Theor Appl Genet (2001) 103:1028–1036 © Springer-Verlag 2001

A.W. Campbell · P.J. Campbell · W.B. Griffin
D.J. Burritt · A.J. Conner

Prediction of sample size to maintain genetic variation in doubled-haploid populations following marker selection

Received: 16 October 2000 / Accepted: 20 March 2001

Abstract Marker selection (MS) and doubled-haploid (DH) technologies have the potential to reduce the time taken to breed new cereal cultivars. However, a limiting factor is the potential increased genetic drift. The aim of this study was to design and test a genetic model for predicting the sample sizes needed to maintain genetic variation among DH plants following marker selection. The model estimates the amount of the genome that is fixed during the production of DH populations of a given size using a given number of markers. To test the model, doubled-haploids were produced from wheat plants selected for three PCR-based markers. When the genetic variation of the DH population (108 plants), produced from 15 selected F_2 plants homozygous at three loci, was compared to the genetic variation of an unselected F_3 population (200 plants), five of the six measured quantitative traits were identical and normally distributed. This model should prove to be a valid breeding tool, allowing a breeder to apply MS to a breeding programme and estimate the minimum DH population sizes required for minimal loss of genetic variation through genetic drift.

Communicated by J.W. Snape

A.W. Campbell (✉) · W.B. Griffin · A.J. Conner
New Zealand Institute for Crop and Food Research Limited,
Private Bag 4704, Christchurch, New Zealand

A.W. Campbell · A.J. Conner
Soil, Plant and Ecological Sciences Division, PO Box 84,
Lincoln University, Christchurch, New Zealand

P.J. Campbell
Flinders Medical Centre, Bedford Park, 5042,
South Australia, Australia

D.J. Burritt
Botany Department, University of Otago, PO Box 56,
Dunedin, New Zealand

Present address:
A.W. Campbell, Ag Research, Invermay Agricultural Centre,
Puddle Alley, Private Bag 50034, New Zealand

Keywords Marker selection · Doubled haploids · Cereal breeding · Genetic model

Introduction

Molecular markers are potentially a powerful tool in cereal breeding programmes, enabling the direct selection for linked desirable traits. The efficiency of breeding strategies used to incorporate marker selection (MS) into breeding programmes is influenced by a number of factors. Firstly, the population size used to incorporate MS into breeding programmes depends on the number of genes under selection (Howes et al. 1998) and the heritability of the traits of interest (Luo et al. 1997). Secondly, the degree of linkage between the marker and the trait of interest is important (Lande and Thompson 1990; Luo et al. 1997). If the associations among marker loci and traits of interest are not close, they are more likely to be disrupted by recombination, random genetic drift or selection events (Luo et al. 1997).

In breeding programmes, large populations need to be screened so that only homozygous progeny are selected, thereby eliminating the need for subsequent re-selection after further generations of segregation. Alternatively, smaller populations can be used if heterozygotes are selected as well, but repeated selection is required after recombination (Ribaut and Betrán 1999). The improved efficiencies in doubled-haploid (DH) production of cereals offer opportunities to improve the efficiency of MS (Howes et al 1998). The production of DH lines in cereal breeding allows the advancement of breeding lines to homozygosity in a single generation. Thus, selected genes can be fixed if a DH step is used following MS, saving repeated selection at each generation.

When integrating DH technology and MS into a breeding programme it is important to balance the economic efficiency associated with screening as few plants as possible with the biological requirement of retaining genetic variation in the selected DH population. By chance alone, all members of a very small population

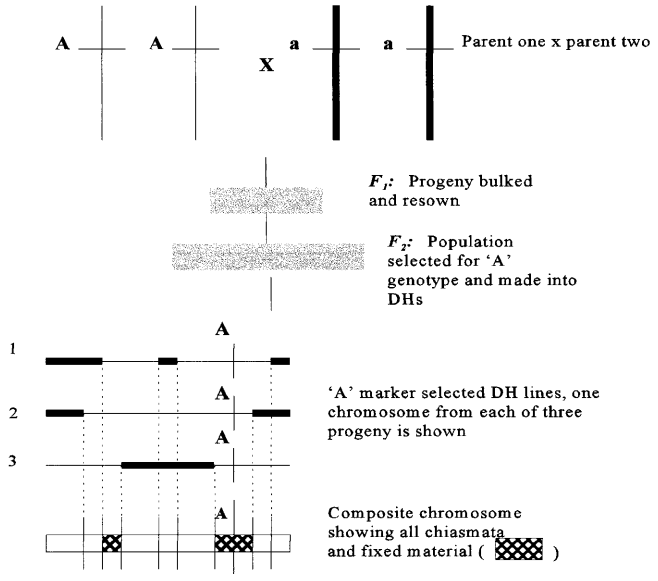


Fig. 1 Diagram of how marker-assisted selection and random genetic drift fixes proportions of a chromosome. Note: '*A*' represents a desired gene for selection from parent one and '*a*' represents a gene selected against

can have the same fixed allele for a given gene. This process is compounded by MS as any genetic material which is closely linked to a marker will have a very high probability of being inherited and fixed in a population along with the marker. Random genetic drift results, in part, from this fixation of genetic material in small populations, as well as gene linkages with restricted haplotypes.

This study presents a model for estimating the proportion of fixed genetic material in the gene pool of a population undergoing MS with DH technology, compared to that of a conventional breeding programme. Phenotypic data are presented to assess the impact of the exaggerated genetic drift due to MS on six morphological quantitative traits in wheat. The aim of this paper is to determine how screening different numbers of plants using different numbers of markers affects genetic drift and, thus, the robustness of the population for subsequent selection.

