_5\ M_6\ M_7\ M_8\ M_9\ M_{10}}$
4(7)	Obtain 4-marker 4-marker homozygotes and self	$\frac{M_1\ M_2\ M_3\ M_4\ M_5\ M_6\ M_7\ M_8}{+\ +\ M_3\ M_4\ M_5\ M_6\ +\ +}$ $\quad \frac{M_3\ M_4\ M_5\ M_6\ M_7\ M_8\ M_9\ M_{10}}{+\ +\ M_5\ M_6\ M_7\ M_8\ +\ +}$
5(8)	Obtain 8-marker homozygotes and cross	$\frac{M_1\ M_2\ M_3\ M_4\ M_5\ M_6\ M_7\ M_8}{M\ M_2\ M_3\ M_4\ M_5\ M_6\ M_7\ M_8} \times \frac{M_3\ M_4\ M_5\ M_6\ M_7\ M_8\ M_9\ M_{10}}{M_3\ M_4\ M_5\ M_6\ M_7\ M_8\ M_9\ M_{10}}$
5(9)	Obtain 4-marker heterozygotes, 6-marker homozygotes and self	$\frac{M_1\ M_2\ M_3\ M_4\ M_5\ M_6\ M_7\ M_8\ M_9\ M_{10}}{+\ +\ M_3\ M_4\ M_5\ M_6\ M_7\ M_8\ +\ +}$
10	Obtain 10-marker homozygotes	$\frac{M_1\ M_2\ M_3\ M_4\ M_5\ M_6\ M_7\ M_8\ M_9\ M_{10}}{M_1\ M_2\ M_3\ M_4\ M_5\ M_6\ M_7\ M_8\ M_9\ M_{10}}$

duals homozygous for the shared markers and heterozygous at the non-shared markers, (2) An F-2 generation produced by intercrosses or selfing involving the F-1 individuals of each cross, separately. From each such F-2 generation individuals homozygous for both shared and non-shared markers can be recovered. In this way a new series of strains is derived that have even more markers in common. This process is continued until the final marker strain having all markers is obtained.

Table 1 shows a series of such crosses for a set of 10 recessive or codominant alleles. Except for the first cycle, it is assumed that the parental strains to be crossed differ at four marker loci, and that sufficient F-2 individuals can be raised out of each cross to give reasonably high probability of recovering at least one F-2 individual homozygous for all four non-shared marker loci. Examination of Table 1 shows that individuals homozygous for 2j markers can be found among the progeny of the second (selfing) cross in the jth cycle. Thus for n markers n/2 cycles of crossing (n generations) are needed in order to incorporate all n markers in a single strain. Starting from j = 2, the number of strains at the jth cycle are $(n/2) - j + 2$ while the number of crosses that will be carried out in the jth cycle is $(n/2) - j + 1$. If N individuals are raised per cross, $N[(n/2) - j + 1]$ individuals will be raised in each F-2 generation. N will generally be chosen as to give a high probability of recovering at least one 4-way homozygote in each cross of a particular cycle. After the first cycle this probability is equal to $[1-(1-\frac{1}{4}^4)^N]^m$ where m is the number of crosses and N the total number of offspring raised per cross.

For dominant markers it will be necessary to progeny test F-2 individuals having the desired phenotype in order to identify those with the desired homozygous genotype. This will involve an extra generation at each cycle and a major increase in resources. If only a few of the n markers are dominant, it will probably be more effective to first produce a series of parental strains, each carrying all of the dominant markers in homozygous condition, and also one or two of the recessive or codominant markers. Using these strains one would proceed to carry out the same sequence of planned matings as above.

Random Mating and Mass Selection

In this approach, individuals from different strains each homozygous for one or more marker alleles, are crossed at random, producing an F-1 and F-2 generation in which the frequency of each marker allele remains at its initial frequency. It is assumed that pooled pollen or semen is used in all crosses so that a random mating situation obtains. Starting with the F-2 generation, all individuals are assigned a score according to the number of desired alleles present in homozygous state (for recessive alleles) or in homozygous or heterozygous state (for codominant or dominant alleles). Those individuals having the highest total scores are selected as parents of the next generation.

Effect of Selection on Gene Frequency

The gene frequency in the t^{th} generation of such a program, p_t, can be calculated directly from the expected distribution of marker genotypes in the population under selection, or can be approximated by the expression developed by Falconer (1960, p. 206) for the effect of selection on the frequency of alleles affecting a quantitative trait. Sample numerical calculations showed that Falconer's expression gave a close approximation to the exact results for dominant and codominant alleles from p = 0.05, and for recessive alleles from p = 0.20.

