# University of Groningen

## On the selection for specific genes in doubled haploids

Jansen, Ritsert

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

Link to publication in University of Groningen/UMCG research database

# On the selection for specific genes in doubled haploids

R. C. JANSEN

*Centre for Plant Breeding and Reproduction Research (CPRO-DLO) P.O. Box 16, 6700 AA Wageningen, The Netherlands*

Doubling haploids from the $F_1$ of a cross between two homozygous lines and selection between the produced homozygous lines is a breeding strategy aimed at producing a population of pure lines quickly and thereby facilitating selection. Different selection goals are discussed. The problem of how to retain all desired genotypes by one or more plants in the population of doubled haploids is considered. For (combinations of) characters which are determined by a small number of unlinked (major) genes, minimum numbers of doubled haploid lines are tabulated that guarantee the breeder retaining the desired genotypes with a given probability. A general expression for this probability is given, which may be used for the case of linked genes.

**Keywords:** doubled haploids, minimum population size.

## Introduction

A plant breeder or geneticist may be interested in exploring and exploiting the genetic variation resulting from crossing two homozygous lines. A population of homozygous plants can be obtained by doubling the chromosomes of haploid gametes from $F_1$ plants, and growing diploid homozygous plants from the doubled haploid gametes (Kasha & Séguin-Swartz, 1983). Therefore, if homozygous lines are desired, early selection is facilitated by using methods of doubling haploids.

Characters or combinations of characters will be considered in this paper which are probably determined by a small number of (major) genes. It is assumed that a resistance test capable of selection for the desired genotypes in the doubled haploids is available. We consider the problem of how to retain all desired genotypes by at least one plant in the population of doubled haploids. Tables are presented, which contain the numbers of doubled haploids required to achieve this with given probability. These tables may support breeders and geneticists in balancing the costs and benefits of this and alternative breeding strategies, such as the single seed descent (SSD) method (Jansen & Jansen, 1990).

Starting from a cross between two pure lines the aim may be to select a specific genotype with a combination of favourable genes, or to fix all genetic variation in

pure lines. Three breeding aims will be considered in this paper.

(a) Selection of a particular homozygous genotype.
(b) Selection of all homozygous genotypes except those of the parents of the original cross.
(c) Selection of all homozygous genotypes, if the parental genotypes are not directly available.

The problems have been solved by applying standard probability theory (Feller, 1968).

## Probabilistic background

1. *The genetics of doubling haploids for two unlinked loci.* Firstly, the case of two unlinked loci will be considered. Extension to any number of linked or unlinked genes is straightforward. An $F_1$ between two homozygous lines AABB and aabb produces four types of gamete AB, Ab, aB and ab. Doubling the haploid gametes gives doubled haploids AABB, AAbb, aaBB and aabb. It is assumed that the gametes of which chromosomes will be doubled represent a random sample of all gametes. It follows by symmetry arguments that the expected frequencies $p_{AABB}$, $p_{AAbb}$, $p_{aaBB}$ and $p_{aabb}$ of the four doubled haploids AABB, AAbb, aaBB and aabb are equal to one-quarter. The numbers of doubled haploids $n_{AABB}$, $n_{AAbb}$, $n_{aaBB}$ and $n_{aabb}$ in the population of $N$ doubled haploids follow a multinomial

distribution. Each of the desired genotypes should be retained in the population of doubled haploids by one or more plants. The probability that this will actually happen is a function of the number of doubled haploids, $N$, and will be denoted by $\pi(N)$. The problem arises how to obtain values for $N$ which provide sufficient confidence for the desired genotypes to be retained in the population of doubled haploids. As the probability $\pi(N)$ increases with the number of doubled haploids, this problem is solved by finding the smallest value of $N$ such that $\pi(N) \geq \alpha$, where $\alpha$ is a (minimum) probability of success specified by the breeder. Once an expression for $\pi(N)$ is derived, the inequality $\pi(N) \geq \alpha$ can be solved by simple techniques.

The probabilities of retaining one new specific genotype, the two new genotypes or all four genotypes are denoted by $\pi_1(N)$, $\pi_2(N)$ and $\pi_3(N)$, respectively. Expressions for the probabilities can be obtained from Jansen & Jansen (1990) by setting the number of generations of SSD to infinity. The probabilities are given by

$$\pi_1(N) = 1 - \left(\frac{3}{4}\right)^N, \tag{1}$$

$$\pi_2(N) = 1 - 2\left(\frac{3}{4}\right)^N + \left(\frac{2}{4}\right)^N, \text{ and} \tag{2}$$

$$\pi_3(N) = 1 - 4\left(\frac{3}{4}\right)^N + 6\left(\frac{2}{4}\right)^N - 4\left(\frac{1}{4}\right)^N. \tag{3}$$