The genetic model

The model outlined estimates the proportion of genetic material that is fixed by the MS and DH processes. For example, if n progeny are produced, the proportion of genetic material which is identical in all progeny is estimated. The fixed genetic material includes the desirable selected genes and may also include genes which co-segregate with the selected gene (Fig. 1). In Fig. 1 the desirable allele A is selected to be present in all three progeny. As the progeny are DHs, each individual has two identical copies of the chromosome. The chiasmata produced during meiosis cause the cross-over of genetic material and the distribution of these across the gene pool of F_2 plants determines what fraction of the chromosome is fixed in the population. The hatched areas in the composite chromosome are portions where all three individuals inherited DNA from only one of the parents. Figure 1 also shows how genetic material can be fixed even when it is derived from chromosomes, or parts of chromosomes, which do not have selected markers. This corresponds to random genetic drift.

In designing the genetic model assumptions were made (the implications of which are discussed later):

(1) That recombination events occur randomly on the chromosomes (i.e. there are no cross-over 'hot-spots').
(2) That all markers segregate independently (no linkage).
(3) That the molecular markers used in selection are always tightly linked to the gene of interest.

The first assumption, expressed statistically, states that the chiasmata are distributed along the length of a chromosome, $[0,L]$, according to a Poisson process with rate 1, where L is the length of the chromosome in Morgans.

Let n be the number of doubled-haploid offspring produced subsequent to marker selection.

Consider a chromosome with a locus for one of the selected genes. Then, from the assumption above, the distribution of chiasmata along this chromosome is a Poisson point process with rate 1, for each individual.

We consider an imaginary 'composite' chromosome where the chiasmata for each of the n individuals are combined. It is known from the theory of point processes that the 'sum' of n independent Poisson processes on $[0,L]$ with rate 1 is a Poisson process with rate n (Cox and Isham 1990).

Let Z be the length, in Morgans, of genetic material on a given chromosome, in the gene pool of the n individuals that is fixed. Let X_i, with $0<X_1<X_2<...\leq L$, $i=1,2,...$, be the order statistics for the distribution of chiasmata along the chromosome. Let P_i be the indicator variable of fixed genetic material between chiasmata i and $i-1$, such that:

$$P_i = \begin{cases} 1, \text{ if the genetic material between chiasmata } i \text{ and } i-1 \text{ is identical for all } n; \\ 0, \text{ if the genetic material between chiasmata } i \text{ and } i-1 \text{ has both parents' genes among } n. \end{cases}$$

It follows that:

$$Z=P_0(X_1-0)+P_1(X_2-X_1)+P_2(X_3-X_2)+... .$$

Because the P_i and the X_i are independent, we know that the expected amount of genetic material that is fixed in the population is given by:

$$\mathrm{E}[Z]=\mathrm{E}[P_0]\,\mathrm{E}[X_1]+\mathrm{E}[P_1]\,\mathrm{E}[X_2-X_1] +\mathrm{E}[P_2]\,\mathrm{E}[X_3-X_2]+... . \quad (1)$$

Calculating P_i

The random variables P_i, for i=1,2,..., can be viewed as a random walk on the chromosome.

Let $p_i(j)$, i=1,2,...; j=0,1,...,n, be defined as the probability that across all n F_1 offspring between chiasmata i and i-1 there are precisely j copies of the maternal chromosome (and hence n–j copies of the paternal genetic material).

Consider the transition probability at a chiasma. Here, one (and only one) of the n F_1 offspring changes from one parent's genetic material to the other parent's. From the theory of Poisson point processes, each individual has an equal probability, 1/n, of changing. Thus, the Markov chain transition probability matrix, Q, can be written as:

$$Q=\begin{bmatrix} 0 & 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 1/n & 0 & (n-1)/n & 0 & \dots & 0 & 0 & 0 \\ 0 & 2/n & 0 & (n-2)/n & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & (n-1)/n & 0 & 1/n \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 & 0 \end{bmatrix}.$$

The entry in position (a, b) of this matrix is the probability of going from a copies of the maternal chromosome to b copies at one chiasma. From this, we know that the vector of probabilities after the ith chiasma is given by:

$$[p_i(0)\ p_i(1)\ \dots\ p_i(n)]=[p_0(0)\ p_0(1)\ \dots\ p_0(n)]Q^i,$$

where

$$[p_0(0)\ p_0(1)\ \dots\ p_0(n)]$$

is the starting distribution.
It follows that:

$$\Pr(P_i=1)=p_i(0)+p_i(n), \quad \text{and therefore,} \quad \mathrm{E}[P_i]=p_i(0)+p_i(n). \tag{2}$$

The starting distribution is simple to calculate. For chromosomes on which we select for a marker, we can consider the random walk starting at the marked gene locus. At this locus, the starting vector will be [1 0 0 ... 0], because all offspring will have the marked gene from just one parent. Of course, the random walk will have to be performed twice, once in each direction from the starting point.

For chromosomes on which we do not select for a marker, the number of copies from each parent will be distributed according to the binomial distribution with parameters n and 0.5, and the random walk can begin from oneend of the chromosome.

Calculating X_i

With a point process, the distance from one chiasma to the next is called a renewal function. With a Poisson process on an infinite line, this renewal function is the simple exponential function. However, when the Poisson process occurs on a finite line such as a chromosome the renewal function is somewhat more complicated.

