

A model for population growth of laboratory animals subjected to marker-assisted introgression: how many animals do we need?

O. DELPHIN KOUDANDÉ^{†‡}, PETER C. THOMSON[§] & JOHAN A.M. VAN ARENDONK^{*†}

[†]Animal Breeding and Genetics group, Wageningen Institute of Animal Sciences, Wageningen Agricultural University, PO Box 338, 6700 AH Wageningen, The Netherlands, [‡]Institut National des Recherches Agricoles du Bénin, B.P. 884, Cotonou, République du Bénin and [§]Biometry Unit, Department of Crop Sciences, University of Sydney, NSW, 2006, Australia.

This study provides methods for calculating the mean and variance of the number of animals with the desired genotype in each backcross generation for a marker-assisted introgression experiment. The ultimate goal is to produce animals which are homozygous for the desired loci. The methods have been developed specifically for experiments with inbred lines. The model assumes a Poisson distribution for litter size, and is similar to that used in stochastic versions of population dynamics models. Certain biological parameters must be specified as well as parameters under the control of the breeder. These methods can be utilized in designing an experiment to determine the number of founder animals required, given the number of animals required at the completion of the backcross process and vice versa. Consideration is given to minimizing the total amount of genotyping over the entire experiment, by varying the number of times each backcrossed male is used. In addition, an outline is given for an adaptive design that allows for changes in male usage to be made during the experiment.

Keywords: experimental design, inbred lines, MAI, Poisson distribution, population growth.

Introduction

Trypanosomiasis transmitted by the tsetse fly occurs in 37 countries in Africa, and there is evidence of genetic resistance to trypanosomiasis in some cattle breeds (Murray & Trail, 1984). A programme has commenced at the International Livestock Research Institute in Nairobi, Kenya, with the aim of identifying genes for trypanotolerance in N'Dama cattle (Teale, 1993). The intention is to use marker-assisted introgression (MAI) to develop trypanotolerant cattle types. Trypanoresistance is well known in some laboratory mouse strains (Morrison *et al.*, 1978), and three autosomal chromosomal regions have been identified (Kemp *et al.*, 1997). Consequently, mice are being used as an animal model in a pilot study to assess the effectiveness of MAI of trypanotolerance genes for cattle.

In the design stage of the experiment, a literature survey was undertaken to identify how many founder donor animals would be needed for the experiment. Although there is a considerable volume of material dealing with MAI, relatively little deals with experimental design. Gama *et al.* (1992) studied introgression strategies of transgenes in pigs and considered the effect of various strategies (number of male and female donor animals) on the number to be genotyped to detect some deleterious transgene effect. Groen & Smith (1995) conducted a stochastic simulation study of the efficiency of marker-assisted introgression in livestock, and investigated the number of animals with the desired genotype. Several studies have reported methods for reducing the number of backcross generations required for the introgression process (e.g. Hospital *et al.*, 1992; Hillel *et al.*, 1993; Yancovich *et al.*, 1996). Hillel *et al.* (1990), using genomic selection based on DNA fingerprints for gene introgression in a breeding

*Correspondence. E-mail: johan.vanarendonk@alg.vf.wau.nl

programme, investigated the theoretical distribution of the proportion of the donor's genome in successive backcross generations. Visscher *et al.* (1996) focussed more on the proportion of the recipient genotypes (background genotypes) existing after specified numbers of backcrossing generations. They studied the simultaneous effect of introgressing an allele and selecting for a desired genomic background. They also looked at a selection index combining markers and phenotypic information. However, none of these studies provides strategies for determining the number of animals required for the experiment. The determination of this number is quite important to plan properly the logistics necessary for such an experiment and to estimate its cost.

Consequently, the present paper intends to provide a mathematical model for estimating the mean and variance of the number of animals at each backcross generation resulting in the targeted number of favourable animals, i.e. those with the desired chromosomal region(s), at the end of the backcross process. The ultimate goal is to intercross those animals to produce sufficient animals that are homozygous for the desired loci in one breeding cycle if possible. Opportunities might exist for the breeder to repeat the intercross process from heterozygous individuals to obtain the required number of animals.

The objective of the paper is to predict the amount of genotyping required during a backcross experiment in order to produce a specified number of individuals with desired genotypes. In addition, the consequences of reducing the risk of not obtaining enough animals of the desired genotype to a set level are studied. Aspects considered here are: the number of founder animals; the number of times each male is used with the objective of minimizing the genotyping workload; the estimated number of animals to be genotyped for the whole experiment and at each generation, as well as the expected number of animals with a favourable genotype. The effect of the number of backcross generations, the number and the size of the regions to introgress as well as the influence of risk limitation will be studied.