By Falconer's expression

$$\Delta p = ipq\alpha/\sigma \qquad (1)$$

where,

- i = standardized selection differential = z/P, where P is the proportion selected and z is the height of the ordinate of the normal curve at the point of truncation,
- p = the frequency of the desired allele
- q = frequency of the alternative allele,
- α = the average effect of a gene substitution (Falconer 1960 p. 129) = a + (q-p)d, where $\underline{a}$ is the deviation of each homozygote phenotype from the midparent mean, and $\underline{d}$ is the deviation of the heterozygote from the midparental mean.
- σ = the phenotypic variance = $n[2pq\alpha^2 + (2pqd)^2]$ (Falconer 1960, p. 136).

In our case $\underline{a}$ can be set equal to 1.0, and $\underline{d}$ to +a, 0, and −a for dominant, codominant and recessive alleles, respectively. Substituting these values in (1) gives:

$\Delta p = ipq/[n(1-p^2)]^{1/2}$ for recessive alleles

$\Delta p = ipq/(2npq)^{1/2}$ for codominant alleles, and

$\Delta p = ipq/[n(1-q^2)]^{1/2}$ for dominant alleles

The frequency of the desired alleles in any particular generation, p_t, can be obtained by repeated application of these expressions.

For recessive alleles, an exact expression for gene frequency of the desired marker alleles in the selected population was used when $p < 0.20$. In this case gene frequency in the m selected individuals of the t^{th} generation ($p_t + 1$) is given by the weighted average of gene frequency in the $\bar{k}$ out of m individuals that are homozygous for

the given marker allele (in these individuals, p = 1.0) and in the $(m-\bar{k})$ individuals that show the alternative marker phenotype (in these individuals, $p = p_t/(1 + p_t)$. That is,

$$p_{t+1} = \frac{\bar{k} + (m-\bar{k})p_t/(1 + p_t)}{m}$$

The average number of individuals homozygous for the recessive marker for any given allele, $\bar{k}$, will be equal to the average number of homozygous markers per selected individual,

$$\bar{k} = \frac{\sum_{k=m-r+1}^{m} kN_k + (m-r)B}{m}$$

where,

N_k = the number of individuals, each homozygous for k markers, in the selected group, and is equal (by the binomial distribution) to $NC_n^k (p_t^2)^k (1-p_t^2)^{n-k}$,

r is chosen so that $\sum_{k=m-r}^{m} N_k \geqslant m \geqslant \sum_{k=m-r+1}^{m} N_k$,

and $B = m - \sum_{k=m-r+1}^{m} N_k$

The Number of Offspring Raised per Generation

In a program of this sort it is essential that each allele be represented a number of times in the group of selected individuals. This is particularly true at the early stages of the selection program when p is small. The sampling variance of p is (pq/2N) where N is the number of individuals raised per generation. Once N is fairly large, and $p \geqslant 0.10$, the probability that any of the alleles will be absent in the offspring population is nil. Nevertheless, by sampling alone, some frequencies might drift to zero in spite of the selection pressure. For this reason the selected group should include some minimum number of individuals carrying each marker allele. For codominant and dominant alleles, it will be relatively simple to ensure that all alleles are represented at more or less equal frequencies in the selected group, even in the early generations. For recessive alleles this means including at least one individual showing each marker phenotype. The probability that at least one of N individuals will show any given recessive marker phenotype will be $1 - (1-p^2)^N$, and the probability that all marker phenotypes will appear at least once is this expression to the nth power, $[1-(1-p^2)^N]^n$. N would then be chosen so as to give a reasonably high probability that at least one such individual would be found for each marker.

In later generations, when gene frequencies are high, it might still be worthwhile to examine the selected individuals to ensure that none of the alleles are present in particularly low frequency, even if this involves some reduction in the nominal selection intensity. At later generations N would need to be large enough, so that m would not fall below some reasonably 'safe' number (say 10 or 20) from the point of view of ensuring the line against chance accidents.

Numerical Results

Formal Crosses

As a specific numerical example, consider the production of a 20-marker strain, starting with 20 single-marker parental strains. Following the expressions developed above, this would take 10 crossing cycles (20 generations) with 9 crosses in the first cycle. In order to have a reasonably high probability that all F-2 crosses would produce at least one or more 4-way homozygotes, it would be necessary to raise N = 1000 offspring per cross in the early cycles. In this case the probability that all crosses produced at least one 4-way homozygote would be 0.83, with 500 offspring per cross this probability is only 0.25. N could be reduced in later cycles. Thus, the total number of offspring per F-2 generation will vary from 9,000 in the second cycle to, say, 500 in the last cycle. Much smaller numbers, of course, will be needed in the alternate F-1 generations of the cycles.

Producing a 10 marker strain in this manner would require 10 generations, and a 30 marker strain would require 30 generations.