Essentially, we consider the end of the chromosome to be a censoring point. If a chiasma has not occurred before L, then the process is truncated at L. Then, with X_0=0, the mixed discrete-continuous renewal function for X_1 is:

$$f_{X_1}(x_1)=\begin{cases} ne^{-nx_1}, & 0\le x_1<L; \\ \delta_L(x_1)e^{-nL}, & x_1=L; \\ 0, & x_1<0 \quad \text{or} \quad x_1>L, \end{cases}$$

where $\delta_L(x)$ is the Dirac delta function at L (Hildebrand 1962).

Because the sum of any i exponential distributions with parameter n is a gamma distribution with parameters n and i (Cox and Isham 1990) we can generalise this to:

$$f_{X_i}(x_i)=\begin{cases} \dfrac{n^i}{(i-1)!}x_i^{i-1}e^{-nx_i}, & 0\le x_i<L \\ \delta_L(x_i)\left[1-\int_0^L \dfrac{n^i}{(i-1)!}y^{i-1}e^{-ny}dy\right], & x_i=L. \end{cases}$$

It follows from this that:

$$\mathrm{E}[X_i]=\int_0^L \frac{n^i}{(i-1)!}x_i^i e^{-nx_i}dx_i+L\left[1-\int_0^L \frac{n^i}{(i-1)!}x_i^{i-1}e^{-nx_i}dx_i\right].$$

With a little manipulation, we can then find an expression for the difference $\mathrm{E}[X_i–X_{i-1}]$:

$$\mathrm{E}[X_i-X_{i-1}]=\frac{1}{n}\left(1-e^{-nL}\sum_{j=0}^{i-1}\frac{(nL)^j}{j!}\right). \tag{3}$$

Completing the model

We can now calculate the expected amount of genetic material that is fixed by MS, using equations (1), (2) and (3) above. Because the value of $\mathrm{E}[X_i–X_{i-1}]\to 0$ as $i\to\infty$, the sum quickly approaches its limit. In using the model we know that wheat has 42 chromosomes with known lengths in Morgans (Leroy et al. 1997). The output from the model, expressed as the proportion of fixed genetic material for different numbers of markers and different population sizes, is presented in Fig. 2.

Material and methods

Plant material

Two wheat cultivars ('Rata' and 'Monad') were used to test the model, based on polymorphisms at the HMW *Glu-1* locus. On the A, B and D genomes of wheat this locus potentially codes for two distinct proteins designated "x-type" and "y-type" on the basis of structural and conformation differences (Payne et al. 1981; Shewry et al. 1989). HMW *Glu-1* x/y locus configurations for 'Monad' and 'Rata' cultivars are as follows: 'Rata' 1A: null, 1B: 7/9, 1D: 3/12, 'Monad' 1A: 2*, 1B: 17/18, 1D: 5/10. 'Rata' and 'Monad' were crossed to produce F_1 seed which was collected

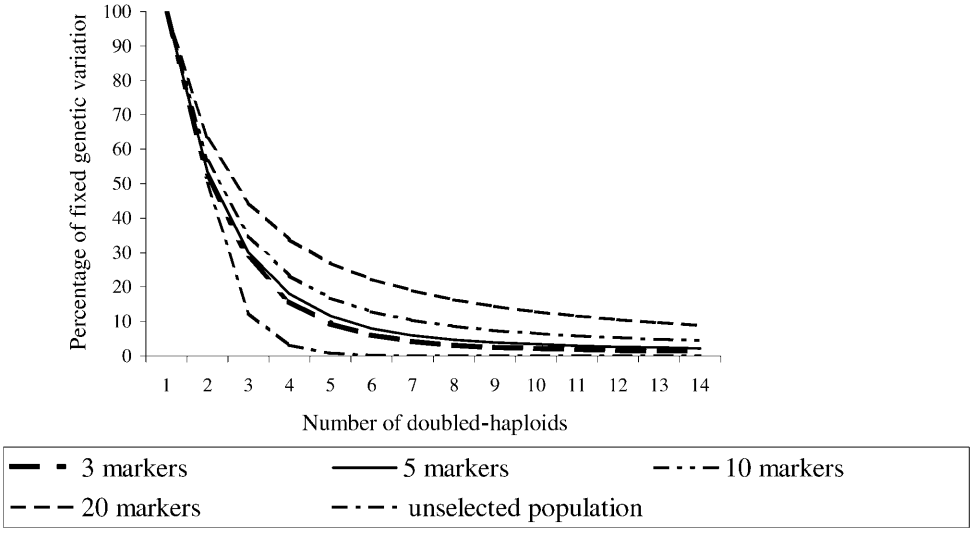


Fig. 2 The expected amount of genetic material which is fixed in doubled-haploid populations produced following marker-assisted selection

and re-sown to produce F_2 plants for DNA extraction and selection.

DNA extraction and marker selection

Genomic DNA was extracted as described by Weining and Langridge (1991) from 364 plants to ensure a minimum of five DH plants homozygous for three loci. The DNA quantity of each sample was measured using a Pharmacia GeneQuant RNA/DNA calculator and diluted to 50 ng μl^{-1}.

PCR was performed in a 25 µl reaction volume containing one unit of *Taq* DNA polymerase (Boehringher), 1×*Taq* PCR buffer, 200 µM of each deoxyribonucleotide, 10 ng each of the two primers and 50 ng of genomic template. Primers were prepared as per D'Ovidio et al. (1995). Amplification of the A and D HMW x-type glutenin genes included an initial denaturation step at 94°C for 2 min, followed by 35 cycles at 94°C for 1 min, 62°C for 2 min, 72°C for 2 min 30 s and 72°C for 7 min. Amplification of the B HMW-*Glu* x-type gene included the same initial denaturation step followed by 35 cycles at 94°C for 1 min, 58°C for 2 min, 72°C for 2 min 30 s and 72°C for 7 min.