Theory

Model of population growth

Some assumptions have been made: the donor and recipient lines used in the experiment are fully inbred; the introgressed alleles or regions are unambiguously identifiable in founder donor as well as in crossbred animals, and they lie on autosomal

chromosomes. Each of a fixed number of males and females selected from the donor line is mated only once with an animal from the recipient line to produce F_1 offspring. The first backcross generation is produced by mating F_1 animals to animals of the recipient line. Selection of animals for the production of subsequent backcross generations is based solely on the presence of all the entire desired chromosomal regions from the donor strain. Animals which did not inherit all regions or animals carrying one or more recombinations in any of the regions are not selected. For the development of the model, it is also assumed that all the favourable animals will be used in subsequent production of the next generation; in practice this does not need to be the case though, depending on the number of animals available. Because of differences in reproductive capacities, favourable females are mated with only one male in the recipient line, whereas selected males can be mated to more than one recipient line female. The number of times that males are used will be treated as a design parameter. Attention will also be paid to the situation in which only males are genotyped and selected during the backcross generations. It will also be assumed that there are always sufficient recipient animals available for mating.

A further assumption is that litter size varies stochastically, and independently, from female to female. As a working approximation, a Poisson distribution for litter size is assumed (see, e.g. Foulley *et al.*, 1987; Matos *et al.*, 1997), the major consequence being that the variance of litter size is taken as being equal to the mean.

In developing the theory, the mean and variance of the number of animals with the desired genotype at each generation are provided. Derivation of the full probability distribution of the number of animals at each generation is a more intractable problem. These issues are considered in mathematical theories of population dynamics (e.g. Renshaw, 1991). However, it is believed that use of summary statistics will characterize the population sufficiently well, and also allows a relatively simple development of the theory.

The following notation will be used here:

t = generation number ($0 = F_1$; 1, 2, 3, ... = backcross generation 1, 2, 3, ...);

M_t = number of favourable males at backcross generation t ;

F_t = number of favourable females at backcross generation t ;

$N_t = M_t + F_t$, the total number of favourable animals at generation t ;
 n = number of males and number of females initially selected from each strain to start the introgression process (i.e. $2n$ donor animals and $2n$ recipient animals);
 p = probability of inheriting the desired chromosomal region(s), i.e. favourable;
 r = number of times each favourable male is used for breeding;
 s = sex ratio in offspring at breeding age;
 λ = mean number of animals produced in a litter and surviving to breeding age.

Although it would usually be assumed that the sex ratio s would equal $\frac{1}{2}$, other values might occur as a result of differential mortality of males and females up to the time of breeding. Note that the probability of inheriting the desired chromosomal region(s) is a function of the number and the size of the region(s) to be introgressed. Let d_i be the length of a region (Morgans) and z the number of independent regions. With unequal chromosomal region lengths and applying Haldane's (1919) Poisson model for crossing-over, p is calculated as $p = (\frac{1}{2})^z \prod_{i=1}^z e^{-d_i}$ which reduces to $p = (\frac{1}{2})^z e^{-zd}$ when all regions are of equal length.

Assuming that the parameters r , p , s and λ are constant over generations, it may also be shown that there are explicit solutions for the mean and variance of N_t , namely $E(N_t) = 2n\lambda [p\lambda(rs + 1 - s)]^t$ and $\text{var}(N_t) = E(N_t)[\beta + \gamma E(N_t)]$ where

$$\beta = 1 - \frac{p\lambda(r^2s + 1 - s)}{(rs + 1 - s)[p\lambda(rs + 1 - s) - 1]}$$

and

$$\gamma = \frac{p(r^2s + 1 - s)}{2n(rs + 1 - s)[p\lambda(rs + 1 - s) - 1]}$$

As an illustration (Fig. 1), 20 simulations were performed, each starting from $2n = 36$ donor animals with $r = 4$; $\lambda = 5$; $s = \frac{1}{2}$; $p = 0.0686$ (based on $z = 3$ and $d = 20$ cM). This also shows the means ± 2 standard deviations, $E(N_t) \pm 2[\text{var}(N_t)]^{1/2}$. As expected, relatively few simulated values lie outside these limits.

Constant breeding strategy

Two parameters (n and r) are under control of the breeder. When specifying that a certain mean number (say N_{mean}) of favourable animals are to be available after k backcross generations and for a specified value of r , the initial number of animals

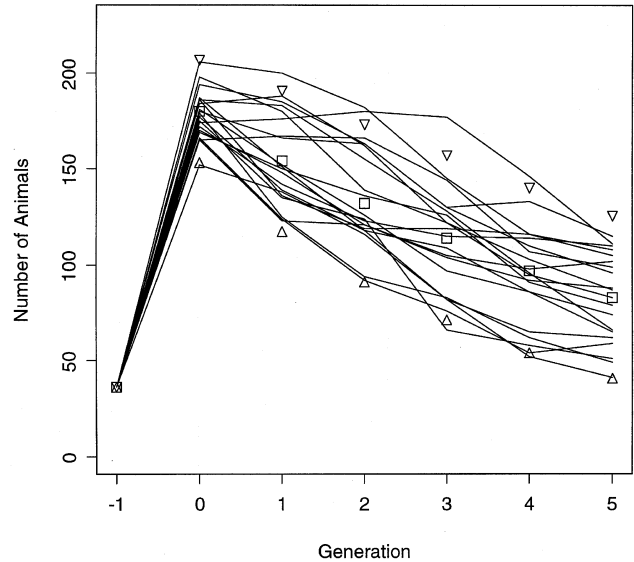


Fig. 1 Simulation of the number of favourable animals at each generation, N_t . Twenty simulations are shown all starting with $2n = 36$ founder animals; other parameters are $r = 4$; $\lambda = 5$; $s = \frac{1}{2}$; $p = 0.0686$ (based on $z = 3$ and $d = 20$ cM). Also shown are the theoretical mean (squares) ± 2 standard deviations (triangles). Generation ‘-1’ refers to the founders, ‘0’ to the F_1 generation, and subsequent generations are backcrosses.