*2. The consequences of linkage in the case of two loci.*
Linkage affects the gamete frequencies. In these cases the genotypes must be grouped into the two categories (1) AABB, aabb and (2) AAbb, aaBB in order to distinguish between coupling and repulsion phase. The recombination frequency between the two loci will be denoted by $r$. If genes are in coupling phase (AB/ab) doubled haploids AABB, AAbb, aaBB and aabb are produced with frequency $p_{AABB} = (1-r)/2$, $p_{AAbb} = r/2$, $p_{aaBB} = r/2$ and $p_{aabb} = (1-r)/2$, e.g. the probability that AABB is not extinct is

$$\pi_1(N) = 1 - \left(1 - \frac{1-r}{2}\right)^N.$$

If genes are in repulsion phase (Ab/aB) the same doubled haploids are produced with frequency $p_{AABB} = r/2$, $p_{AAbb} = (1-r)/2$, $p_{aaBB} = (1-r)/2$ and $p_{aabb} = r/2$. In this case

$$\pi_1(N) = 1 - \left(1 - \frac{r}{2}\right)^N.$$

Similarly, expressions for the probabilities $\pi_2(N)$ and $\pi_3(N)$ can be obtained. However, we confine ourselves to presenting a general formula in the next section.

*3. General expressions.* In the general case of $K$ loci the number of different homozygous genotypes equals $2^K$. As in the preceding sections it is possible to calculate the expected gamete frequencies and derive formulae for the probabilities $\pi_1(N)$, $\pi_2(N)$ and $\pi_3(N)$. However, a general expression is given by Feller (1968, p. 99) for occupancy problems. Write $p_1$, $p_2$ ... $p_L$ for the expected frequencies of the $L$ desired genotypes. The probability $\pi(N)$ that all desired genotypes are retained is

$$\pi(N) = 1 - \sum_i (1-p_i)^N + \sum_{i<j} (1-p_i-p_j)^N$$
$$- \sum_{i<j<k} (1-p_i-p_j-p_k)^N$$
$$+ \ldots \pm (1-p_1-p_2- \ldots -p_L)^N. \tag{4}$$

The expressions (1), (2) and (3) for unlinked loci can be generalized by using (4) and $p_i = (\frac{1}{2})^K$ for each of the $2^K$ possible genotypes to expressions (1'), (2') and (3')

$$\pi_1(N) = 1 - \left[1 - \left(\frac{1}{2}\right)^K\right]^N, \tag{1'}$$

$$\pi_2(N) = \sum_{i=0}^{2^K-2} (-1)^i \binom{2^K-2}{i}\left[1 - i\left(\frac{1}{2}\right)^K\right]^N, \text{ and} \tag{2'}$$

$$\pi_3(N) = \sum_{i=0}^{2^K} (-1)^i \binom{2^K}{i}\left[1 - i\left(\frac{1}{2}\right)^K\right]^N. \tag{3'}$$

**Table 1** Minimum population sizes in the case of selection of a specific homozygous genotype for unlinked loci. Example: a character is determined by two ($\ell$) unlinked loci. By producing at least 11 ($N$) doubled haploids it can be guaranteed with probability 0.95 ($\alpha$) that a specific homozygote, e.g. AABB, is retained by at least one plant

| | $\alpha$ | | |
|---|---|---|---|
| $\ell$ | 0.90 | 0.95 | 0.99 |
| 1 | 4 | 5 | 7 |
| 2 | 9 | 11 | 17 |
| 3 | 18 | 23 | 35 |
| 4 | 36 | 47 | 72 |
| 5 | 73 | 95 | 146 |

$\alpha$ = Probability level.
$\ell$ = Number of loci.
$N$ = Number of doubled haploid lines.

**Table 2** Minimum population sizes in the case of selection of all homozygous genotypes except for parents, for unlinked loci. Example: a character is determined by two ($\ell$) unlinked loci. By producing at least 13 ($N$) doubled haploids it can be guaranteed with probability 0.95 ($\alpha$) that each new homozygote is retained by at least one plant

| $\ell$ | $\alpha$ | | |
|---|---|---|---|
| | 0.90 | 0.95 | 0.99 |
| 2 | 11 | 13 | 19 |
| 3 | 31 | 36 | 48 |
| 4 | 77 | 88 | 113 |
| 5 | 179 | 201 | 253 |

**Table 3** Minimum population sizes in the case of selection of all homozygous genotypes, for unlinked loci. Example: a character is determined by two ($\ell$) unlinked loci. By producing at least 16 ($N$) doubled haploids it can be guaranteed with probability 0.95 ($\alpha$) that all homozygotes are retained by at least one plant

| $\ell$ | $\alpha$ | | |
|---|---|---|---|
| | 0.90 | 0.95 | 0.99 |
| 1 | 5 | 6 | 8 |
| 2 | 13 | 16 | 21 |
| 3 | 33 | 38 | 51 |
| 4 | 79 | 90 | 115 |
| 5 | 181 | 203 | 255 |

## Results

Tables with a list of minimum numbers of doubled haploid lines are presented for a maximum of five unlinked loci and for $\alpha = 0.90$, 0.95, and 0.99 ($\alpha$ = probability of successfully retaining desired genotypes). Table 1 provides information about the case of unlinked loci in which the breeder wants to retain one specific homozygous genotype, Table 2 about the case in which the breeder wants to retain all homozygous genotypes except the parental genotypes, and Table 3 about the case in which the breeder wants to retain all homozygous genotypes. Table 4 provides information for the case of two linked loci, in which the breeder wants to retain one specific homozygote. This table is added to show the effect of linkage.