Amplified products were analysed on a 1.5% agarose gel in TBE (0.085 M Trizma base, 0.09 M boric acid and 0.05 M EDTA) and stained with ethidium bromide (0.5 mg ml^{-1}) for half an hour before viewing under a UV light source.

Doubled-haploid production and measurement of genetic variation

DH plants from the selected 'Rata'×'Monad' F_2 plants (15 plants) were produced (3–15 DHs produced from each F_2 parent, totalling 108 DHs) as follows: wheat spikes were emasculated 1–2 days before anthesis and pollinated with maize (Yates 'Early Miracle') on the approximate day of anthesis. Spikes were sprayed with 2,4-dichlorophenoxyacetic acid (2,4-D) (100 mg l^{-1}), until run off, 1-day after pollination. Seeds were dissected 14 days after pollination and developing embryos were rescued onto 1/2 MS (Murashige and Skoog 1962) salts supplemented with 20 g l^{-1} of sucrose and 12 g l^{-1} of agar. Embryos were incubated at 20–24°C (in the dark) until germination and then transferred to light (65 µm ol m^{-2} s^{-1}, photosynthetically active radiation). When they were 20–40 mm in height, haploid plants were transplanted to soil mix and grown in a glasshouse (described in Campbell et al. 2000). Haploid plants were removed from soil mix when they had produced 3–4 tillers and were placed in aerated colchicine (500 mg l^{-1}) and a dimethy sulphoxide (DMSO) (10%) solution for 5 h, at 30°C, to double chromosome numbers (Campbell et al. 2000).

Three populations were developed and evaluated in a completely randomised design in the glasshouse:

(1) An unselected population of 200 F_3 seeds which were randomly harvested from 200 F_2 plants.
(2) A DH population which consisted of 108 DH lines (derived from 15 parental plants) homozygous for markers 'Rata' HMW *Glu1* x-type, A, B and D.
(3) Two parent ('Monad' and 'Rata') populations (15 plants per parent).

Plants were grown and maintained as described by Campbell et al. (2000). Six quantitative traits were measured in all three populations 1 week after anthesis: plant height (crown to terminal spikelet of leading tiller, not including awns), flag leaf length (base of leaf to tip of leaf of leading tiller), flag leaf width (width at base of leaf of leading tiller), number of nodes (from crown to spike of leading tiller), spike length (spike of leading tiller, not including awns) and awn length (base to tip of awn from the middle spikelet of the leading tiller).

Statistical analysis

Chi-square tests were used to confirm that the observed segregation of HMW *Glu1* x-type markers fitted the expected ratios for the F_2 population. The means of the parent populations were compared using *t*-tests. The variances of the selected DH and unselected F_3 populations were compared using *F*-tests and population means were compared using *t*-tests.

DH plants were grouped into families according to which of the 15 selected plants they were derived from. One-way ANOVA was then used to compare the different DH lines to test if the selected plant from which they were derived had any effect on the overall genetic variation in the selected population. A non-parametric bootstrapping method was also used to analyse one trait (DH height) which was not normally distributed.

Results

Marker selection

PCR products of 'Monad' and 'Rata' cultivars showed polymorphisms at all HMW-*Glu1* x-type loci and all loci segregated as expected for an F_2 population in Chi-square analysis. The scheme used to select progeny is described in Fig. 3.

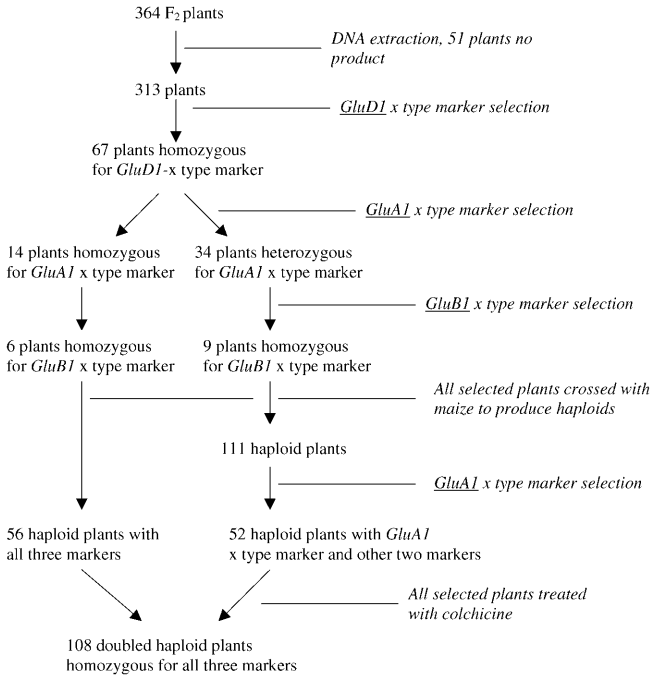


Fig. 3 Schematic representation of marker selection and doubled-haploid processes used to generate the selected doubled-haploid population

Analysis of genetic variation among populations

The two parent populations, 'Rata' and 'Monad', were compared using *t*-tests. For the six traits measured, four traits were significantly different (Table 1). Comparisons between the unselected F_3 population and the selected DH populations showed that height was the only trait to significantly differ in both *F*-tests and *t*-tests (Fig. 4). The distribution of the height trait in the selected population was bimodal. Distribution curves were normal for both selected and unselected populations for the other five traits (flag leaf length, flag leaf width, node number, spike length and awn length) (Figs. 5–9).