required is clearly $n = N_{\text{mean}} / \{2\lambda [p\lambda(rs + 1 - s)]^k\}$. For this strategy, however, there will be (approximately) a 50% risk of obtaining fewer animals than this at the end of the backcross process. An alternative strategy would be to specify the minimum number of favourable animals, N_{min} , and the nominated risk of obtaining fewer animals than this, α .

Provided that N_{min} is not too small, the distribution of the number of favourable animals at generation k can be taken as approximately normal, with mean $\mu = E(N_k)$ and variance $\sigma^2 = \text{var}(N_k)$. So for a chosen value of r , we select the value of n so that $P(N_k < N_{\text{min}}) \leq \alpha$. This may be approximated as the value n satisfying

$$\frac{N_{\text{min}} - E(N_k)}{\sqrt{\text{var}(N_k)}} \leq \Phi^{-1}(\alpha),$$

where $\Phi^{-1}(\alpha)$ is the 100 α percentage point of the standard normal distribution. However, when N_{min} is small, a better approximation to normality is achieved by using a square-root transformation. These means and variances may be approximated as $E(\sqrt{N_k}) \approx \sqrt{E(N_k)} + \frac{1}{4}\text{var}(N_k)[E(N_k)]^{-3/2}$ and $\text{var}(\sqrt{N_k}) \approx \frac{1}{4}\text{var}(N_k)/E(N_k)$, and n then chosen to satisfy

$$\frac{\sqrt{N_{\min}} - E(\sqrt{N_k})}{\sqrt{\text{var}(\sqrt{N_k})}} \leq \Phi^{-1}(\alpha).$$

Note that different values of male usage, r , result in different minimum values of n . One possible option is to choose the value of r that would require the smallest expected number of animals to be genotyped during the entire experiment (amount of genotyping). Because for any backcross generation, the amount of genotyping required has a mean of $p^{-1}E(N_t)$ if males and females are genotyped, the total expected amount of genotyping over k backcross generations would be

$$EG_k = p^{-1} \sum_{t=1}^k E(N_t) \\ = 2n\lambda^2(rs + 1 - s) \frac{[p\lambda(rs + 1 - s)]^k - 1}{p\lambda(rs + 1 - s) - 1}.$$

Instead of genotyping males and females, one could only genotype and select males from generation 1 to $k - 1$, and subsequently males and females in generation k , which results in a total expected amount of genotyping of:

$$EG_k = p^{-1} \left\{ E(N_k) + \sum_{t=1}^{k-1} E(M_t) \right\} \\ = 2n\lambda^2sr \left[(\lambda spr)^{k-1} + s \frac{[(\lambda spr)^{k-1} - 1]}{\lambda spr - 1} \right].$$

Alternatively, a constant amount of genotyping each generation may be the preferred option. In this situation, we need the growth rate to be (close to) one; this is achieved by choosing r to be the nearest integer to $(p\lambda s)^{-1} - s^{-1} + 1$, that is $2/(p\lambda) - 1$ assuming $s = \frac{1}{2}$.

Adaptive breeding strategies

Another issue is that the breeding strategy may be adaptive in the sense that it might be changed between generation 0 and k , or perhaps extended for additional generations. Although n cannot be altered, r may be altered during the programme, particularly when N_t at generation t becomes critically small. For this, we need to calculate the conditional mean and variance of N_k given the actual number of males and females at generation t . These may be calculated as

$$E(N_k | M_t, F_t) = p\lambda(rM_t + F_t)[p\lambda(rs + 1 - s)]^{k-t-1}$$

and

$$\text{var}(N_k | M_t, F_t) = E(N_k | M_t, F_t)[\beta + \gamma E(N_k | M_t, F_t)],$$

where β and γ are as defined above. So at each intervening generation, $t = 0, \dots, k - 1$, we may select the value of r so that

$$\frac{N_{\min} - E(N_k | M_t, F_t)}{\sqrt{\text{var}(N_k | M_t, F_t)}} \leq \Phi^{-1}(\alpha).$$

Another adaptation to face situations in which the number of favourable animals has dropped below the desired level is to produce more offspring than planned in subsequent generations.

Numerical application

In this section, consequences of the theory outlined previously will be investigated numerically. For all calculations, regions to be introgressed are assumed to be of equal size. The sex ratio (s) was assumed to be $\frac{1}{2}$ and the mean litter size (λ) assumed to be 5. First the situation in which both males and females are genotyped will be considered. Subsequently, results for situations in which only males are genotyped during the backcrossing process are presented.