Random Mating and Selection

Again consider the production of a 20 marker strain, starting with 20 single-marker parental strains. It is assumed that 400 offspring are raised and 20 individuals chosen each generation, so that the proportion selected is equal to 0.05, and the standardized selection differential, i = 2.0. Figure 1 shows gene frequencies by generations on these assumptions, separately for recessive, codominant and dominant alleles. Preliminary calculations showed that increase in gene frequency for the recessive alleles was slow in the initial one or two generations when gene frequencies were low. Also at these low frequencies, it is necessary to raise large numbers of offspring in order to ensure good representation of recessive alleles in the selected population. For this reason a more optimal program in the case of recessive alleles is to devote two generations at the start of the program to the production of

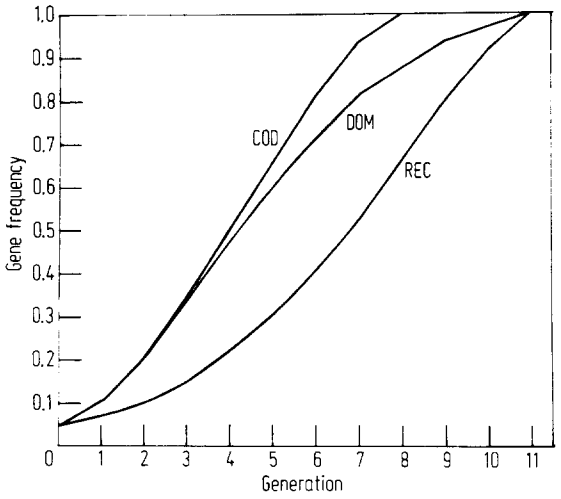

Fig. 1. Production of a 20-marker strain by random mating and selection. Gene frequencies by generations for recessive (REC), codominant (COD) and dominant (DOM) alleles

ten, 2-marker strains, and then proceed by random mating and selection from this point. In this way gene frequency in the initial generation would be 0.10. The 20-marker strain would be reached in 9 additional generations giving 11 generations in all for recessive alleles. In the case of codominant alleles, it would require a total of 8 generations to reach a 20-marker strain starting with 20 single marker parent strains. For dominant alleles, increase in gene frequency under these conditions is rapid until generation 7, but slows significantly thereafter. Also, from this point the desired selection intensity cannot be maintained, since more than 5-10% of the population show all desired marker phenotypes. From this point therefore, it would be necessary to progeny test in order to reach fixation. This could be achieved by two generations of progeny testing starting in generation 8, giving 11 generations in all for the production of a 20-marker strain for the case of dominant alleles.

In the initial stages of selection for n recessive alleles, it would be necessary to raise 500-600 total offspring to ensure that each marker appeared at least once in homozygous condition. This number could be reduced considerably as gene frequencies increase. For codominant or dominant alleles, even in the initial generations, 200 offspring would be adequate to ensure that all alleles were well represented in the selected population.

In all cases reducing selection intensities to 10% adds one generation until fixation. Also, in all cases, producing a 10-marker strain would require two generations less, and producing a 30 marker strain would require two generations more, than the number of generations required for a 20-marker strain.

Discussion

The results of this study show that production of a multi-marker strain will require a considerable investment in time and resources when based on a series of formal crosses, but can be carried out with impressive rapidity and only minimal resources by a program of random mating and selection. It is recognized that in practice the production of such strains will depend on the availability of suitable markers, from the point of view of chromosomal location, and on their biological interactions when present in the same individual. Biochemical markers extend both possibilities considerably.

Acknowledgement

This research was supported by a grant from The Leonard Wolfson Foundation for Scientific Research. The authors thank Dr. T. Brody and Dr. A. Beiliss for useful discussions.

Literature

Chai, C.K. (1975): Genes associated with leucocyte production in mice. J. Hered. 66, 301-308

Falconer, D.S. (1960). Introduction to Quantitative Genetics. The Ronald Press Co., New York

Patterson, F.L.; Schafer, J.F.; Caldwell, R.M. (1968): Effects of selected linkage blocks on yield and yield components in wheat. In: Third Intern. Wheat Genetics Symposium, Canberra, 1968 (eds.: Finlay, K.W.; Shepherd, K.W.). Sydney: Butterworth

Soller, M.; Brody, T.; Genizi, A. (1979): The expected distribution of genes determining quantitative traits next to marker loci in crosses between inbred and outbred lines. Heredity 43, 179-180

Soller, M.; Genizi, A.; Brody, T. (1976): On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. Theor. Appl. Genet. 47, 35-39

Spickett, S.G.; Thoday, J.M. (1966): Regular responses to selection: 3. Interaction between located polygenes. Genetic Res. 7, 96-121

Thoday, J.M. (1961): Location of polygenes. Nature 191, 368-370

Zhuchenko, A.A.; Samovol, A.P.; Korol, A.B.; Andryuschenko, V.K. (1979): Linkage between loci of quantitative characters and marker loci. 2. Influence of three tomato chromosomes on variability of five quantitative characters in backcross progenies. Soviet Genet. 15, 433-443 Translated from Genetika 15, 672-683

Received September 10, 1980
Communicated by R.W. Allard

J. Weller
Dr. M. Soller
Dept. of Genetics
The Hebrew University of Jerusalem
Jerusalem (Israel)