Table 1 Means and standard deviations of the 'Rata' and 'Monad' parent populations. Note: ns, *, *** represent non-significance and significance at 5% and 0.1% probability levels respectively

Trait	'Rata' parent population	'Monad' parent population	*t*-test *p* value
Height (cm)	56.5±5.5	91.5±10.5	<0.001 ***
Flag leaf length (cm)	25.0±5.5	32.5±6.0	0.002 ***
Flag leaf width (mm)	18.0±2.0	19.5±1.5	0.014 *
Node number	4±1	4±1	0.717 ns
Spike length	3.5±0.5	4.0±0.5	0.049 *
Awn length	93.0±9.0	89.0±5.0	0.163 ns

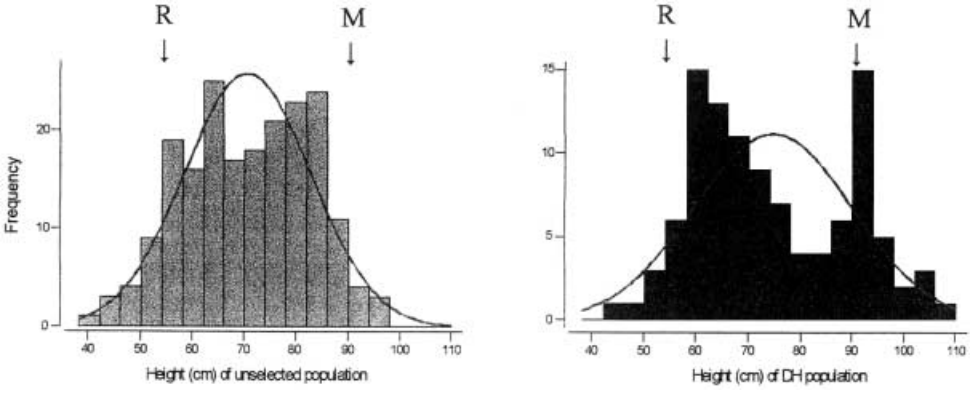
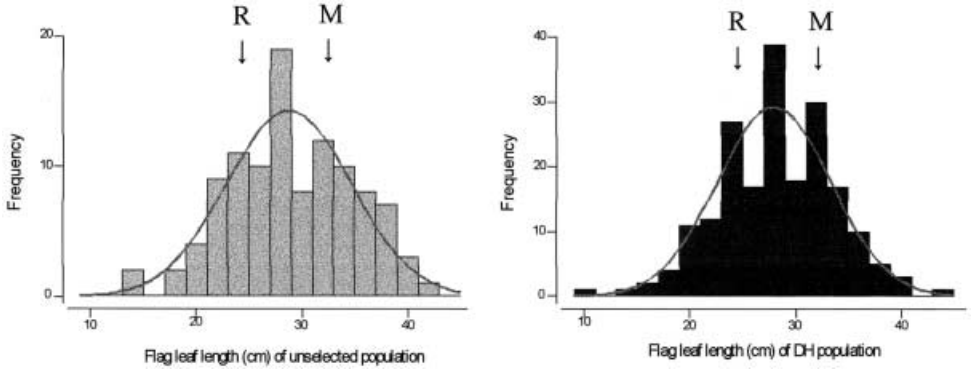


Fig. 4 The distribution of means for height in the unselected F_3 population (*left*) and the selected DH population (*right*). *R* and *M* represent the means of Rata and Monad respectively. Statistics for the unselected population: mean=70.5 cm, variance=149.5 cm, skewness=–0.12, kurtosis=–0.82. Statistics for the selected DH population: mean=74.5 cm, variance=230.5 cm, skewness=0.311, kurtosis=–0.97. Population comparisons: *F*-test, populations significantly different, *t*-test, populations significantly different ($p<0.01$). Note: both *F*-ratios and *t*-tests require normality, therefore tests may not be valid because of the bimodal distribution of the selected population

Fig. 5 The distribution of means for flag leaf length in the unselected F_3 population (*left*) and the selected DH population (*right*). *R* and *M* represent the means of Rata and Monad respectively. Statistics for the unselected population: mean=28.0 mm, variance=29.0 mm, skewness=–0.11, kurtosis=0.176. Statistics for the selected DH population: mean=29.0 mm, variance=35.0, skewness=–0.11, kurtosis=–0.33. Population comparisons: *F*-test, non-significant; *t*-test, non-significant



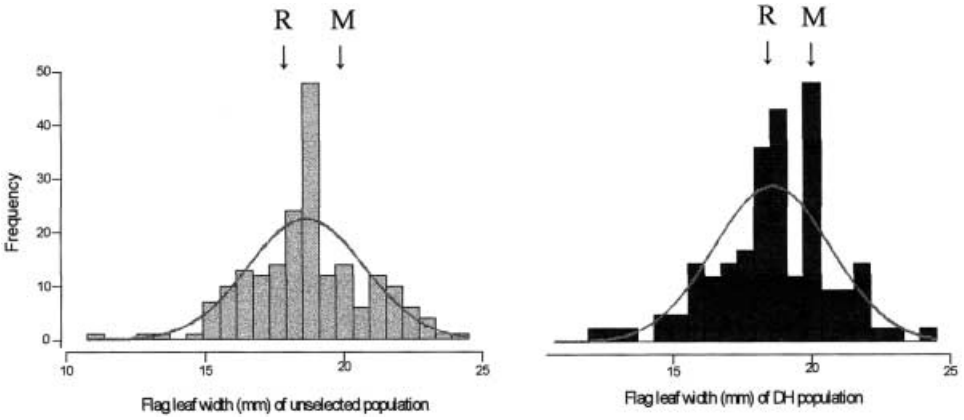


Fig. 6 The distribution of means for flag leaf width in the unselected F_3 population (*left*) and the selected DH population (*right*). *R* and *M* represent the means of Rata and Monad respectively. Statistics for the unselected population: mean= 18.5 mm, variance=4.5 mm, skewness=–0.06, kurtosis=0.56. Statistics for the selected DH population: mean=18.5 mm, variance=4.5, skewness=–0.45, kurtosis=0.77. Population comparisons: *F*-test, non-significant; *t*-test, non-significant