Genotyping males and females

In the first investigation, the mean number of favourable animals was specified ($N_{\text{mean}} = 50$), and the initial number of donor animals (n males and n females) determined (Table 1). Also shown is the expected number of animals to be genotyped for the whole experiment (EG_k). The probability p of inheriting the desired genotype decreases when z or d increases. Note that in some situations, values of n less than one result in the required number of animals in the target generation. In these situations, entries for higher values of r have been left empty, as higher usage of males is clearly not necessary. Both n and the number of animals to be genotyped for the whole experiment (EG_k) decrease with increasing r and p , but increase when the number of backcrosses k becomes higher. Figure 2 shows the pattern of $E(N_t)$ over the backcross generations which converges to a fixed value of $N_{\text{mean}} = 50$. Note that a plot of the expected amount of genotyping over time would show the same pattern, converging to 729 animals after $k = 5$ backcross generations.

For the second investigation, the minimum number of animals after k backcross generations was specified ($N_{\min} = 50$), with a risk of $\alpha = 0.05$ of obtaining fewer than this number (Table 2). This results in an increase of all corresponding values of EG_k and n compared with the values in Table 1 where the mean was set at 50; similarly n decreases with increasing p , increasing r and for greater number of backcross generations. The expected

Table 1 Expected number of animals to be genotyped (EG_k) for breeding scheme (r), the expected number (N_{mean}) of favourable animals required at the end of backcrosses being 50. (p = probability of inheriting the desired chromosomal region(s))

Backcrosses	No. of regions	Size of region (cM)	p	r				
				1	2	3	4	5
5	1	0	0.5	165 (0.051)	—	—	—	—
		10	0.4524	195 (0.085)	—	—	—	—
		20	0.4094	232 (0.139)	—	—	—	—
	2	0	0.25	672 (1.63)	410 (0.215)	—	—	—
		10	0.2047	1166 (4.45)	618 (0.586)	—	—	—
		20	0.1676	2191 (12.09)	994 (1.59)	684 (0.378)	—	—
	3	0	0.125	6324 (52.42)	2285 (6.90)	1345 (1.63)	992 (0.536)	—
		10	0.0926	21415 (234.99)	6370 (30.94)	3167 (7.34)	2058 (2.40)	1555 (0.967)
		20	0.0686	79770 (1053.17)	20652 (138.68)	8889 (32.91)	5074 (10.78)	3445 (4.33)
10	1	0	0.5	166 (0.000)	—	—	—	—
		10	0.4524	198 (0.001)	—	—	—	—
		20	0.4094	239 (0.003)	—	—	—	—
	2	0	0.25	893 (0.536)	—	—	—	—
		10	0.2047	2205 (3.96)	691 (0.068)	—	—	—
		20	0.1676	7493 (29.27)	1311 (0.507)	—	—	—
	3	0	0.125	72634 (549.75)	5440 (9.53)	1785 (0.536)	—	—
		10	0.0926	1027920 (11044.8)	45793 (191.53)	7819 (10.78)	3049 (1.15)	1856 (0.187)
		20	0.0686	16881900 (221834)	593469 (3846.72)	67399 (216.63)	16018 (23.26)	6430 (3.75)

The numbers in parentheses are n , i.e. n donor males and n donor females are mated to n recipient females and n recipient males to produce the F_1 generation.

amount of genotyping (EG_k) behaves differently however; for smaller values of p , EG_k initially decreases with increasing r , reaches a minimum value and then starts increasing (Table 2). However,

the initial decrease is not observed when p exceeds 0.216 (for the given values of λ, s).

Figure 3 shows the behaviour of $E(N_i)$ over the backcross generations. The mean population trajectories do not converge to the same value for all values of r ; Table 3 shows the means and variances of N_i at the fifth generation for a given N_{mean} or N_{min} . With high r -values, the initial number of founders ($2n$) is relatively small, but the final mean ($E(N_k)$) tends to be higher, compared with the designs with a lower r -value. When the number of founders is small, there is relatively more variability in the number of favourable animals in F_1 and subsequent backcross generations. Consequently, the mean in the target backcross generation, $E(N_k)$, needs to be relatively high in order to ‘ensure’ that the minimum number of favourable animals is achieved.

As expected in the third investigation (Table 4), increasing the number of animals required results in an increased amount of genotyping, as does reducing the risk level. Further, doubling the number of animals required doubles the amount of genotyping when N_{mean} is specified. However, when the risk is considered (i.e. for $\alpha < 0.50$), doubling N_{min} leads to a less than doubling of the amount of genotyping.

Genotyping males only

Males in general have a higher reproductive rate than females and consequently it might be interest-