## Example

Mapped genes from different sources can be combined efficiently by using indirect selection via linked

**Table 4** Minimum population sizes in the case of selection of a specific homozygous genotype, for two linked loci. Example: a character is determined by two linked loci in repulsion phase (Ab/aB) with recombination frequency 0.1 ($r$). By producing at least 59 ($N$) doubled haploids it can be guaranteed with probability 0.95 ($\alpha$) that AABB is retained by at least one plant

| Phase | Coupling | | Repulsion | |
|---|---|---|---|---|
| $r$ | 0.1 | 0.4 | 0.1 | 0.4 |
| $N$ | 6 | 9 | 59 | 14 |

$r$ = recombination frequency.

markers. We consider as an example the case of two quantitative trait loci (QTLs) on different chromosomes using for each QTL one marker at a short distance (recombination frequency between QTL and marker equals 0.1). Doubled haploids are produced from an $F_1$ $M_1Q_1m_2q_2/m_1q_1M_2Q_2$ between parental genotypes $M_1Q_1m_2q_2/M_1Q_1m_2q_2$ and $m_1q_1M_2Q_2/m_1q_1M_2Q_2$. The expected frequencies of gametes $M_1Q_1M_2Q_2$ and $m_1q_1m_2q_2$ equal $(\frac{1}{2}0.9)^2$. Application of expression (4) gives

$$P(n_{M_1Q_1M_2Q_2/M_1Q_1M_2Q_2} \geq 1) = 1 - \left[1 - \left(\frac{1}{2}0.9\right)^2\right]^N,$$

and

$$P(n_{M_1Q_1M_2Q_2/M_1Q_1M_2Q_2} \geq 1; n_{m_1q_1m_2q_2/m_1q_1m_2q_2} \geq 1)$$
$$= 1 - 2\left[1 - \left(\frac{1}{2}0.9\right)^2\right]^N + \left[1 - 2\left(\frac{1}{2}0.9\right)^2\right]^N.$$

A minimum of 14 doubled haploids suffices to retain $M_1Q_1M_2Q_2/M_1Q_1M_2Q_2$, a minimum of 17 doubled haploids suffices to retain the two new QTL genotypes, in both cases with probability equal to 0.95.

## Discussion

The results of this paper are intended to be support tools for breeders and geneticists in decision making on a doubled haploids programme. The tables list minimum numbers of doubled haploids which are sufficient to retain (with specified probability) all desired genotypes. However, raising a large number of doubled haploids may be time consuming and labour intensive. By means of the given tables the breeder can choose a procedure by considering the costs and benefits of using doubled haploids or alternative breeding strategies. One such alternative breeding strategy is the SSD

method (Jansen & Jansen, 1990). Inbreeding a large number of plants for several generations will probably take a greenhouse compartment for several seasons. A population of doubled haploid lines is nearly equivalent to a late-generation of an SSD programme. In the case of unlinked genes, mimimum numbers of doubled haploids are only slightly smaller than minimum numbers of SSD lines in the seventh to eleventh generation (Jansen & Jansen, 1990). However, in case of linkage large differences between minimum population sizes may exist due to linkage of favourable and non-favourable alleles. Doubled haploids are produced directly from the $F_1$ and therefore recombination occurs only once. Conversely, recombination may occur in each generation of an SSD programme. The consequences of this may be seen in the following example. In the simple case of two favourable alleles in repulsion phase (Ab/aB) with recombination frequency equal to 0.1, the minimum number of doubled haploids equals 59, whereas the minimum number of SSD lines equals 36 in the eleventh generation (Jansen & Jansen, 1990).

## Acknowledgements

## References

FELLER, W. 1968. *An Introduction to Probability Theory and its Applications*, Vol. 1. Wiley, New York.

JANSEN, R. C. AND JANSEN, J. 1990. On the selection for specific genes by Single seed descent. *Euphytica*, **51**, 131–140.

KASHA, K. J. AND SÉGUIN-SWARTZ, G. 1983. Haploidy in crop improvement. In: Swaminathan, M. S., Gupta, P. K. and Sinha, U. (eds) *Cytogenetics of Crop Plants*. MacMillan, Delhi.