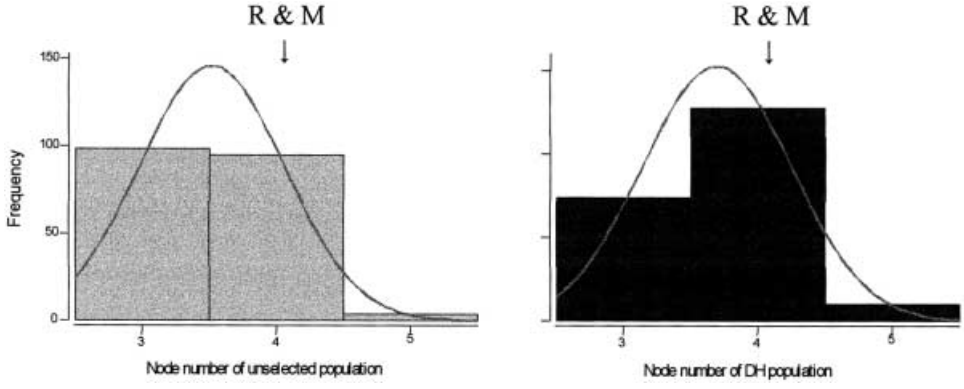


Fig. 7 The distribution of means for number of nodes in the unselected F_3 population (*left*) and the selected DH population (*right*). *R* and *M* represent the means of Rata and Monad respectively. Statistics for the unselected population: mean=4, variance=1, skewness=0.31, kurtosis=–1.12. Statistics for the selected DH population: mean=4, variance=1, skewness=–0.02, kurtosis=–0.57. Population comparisons: *F*-test, non-significant; *t*-test, non-significant

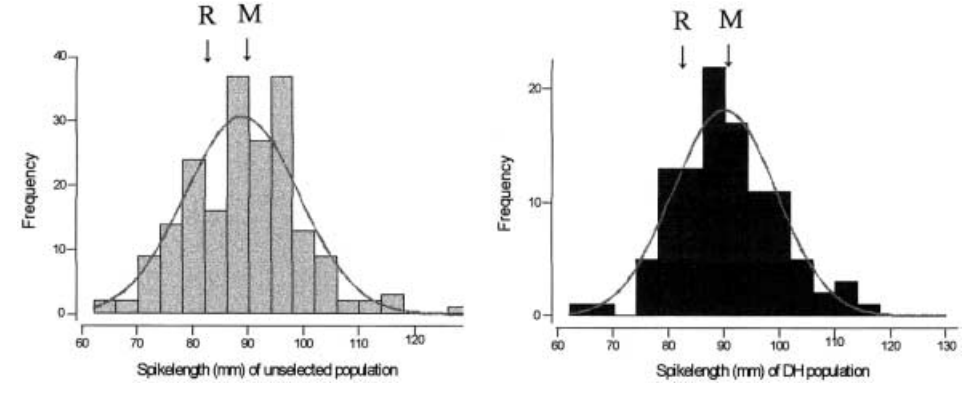


Fig. 8 The distribution of means for spike length in the unselected F_3 population (*left*) and the selected DH population (*right*). *R* and *M* represent the means of Rata and Monad respectively. Statistics for unselected the population: mean=89.0 mm, variance= 105.5 mm, skewness=0.22, kurtosis=0.61. Statistics for selected DH population: mean=90.0 mm, variance=86.5 mm, skewness= 0.31, kurtosis=0.31. Population comparisons: *F*-test, non-significant; *t*-test, non-significant

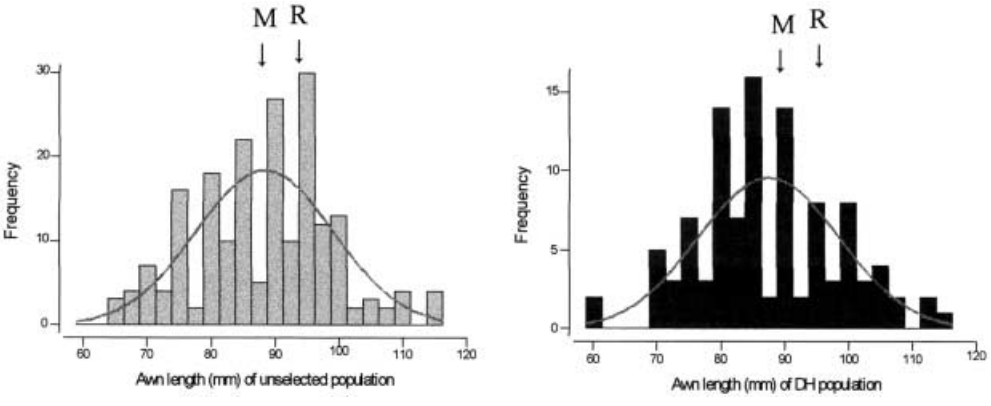


Fig. 9 The distribution of means for awn length in the unselected F_3 population (*left*) and the selected DH population (*right*). *R* and *M* represent the means of Rata and Monad respectively. Statistics for the unselected population: mean=88.5 mm, variance=115.5 mm, skewness=–0.02, kurtosis=–0.22. Statistics for the selected DH population: mean=87.5 mm, variance=121.5 mm, skewness=0.19, kurtosis=–0.14. Population comparisons: *F*-test, non-significant; *t*-test, non-significant