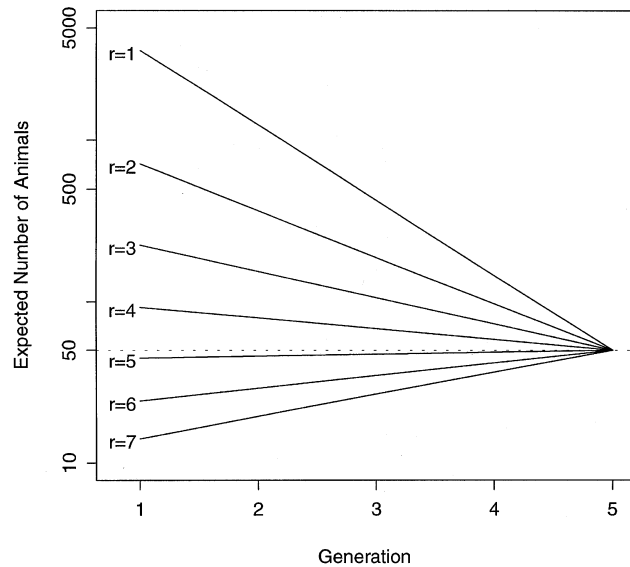


Fig. 2 Expected number of favourable animals ($E(N_i)$) at each generation of backcrossing for different breeding schemes (r). Fixed parameters were $N_{\text{mean}} = 50$ at $k = 5$ backcross generations; $\lambda = 5$; $s = \frac{1}{2}$; $p = 0.0686$ (based on $z = 3$ and $d = 20$ cM). The broken line is drawn at $N_{\text{mean}} = 50$.

Table 2 Expected number of animals to be genotyped (EG_k) for breeding scheme (r), the minimum number (N_{\min}) of favourable animals required at the end of backcrosses being 50, $\alpha = 0.05$. (p = probability of inheriting the desired chromosomal region(s))

Backcrosses	No. of regions	Size of region (cM)	p	r				
				1	2	3	4	5
5	1	0	0.5	1761 (0.546)	—	—	—	—
		10	0.4524	1474 (0.639)	—	—	—	—
		20	0.4094	1293 (0.775)	—	—	—	—
	2	0	0.25	1440 (3.51)	1927 (1.01)	4247 (0.659)	—	—
		10	0.2047	2081 (7.94)	1869 (1.77)	3028 (0.907)	—	—
		20	0.1676	3482 (19.22)	2223 (3.56)	2628 (1.45)	4153 (0.924)	—
	3	0	0.125	9122 (75.62)	3935 (11.89)	3115 (3.79)	3436 (1.86)	4584 (1.20)
		10	0.0926	29332 (321.87)	9578 (46.53)	5556 (12.88)	4483 (5.24)	4462 (2.77)
		20	0.0686	106157 (1401.56)	28966 (194.51)	13535 (50.11)	8652 (18.38)	6798 (8.55)

The numbers in parentheses are n , i.e. n donor males and n donor females are mated to n recipient females and n recipient males to produce the F_1 generation.

ing to look at the situation in which only males are genotyped and selected. During the backcrossing generations these selected males can be mated to ungenotyped females from the recipient line. Males and females in the final backcross generation need to be genotyped and selected for the intercross. Table 5 gives the number of founder animals n that are needed in order to have a risk of $\alpha = 0.05$ of obtaining fewer than $N_{\min} = 50$ animals after five generations of backcrossing for the alternative of genotyping all animals ($r_F = 1$) or males only ($r_F = 0$). For higher values of r (≥ 4) the expected

amount of genotyping is smaller by genotyping males only. However, if there is a limit to r then the Table indicates there are situations where genotyping females would be optimal with respect to the amount of genotyping. Genotyping males only results in an increase in the number of animals being produced in the different generations as illustrated by the number of founder animals n in Table 5.

Table 3 Means and variances of N_5 for $\alpha = 0.50$ and $\alpha = 0.05$; $z = 3$ and $d = 20$ cM

r	$\alpha = 0.50$		$\alpha = 0.05$	
	$E(N_5)$	$\text{var}(N_5)$	$E(N_5)$	$\text{var}(N_5)$
1	50	76	67	101
2	50	107	70	150
3	50	166	76	252
4	50	269	85	460
5	50	444	99	876
6	50	725	118	1713
7	50	1163	146	3389

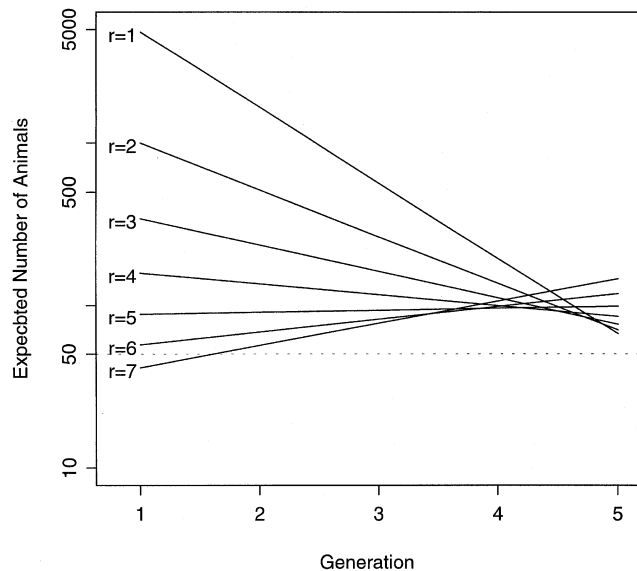


Fig. 3 Expected number of favourable animals ($E(N_i)$) at each generation of backcrossing for different breeding schemes (r). Fixed parameters were $N_{\min} = 50$ at $k = 5$ backcross generations with risk $\alpha = 0.05$; $\lambda = 5$; $s = \frac{1}{2}$; $p = 0.0686$ (based on $z = 3$ and $d = 20$ cM). The broken line is drawn at $N_{\min} = 50$.

Table 4 Expected number of animals to be genotyped (EG_k) for $r = 4, k = 5, z = 3$ and $d = 20$ cM

α	N_{\min}		
	25	50	100
0.50†	2537	5074	10148
0.10	4561	7705	13650
0.05	5351	8652	14833

† N_{\min} corresponds to N_{mean} when $\alpha = 0.50$.

Table 5 Expected number of animals to start the experiment (n) and expected number of animals to be genotyped (EG_k) for a situation with ($r_F = 1$) and without ($r_F = 0$) genotyping and selection of females for $\alpha = 0.05$, $N_{\min} = 50$, $k = 5$, $z = 3$ and $d = 20$ cm

r	$r_F = 1$		$r_F = 0$	
	n	EG_k	n	EG_k
1	1401.41	106 147	42 071.30	635 115
2	194.50	28 963	1403.97	53 656
3	50.11	13 534	200.54	15 458
4	18.39	8652	52.79	7714
5	8.55	6798	19.73	5309
6	4.73	6172	9.33	4492

Discussion

The model developed in this study can be used to assist the experimentalist in designing an MAI experiment. It provides important insights into parameters that need to be considered and the scale at which the introgression needs to be carried out in order to be successful. When the aim is to minimize the amount of genotyping there might be options to reduce this amount by using an adaptive breeding scheme when r is lower than the maximum reproductive rate of males. Using low levels of r during the initial generations and high values in later generations will lead to smaller values of EG_k than those given in Tables 1 and 2. The results obtained for a constant breeding scheme will serve as a good starting point for the calculations in that case. An additional reason for a departure from a constant breeding strategy might be if the breeder wishes to introgress two regions and exclude a third region. Once the breeder has eliminated that third region, p would be changed to the 'original' level for two regions. This situation can be accommodated by allowing p to vary over the generations using this adaptive approach. Setting the mean (N_{mean}) at 50, one would expect a number of progeny of 125 after one generation of intercrossing. Introgressing $z = 3$ regions each of length 20 cm, then $(\frac{1}{2}e^{-0.2})^6 = 0.47\%$ (corresponding to 0.6 animal) of the offspring is expected to be homozygous for the three-locus genotype. This demonstrates that nominating an expected number of 50 animals at the end of the backcross generations is too low. Now consider the case where the minimum (N_{\min}) is set at 50, and assume that the breeder used each male $r = 7$ times during the backcrossing process (Table 3). This would result in approximately 1.7 animals, still a low number. Nevertheless, opportunities might exist for the breeder to repeat the intercross process, or addi-

tionally to intercross any heterozygous animals resulting from the intercross.

During the experiment, the experimenter will try to modify the scheme when the results would permit this. Genotyping of animals will be stopped as soon as sufficient animals with a desirable genotype have been found. It is clearly demonstrated that the amount of genotyping can be reduced by using such an adaptive breeding scheme (Table 5). There might be practical limitations in implementing a scheme which results in minimum genotyping. To make efficient use of genotyping facilities, animals are genotyped in batches rather than individually. Because of culling of surplus animals at an early age, it may be difficult to go back to genotyping more males or females if insufficient males with favourable genotypes are found. These practical constraints need to be taken into account in executing a MAI programme in order to make efficient use of experimental and genotyping facilities.

The number of founder animals from the donor as well as the recipient strain (n female and n male) are represented as a decimal number (Tables 1 and 3). In practice this cannot be realized. Imposing a minimum value of 1 for n would under a constant breeding strategy lead to too large numbers at the end of the experiment. Values smaller than 1 indicate that during the initial generations of the backcrossing process the number of selected animals should be kept constant instead of increased, perhaps by not using all the available favourable animals.

In MAI the major aim of backcrosses is the recovery of the recipient genome (Hospital *et al.*, 1992; Hillel *et al.*, 1993; Yancovich *et al.*, 1996). The number of generations of backcrossing is determined by the desired proportion of genes coming from the donor line for other parts of the genome. With five backcrosses, the expected proportion of the donor genome in the recipient animal is still 1.562%, and it is 0.049% after 10 backcross generations (Soller & Plotkin-Hazan, 1977; Hillel *et al.*, 1990). In this paper, selection is entirely based on the genotypic information for the identified regions. To reduce the number of backcross generations, genomic selection against the donor genes from other parts of the genome may be applied (Hospital *et al.*, 1992; Hillel *et al.*, 1993; Yancovich *et al.*, 1996). Introgressing alleles simultaneously with genomic selection on the recipient genotype (Visscher *et al.*, 1996) would be another way of reducing backcross generations. These techniques accelerate the recovery of the recipient genome but will lead to an increase in the number of animals to be produced within a genera-

tion and the amount of genotyping per animal; indeed the genomic selection is applied to those individuals known to carry the relevant segments.