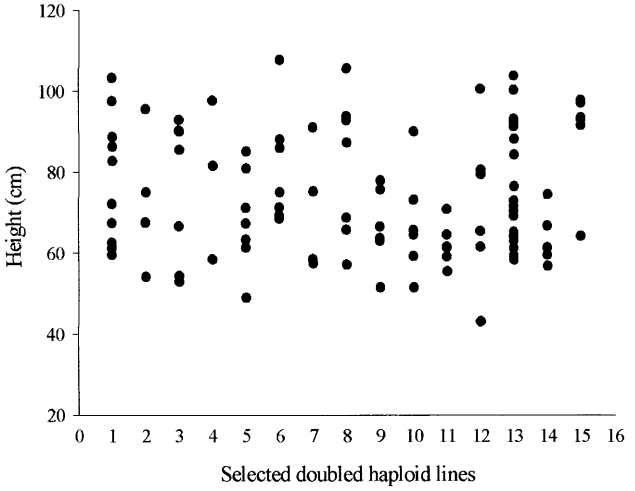


Fig. 10 Comparison of height of selected DH lines. Analysis of variance established a significant ($p<0.05$) difference in height among the DH lines. Non-parametric bootstrapping established non significance ($p>0.05$)

The groups of DH lines derived from each of the 15 selected plants were compared for all six traits (e.g. Fig. 10). Plant height and awn length were significantly different among the 15 lines on ANOVA, but not on non-parametric testing. The other four traits all showed no significant difference among the 15 lines.

Discussion

Limitations and extensions of model

The model presented here estimates the amount of genetic material fixed by MS in a DH population. Three assumptions are made in deriving this model.

The first assumption is that recombination events occur randomly and independently along the chromosome, described by a Poisson process. It is, however, well-known that there are hot-spot regions for recombination events in many eukaryotic chromosomes (Brown and Sundaresan 1991; Shiroishi et al. 1993) including wheat (Faris et al. 2000). This does not invalidate the model because we have used Morgans as the unit of chromosome distance, rather than kilobases. Thus, zones of high recombination frequency will have disproportionately high measures of length in centimorgans relative to kilobases, and the Poisson assumption will hold. For example, in wheat, estimates of kilobase pairs per centi-morgan range from 118 kb for gene-rich regions to 22,000 kb for gene-poor regions (Gill et al. 1996). Therefore, using centimorgans to express the amount of fixed genetic material as a proportion of the total genome gives only an approximate estimate of the number of actual genes fixed in the selection process.

The second assumption in the model is that all markers segregate independently (no linkage). In screening small numbers of markers this assumption is not necessarily untenable as shown with the data presented here where all markers segregate independently. However, when using large numbers of markers, linkage will inevitably occur between markers. For the breeder, two situations could exist depending on whether markers are linked in coupling or repulsion:

(1) If two markers come from the same parent and are in coupling linkage phase, it is likely that the amount of fixed genetic material will be more than for unlinked markers, because all material between the two markers will probably be fixed.
(2) If the two desirable markers come from different parents and are in repulsion linkage phase, then the number of plants the breeder needs to screen will be increased in order to produce a sufficient number of DH lines to find recombinants with both markers. For example, as the distance between the two markers decreases the chance of linking them together decreases proportionally and the required DH population size needed increases. The actual amount of fixed genetic material is not necessarily increased.

The third assumption is that the molecular markers used in selection are always tightly linked to the gene of interest. If the linkage is not tight, then phenotypic screening may need to be performed on the DH lines to ensure no intervening cross-over has occurred. In the data presented here, the actual genes are used as markers, so this assumption was valid. In the case of selection for quantitative trait locus (QTL) chromosome regions, selection for genes themselves may not be viable. A breeding scheme to reduce linkage drag when selecting for QTL regions has been proposed by Ribaut and Betrán (1999); additional markers that flank the selected regions can be used to identify genotypes that contain the least linkage drag due to recombination events occurring close to the target regions. These genotypes can then be crossed with others to increase the population size, before making doubled-haploids.

Impact of fixed genetic material on the distribution of quantitative and qualitative phenotypic traits in a DH population

Since qualitative traits are determined by only a few genes, the effect will be either minimal (if the genes are not fixed) or dramatic (if one of the genes is fixed). A major phenotypic shift in the DH lines may occur if one of the major determinant genes is close to one of the markers. The probability that this will occur rises as the proportion of fixed genetic material rises.

For QTLs, fixing contributory genes will not necessarily affect the expected sample mean, but may affect the sample variance. As a greater percentage of genes is fixed, fewer genes will contribute variation to the trait

and it will become more tightly distributed about the mean. If, for example, a number of genes contribute equally, independently and additively to a trait, the variance of a trait is directly proportional to the number of genes. Fixing 20% of genetic material will result in the variance of such a trait being 20% less. It is unlikely, however, that genes contributing to QTLs will have equal effects and this will need to be taken into account when estimating the amount of fixed genetic material, as well as other factors such as epistasis.