One of the goals of this study is to minimize the number of animals to be genotyped throughout the experiment. If we ignore the risk of obtaining too few animals in the target backcross generation, then Table 1 indicates that we need to select a strategy with a small n and large r . In fact the expected amount of genotyping would be minimized further by a low constant population size up to the second last generation ($k-1$), then the highest possible increase in the last generation (i.e. high r). However, when risk is considered, this may no longer be the optimal strategy. As is seen in Table 2, the expected amount of genotyping is minimized by selecting a low or intermediate value of r ; higher values of r can lead to an increase in the amount of genotyping involved. An option to minimize the amount of genotyping would be to only type, select and use favourable males during the backcross process. From Table 5 it follows that when the reproductive rate of males is sufficiently large, the expected total amount of genotyping can be reduced by genotyping males only, at the expense of producing a larger number of individuals. In designing an experiment, the amount of genotyping needs to be balanced against the number of experimental animals needed. One will first determine the number of backcross generations that is needed to obtain the acceptable proportion of donor genome in the newly formed line. To determine the optimum choice for the other parameters one will need to incorporate the costs of producing and keeping an animal, the genotyping costs, the costs of other resources and the time required to achieve the target.

Alternatively, the choice of 'optimum' may also be driven by the genotyping capacity of the laboratory. Low values of r are usually associated with high rates of population decrease, resulting in a large volume of genotyping being conducted in the early generations (Figs 2 and 3). Conversely, when risk is considered, a high value of r may lead to an increased amount of genotyping in the later generations (Fig. 3). Clearly, for the parameters nominated for Fig. 3, a choice of $r=5$ would be optimal in terms of approximate constant genotyping over all backcross generations.

This model was developed to assist in the design stage of a marker-assisted introgression experiment of trypanoresistance genes in mice. The findings are now to be exploited in this programme. More generally, this model when implemented in a computer spreadsheet programme, allows the breeder to assess

the effect of changes in biological and design parameters on the population growth, as well as effects of random causes. Such a system could be incorporated into a computer-aided decision-support system for use by breeders. A spreadsheet program (Microsoft Excel) for this model is available from the corresponding author.

Acknowledgements

O.D.K. is grateful to WOTRO (Netherlands Foundation for the Advancement of Tropical Research) and ILRI (International Livestock Research Institute) for funding this Ph.D. project. He also acknowledges the leave provided by INRAB (Institut National des Recherches Agricoles du Bénin) to undertake the research. P.C.T. acknowledges the study leave provided by the University of Sydney to undertake this research at Wageningen Agricultural University. The authors would also like to thank Luc Janss, Alan Teale and two anonymous reviewers for helpful comments on an earlier version of the manuscript.

References

- FOULLEY, J. L., GIANOLA, D. AND IM, S. 1987. Genetic evaluation of traits distributed as Poisson-binomial with reference to reproductive characters. *Theor. Appl. Genet.*, **73**, 870–877.
- GAMA, L. T., SMITH, C. AND GIBSON, J. P. 1992. Transgene effects, introgression strategies and testing schemes in pigs. *Anim. Prod.*, **54**, 427–440.
- GROEN, A. F. AND SMITH, C. 1995. A stochastic simulation study of the efficiency of marker-assisted introgression in livestock. *J. Anim. Breed. Genet.*, **112**, 161–170.
- HALDANE, J. B. S. 1919. The combination of linkage values and the calculation of distance between the loci of linked factors. *J. Genet.*, **8**, 299–309.
- HILLEL, J., SCHAAP, T., HABERFELD, A., JEFFREYS, A. J., PLOTZKY, Y., CAHANER, A. AND LAVI, U. 1990. DNA fingerprints applied to gene introgression in breeding programs. *Genetics*, **124**, 783–789.
- HILLEL, J., VERRINDER GIBBINS, A. M., ETCHES, R. J. AND SHAVER, D. McO. 1993. Strategies for the rapid introgression of a specific gene modification into a commercial poultry flock from a single carrier. *Poultry Sci.*, **72**, 1197–1211.
- HOSPITAL, F., CHEVALET, C. AND MULSANT, P. 1992. Using markers in gene introgression breeding programs. *Genetics*, **132**, 1199–1210.
- KEMP, S. J., IRAQI, F., DARVASI, A., SOLLER, M. AND TEALE, A. J. 1997. Localization of genes controlling resistance to trypanosomiasis in mice. *Nature (Genet.)*, **16**, 194–196.
- MATOS, C. A. P., THOMAS, D. L., GIANOLA, D., TEMPELMAN, R. J. AND YOUNG, L. D. 1997. Genetic analysis of discrete reproductive traits in sheep using linear and nonlinear

- models: I. Estimation of genetic parameters. *J. Anim. Sci.*, **75**, 76–87.
- MOOD, A. M., GRAYBILL, F. A. AND BOES, D. C. 1974. *Introduction to the Theory of Statistics*. McGraw-Hill, New York.
- MORRISON, W. I., ROELANTS, G. E., MAYOR-WITHEY, K. S. AND MURRAY, W. I. 1978. Susceptibility of inbred strains of mice to *Trypanosoma congolense*: correlation with changes in spleen lymphocyte populations. *Clin. Exp. Immunol.*, **32**, 25–40.
- MURRAY, M. AND TRAIL, J. C. M. 1984. Genetic resistance to animal trypanosomiasis in Africa. *Prev. Vet. Med.*, **2**, 541–551.
- RENSHAW, E. 1991. *Modelling Biological Populations in Space and Time*. Cambridge University Press, Cambridge.
- SOLLER, M. AND PLOTKIN-HAZAN, J. 1977. The use of marker alleles for the introgression of linked quantitative alleles. *Theor. Appl. Genet.*, **51**, 133–137.
- TEALE, A. J. 1993. Improving control of livestock diseases. Animal biotechnology in the Consultative Group for International Agricultural Research. *Bioscience*, **43**, 475–483.
- VISSCHER, P. A., HALEY, C. S. AND THOMPSON, R. 1996. Marker-assisted introgression in backcross breeding programs. *Genetics*, **144**, 1923–1932.
- YANCOVICH, A., LEVIN, I., CAHANAR, A. AND HILLEL, J. 1996. Introgression of the avian naked neck gene assisted by DNA fingerprints. *Anim. Genet.*, **27**, 149–155.

Appendix: Derivation of means, variances, and covariances

F_1 generation ($t = 0$)

Because there are $2n$ donor animals mated with $2n$ recipient animals, each mating giving rise to a litter with mean size λ , all of which are favourable, the mean number of male and female favourable animals, respectively, will be $E(M_0) = 2ns\lambda$ and $E(F_0) = 2n(1-s)\lambda$. Also, $\text{var}(M_0) = E(M_0)$ and $\text{var}(F_0) = E(F_0)$ from the Poisson litter size assumption, and $\text{cov}(M_0, F_0) = 0$ because of the independence (for a given n). Clearly, $E(N_0) = \text{var}(N_0) = 2n\lambda$.

Backcross generations ($t = 1, 2, \dots$)

Because there are $2n$ donor animals mated with $2n$ recipient animals, each mating giving rise to a litter with mean size λ , all of which are favourable, the mean number of male and female favourable animals, respectively, will be $E(M_0) = 2ns\lambda$ and $E(F_0) = 2n(1-s)\lambda$. Also, $\text{var}(M_0) = E(M_0)$ and $\text{var}(F_0) = E(F_0)$ from the Poisson litter size assumption,

and $\text{cov}(M_0, F_0) = 0$ because of the independence (for a given n). Clearly, $E(N_0) = \text{var}(N_0) = 2n\lambda$.

Backcross generations ($t = 1, 2, \dots$)

Because the M_{t-1} males are used r times and the F_{t-1} females only once, there will be $rM_{t-1} + F_{t-1}$ litters resulting from matings with the recipient line, each litter with a mean and variance of λ . However, there is now a probability p of an offspring being favourable, so the conditional moments are

$$E(M_t | M_{t-1}, F_{t-1}) = ps\lambda(rM_{t-1} + F_{t-1})$$

and

$$E(F_t | M_{t-1}, F_{t-1}) = p(1-s)\lambda(rM_{t-1} + F_{t-1}),$$

with

$$\text{var}(M_t | M_{t-1}, F_{t-1}) = E(M_t | M_{t-1}, F_{t-1}),$$

$$\text{var}(F_t | M_{t-1}, F_{t-1}) = E(F_t | M_{t-1}, F_{t-1}),$$

and

$$\text{cov}(M_t, F_t | M_{t-1}, F_{t-1}) = 0.$$

The marginal (unconditional) moments are obtained using standard results (see, e.g., Mood *et al.*, 1974, p. 157), namely, means of

$$E(M_t) = ps\lambda \times [rE(M_{t-1}) + E(F_{t-1})],$$

$$E(F_t) = p(1-s)\lambda \times [rE(M_{t-1}) + E(F_{t-1})],$$

and variances

$$\text{var}(M_t) = (ps\lambda)^2 \times [r^2 \text{var}(M_{t-1}) + \text{var}(F_{t-1}) + 2r \text{cov}(M_{t-1}, F_{t-1})] + ps\lambda \times [rE(M_{t-1}) + E(F_{t-1})],$$

$$\text{var}(F_t) = (p(1-s)\lambda)^2 \times [r^2 \text{var}(M_{t-1}) + \text{var}(F_{t-1}) + 2r \text{cov}(M_{t-1}, F_{t-1})] + p(1-s)\lambda \times [rE(M_{t-1}) + E(F_{t-1})],$$

and covariance

$$\text{cov}(M_t, F_t) = (p\lambda)^2 s(1-s) [r^2 \text{var}(M_{t-1}) + \text{var}(F_{t-1}) + 2r \text{cov}(M_{t-1}, F_{t-1})].$$

Also,

$$E(N_t) = p\lambda \times [rE(M_{t-1}) + E(F_{t-1})],$$

and

$$\text{var}(N_t) = (p\lambda)^2 \times [r^2 \text{var}(M_{t-1}) + \text{var}(F_{t-1}) + 2r \text{cov}(M_{t-1}, F_{t-1})] + p\lambda \times [rE(M_{t-1}) + E(F_{t-1})].$$

These means, variances, and covariance for M_t, F_t , and hence N_t , can be obtained by calculating the moments recursively. Alternatively, solutions for these may be obtained explicitly, as reported in the main text. These results may be proven by induction.