Phenotypic data

The data presented here compare six phenotypic traits in two populations, one an unselected mixed heterozygote/homozygote F_3 population of wheat (representative of the equivalent time stage in a breeding programme) and the other, a population of doubled-haploid plants subjected to selection using three markers. The model predicts that for three markers, 10% of the genome is fixed in a population of five DH plants and less than 1% for 100 DH plants (Fig. 2). This includes material close to the markers of interest, as well as material on other chromosomes that becomes fixed due to random variation and small population size (Fig. 2). For the unselected population, less than 1% of the genome is fixed for five DH plants and close to zero for 100 plants (Fig. 2).

Statistical tests revealed no significant difference in the means or variances between the two populations in five of the six traits (Figs. 5–9). The almost identical distributions suggest that the effect of fixed genetic material is minimal for these traits, and that even populations produced from only five DH plants using these three markers are unlikely to suffer from excessive genetic drift over the unselected F_3 cross (Fig. 10). Therefore, by selecting for three markers and producing a DH population of 108 plants, we have fixed 1% of the genome without causing any major changes in QTL distribution.

For height, the only trait to differ significantly, the doubled-haploid population had a bimodal distribution, compared to the normal distribution seen for the unselected cross (Fig. 4). The bimodality results in part from the fact that the doubled-haploid population consists entirely of homozygotes. There is a major determinant gene in the 'Rata' parent (*Rht-D1*) that influences height in a qualitative semi-dominant manner. The heterozygotes that would exist in the unselected F_3 population would have an intermediate height, resulting in the observed more-normal distribution (Fig. 4). Small populations of the DH plants could have a major phenotypic bias in such traits, a problem increased if the major determinant gene is linked to one of the markers used for selection. It is also important to consider that the traits measured in this study are unlikely to be completely independent of each other and there may be associations amongst the traits, which may affect distributions.

Conclusions

The model presented here estimates the amount of genetic material that is fixed in a DH population of a given size, selected for a given number of markers. The aim is to guide the breeder in finding the population size that optimally balances competing economic efficiency and biological constraints. The amount of genetic material that can be fixed without major impact on phenotypic traits is unclear. The preliminary data presented here suggest that up to 1% may be fixed without significantly affecting many QTLs, but oligogenic or monogenic traits may be more unpredictable. More research into this area, using saturation genetic mapping, different numbers of markers, linked markers and field trials, would be helpful in further establishing guidelines for the breeder as to the minimum acceptable population sizes. This model has been tested with wheat, but is a model that could also provide useful guidelines as to the minimum populations sizes for using marker and DH technologies in other crops.

References

Brown J, Sundaresan V (1991) A recombination hotspot in the maize *al* intragenic region. Theor Appl Genet 81:185–188

Campbell AW, Griffin WB, Burritt DJ, Conner AJ (2000) Production of wheat doubled-haploids via wide crosses in New Zealand wheat. NZ J Crop Hort Sci 28:185–194

Cox DR, Isham V (1990) Point processes. Chapman and Hall, London

D'Ovido RD, Masci S, Porceddu E (1995) Development of a set of oligonucleotide primers specific for genes at the *Glu-1* complex loci of wheat. Theor Appl Genet 91:189–194

Faris JD, Haen KM, Gill BS (2000) Saturation mapping of a gene-rich recombination hot spot region in wheat. Genetics 154: 823–835

Gill KS, Gill BS, Endo TR Boiko EV (1996) Identification and high-density mapping of gene-rich regions in chromosome group 5 of wheat. Genetics 143:1001–1012

Hildebrand FB (1962) Advanced calculus for applications. Prentice-Hall, New Jersey

Howes NK, Woods SM, Townley-Smith TF (1998) Simulations and practical problems of applying multiple marker-assisted selection and doubled-haploids to wheat breeding programmes. Euphytica 100:225–230

Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124: 743–756

Leroy P, Nègre S, Tixier MH, Perretant MR, Sourdille P, Gay G, Bernard M, Coville JL, Quètier F, Sorrells M, Marino CL, Hart G, Friebe B, Gill BS, Röder M (1997) A genetic reference map for the bread wheat genome, *Triticum aestivum* L. em. Thell. In: McGuire PE, CO Qualset (eds) Progress in genome mapping of wheat and related species. Joint Proc 5th 6th Public Wkshp Int Triticeae Mapping Initiative, Norwich, UK, and Sydney, Australia, pp 134–140

Luo ZW, Thompson R, Wooliams JA (1997) A population genetics model of marker-assisted selection. Genetics 146:1173–1183

Murashige T, Skoog F (1962) A revised medium for rapid growth and bioassays with tobacco tissue cultures. Physiol Plantarum 15:473–497

Payne PI, Corfield KG, Holt LM, Blackman JA (1981) Correlations between the inheritance of certain high-molecular-weight subunits of glutenin and bread-making quality in progenies of six crosses of bread wheat. J Sci Food Agric 32:51–60

Ribaut J-M, Betrán J (1999) Single large-scale marker-assisted selection (SLS-MAS). Mol Breed 5:531–541

Shewry PR, Halford NG, Tatham AS (1989) The high-molecular-weight subunits of wheat, barley and rye: genetics, molecular biology, chemistry and role in wheat gluten structure and functionality. Oxford Surveys Plant Mol Cell Biol 6:164–219

Shiroishi T, Sagai T, Moriwaki K (1993) Hotspots of meiotic recombination in the mouse major histocompatibility complex. Genetica 88:187–196

Weining S Langridge P (1991) Identification and mapping of polymorphisms in cereals based on the polymerase chain reaction. Theor Appl Genet 82:209–216